Ad Hoc Microphone Array Calibration: Euclidean Distance Matrix Completion Algorithm and Theoretical Guarantees

Mohammad J. Taghizadeh[a,b], Reza Parhizkar[b], Philip N. Garner[a], Hervé Bourlard[a,b]

Emails: {mohammad.taghizadeh, phil.garner, herve.bourlard}@idiap.ch, reza.parhizkar@epfl.ch

[a]*Idiap Research Institute, Martigny, Switzerland*
[b]*École Polytechnique Fédérale de Lausanne (EPFL), Switzerland*

Abstract

This paper addresses the problem of ad hoc microphone array calibration where only partial information about the distances between the microphones is available. We construct a matrix consisting of the pairwise distances and propose to estimate the missing entries based on a novel Euclidean distance matrix completion algorithm by alternative low-rank matrix completion and projection onto the Euclidean distance space. This approach confines the recovered matrix to the EDM cone at each iteration of the matrix completion algorithm. The theoretical guarantees of the calibration performance are obtained considering the random and locally structured missing entries as well as the measurement noise on the known distances. The proposed approach has been evaluated using real data recordings where the distance of the close-by microphones are estimated based on the coherence model of an enclosure diffuse noise field. The results confirm that the proposed algorithm outperforms the state-of-the-art calibration techniques.

Keywords: Ad hoc microphone array calibration, Diffuse noise coherence model, Euclidean distance matrix completion, Cadzow algorithm, EDM cone

---

1. Introduction

Ad hoc microphone arrays consist of a set of sensor nodes spatially distributed over the acoustic field, in an ad hoc fashion. Processing of the data acquired with distributed sensors involves challenges attributed to the issues such as asynchronous sampling and unknown microphone positions. In this paper, we address the problem of finding the microphone positions also referred to as *microphone calibration*. Finding the correct positioning of the microphones plays a key role in distant audio processing tasks such as source localization and tracking [1], high-quality acquisition for distant speech separation [2] and recognition [3]. Recent advances in mobile computing and communication technologies enable using cell phones, PDA's or tablets as a flexible acquisition set-up providing an ad hoc network of microphones. However, the unknown prior information on relative positions of the microphones is a key problem to achieve effective data processing. In the following, we review some of the prior approaches for microphone array calibration.

Sachar et al. [4] presented a set-up using a pulsed acoustic excitation generated by five domed tweeters. The transmit times between speakers and microphones were used to calibrate the microphones. Raykar et al. [5] considered a maximum length sequence or chirp signal in a distributed computing platform. The time difference of arrival of microphone signals were then computed by cross-correlation and used for estimating the microphone locations. Since the original signal is known, these techniques are robust to noise and reverberation. Flanagan and Bell [6] proposed a joint source-sensor localization scheme based on the Weiss-Friedlander method where the sensor location and direction of arrival of the sources are estimated alternately until the algorithm is converged. Shang et al. [7] formulated an energy-based method for maximum likelihood estimation of joint source-sensor positions. This method requires several active sources for accurate localization. Recently, McCowan and Lincoln [8] exploited characteristics of a diffuse noise field model for microphone calibration. A diffuse noise field is characterized by noise signals that propagate with equal probability in all directions and its coherence is defined by the sinc function of the distance of the two microphones. The distances can thus be estimated by fitting the computed noise coherence with the sinc function in the least square sense. In this paper, we use the coherence model of a diffuse field for pairwise distance estimation due to its practical assumptions for distant audio applications [9, 10, 11] and no requirement for activating a specific source signal. However, the proposed algorithm and theoretical results are applicable for calibration of a general ad hoc microphone array network.

The state-of-the-art techniques for distance estimation and microphone calibration are usually appropriate for conventional compact arrays. Estimation of the pairwise distances becomes unreliable as the distances between the microphones are increased. Hence, the purpose of this paper is to enable microphone calibration when some of the pairwise distances are missing. The matrix consisted of the squared pairwise distances has very low rank (explained in Section 3.1). The low-rank property has been investigated in the past years to devise efficient optimization schemes for matrix completion, i.e. recovering a low-rank matrix from randomly known entries. Candès et al. [12] showed that a small random fraction of the entries are sufficient to reconstruct a low-rank matrix *exactly*. Keshavan et al. proposed a matrix completion algorithm known as OptSpace and showed its optimality [13]. Furthermore, they proved that their algorithm is robust against noise [14]. Drineas et al. [15] exploited the low rank property to recover the distance matrix. However, they assume a nonzero probability of obtaining accurate distances for any pair of sensors regardless of their distance. This assumption severely restricts the applicability of their result for the microphone array calibration problem.

In this paper, we first estimate the pairwise distances of the microphones in close proximity using the coherence model of the signals of the two microphones in a diffuse noise field using the improved method described in [16]; this approach implies a local connectivity constraint as the pairwise distances of the further microphones can not be estimated. We construct a matrix of all the pairwise distances with missing entries corresponding to the unknown distances. We exploit the low-rank property of the square of this matrix to enable estimation of all the pairwise distances using matrix completion approach. The goal of this paper is to show that exploiting the combination of the rank condition of Euclidean distance matrices (EDMs), similarity in the measured distances, and projection on the EDM cone enables us to estimate the microphone array geometry effectively from only partial measurements of the pairwise distances. To this end, we show that matrix completion is capable of finding the missing entries in our scenario and provide the theoretical guarantees to bound the error for ad hoc microphone calibration considering the local connectivity of the noisy known entries. To increase the accuracy, we incorporate the properties of EDMs in the matrix completion algorithm. We show that imposing EDM characteristics on matrix completion improves the robustness and accuracy of extracting the ad hoc microphone geometry.

The rest of the paper is organized as follows. In Section 2, we explain how pairwise distances of the microphones are estimated using the coherence model of the diffuse noise field as an example use case of the proposed method. Section 3 describes the mathematical basis and the model used for the calibration problem. The proposed Euclidean distance matrix completion algorithm is elaborated in Section 4. Section 5 is dedicated to the theoretical guarantees for ad hoc microphone array calibration based on matrix completion. The related methods are investigated in Section 6 and the experimental results on real data recordings are presented in section 7. The conclusions are drawn in Section 8.

2. Example Use Case

We consider N microphones located at random positions on a large circular table in a meeting room with homogeneous reverberant acoustics. In the time intervals that there is no active speaker, diffuse noise is the dominant signal in the room. The table is located at the center of the room, hence deviation from diffuseness near the walls can be neglected. Based on the theory of the diffuse noise model, the distance of each two close microphones can be estimated by computing the coherence of their signals Γ, and fitting a sinc function based on the relation expressed as

$$\Gamma_{ij}(\omega) = \text{sinc}\left(\frac{\omega d_{ij}}{c}\right), \tag{1}$$

where ω is the frequency, d_{ij} is the distance between the two microphones i and j and c is the speed of sound [17]. Figure 1 represents an example of the coherence and the fitted sinc function.

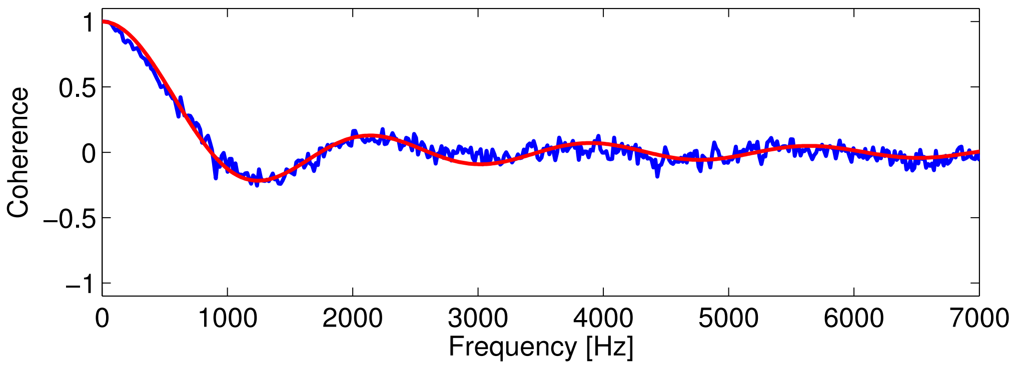

Figure 1: Coherence of the signal of two microphones at $d_{ij} = 20$ cm and the fitted sinc function using real data recordings.

In practice, if the distance between the sensors is large (e.g. greater than 73 cm [16]) we observe deviations from the diffuse characteristics. The maximum distance that can be computed by this method is assumed to be d_{max}. Therefore, pairwise distances greater than d_{max} are missing implying a locality structure in the missing entries in the distance matrix $\boldsymbol{D}$ consisted of the pairwise distances. This locality constraint in distance estimation is a typical problem in ad hoc microphone arrays [18]. In addition, the computation algorithm can lead to deviation from the model resulting in unreliable estimates of the short distances causing random missing entries in $\boldsymbol{D}$. Furthermore, the known entries are noisy due to measurement inaccuracies and violation of diffuseness.

3. Problem Formulation

3.1. Distance Matrix

Consider a distance matrix $\boldsymbol{D}_{N\times N}$ consisting of the distances between N microphones constructed as

$$\boldsymbol{D} = \left[d_{ij}\right], \quad d_{ij} = \left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|, \quad i, j \in \{1, \ldots, N\}, \tag{2}$$

where d_{ij} is the Euclidean distance between microphones i and j located at $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. Therefore, $\boldsymbol{D}$ is a symmetric matrix and it is often full rank.

Let $\boldsymbol{X}_{N\times\zeta}$ denote the position matrix whose i^{th} row, $\boldsymbol{x}_i^T \in \mathbb{R}^\zeta$, is the position of microphone i in ζ-dimensional Euclidean coordinate where microphones are deployed and $.^T$ denotes the transpose operator. By squaring the elements of $\boldsymbol{D}$, we construct a matrix $\boldsymbol{M}_{N\times N}$ which can be written as

$$\boldsymbol{M} = \mathbf{1}_N \boldsymbol{\Lambda}^T + \boldsymbol{\Lambda} \mathbf{1}_N{}^T - 2\boldsymbol{X}\boldsymbol{X}^T \ , \tag{3}$$

where $\mathbf{1}_N \in \mathbb{R}^N$ is the all ones vector and $\boldsymbol{\Lambda} = (\boldsymbol{X} \circ \boldsymbol{X})\mathbf{1}_\zeta$ where $\circ$ denotes the Hadamard product. We observe that $\boldsymbol{M}$ is the sum of three matrices of rank 1, 1 and at most ζ respectively. Therefore, the rank of the squared distance matrix constructed of the elements $\boldsymbol{M}_{ij} = \left[d_{ij}^2\right]$ is at most $\zeta + 2$ [15]. For instance, if the microphones are located on a plane or shell of a sphere, $\boldsymbol{M}$ has rank 4 and if they are placed on a line or circle, the rank is exactly 3. Hence, there is significant dependency between the elements of $\boldsymbol{M}$ and exploiting this low-rank property is the core of the proposed method in this paper.

3.2. Objective

The noisy estimates of the pairwise distances is modeled as

$$\tilde{d}_{ij} = d_{ij} + w_{ij} \quad ; \quad \widetilde{\boldsymbol{D}} = \boldsymbol{D} + \boldsymbol{W} \ , \tag{4}$$

where w_{ij} is the measurement noise for distance d_{ij} and $\boldsymbol{W}$ is the corresponding measurement noise matrix. We introduce a noise matrix on the squared distance matrix as

$$\boldsymbol{Z} = \widetilde{\boldsymbol{M}} - \boldsymbol{M} = \widetilde{\boldsymbol{D}} \circ \widetilde{\boldsymbol{D}} - \boldsymbol{D} \circ \boldsymbol{D} \ , \tag{5}$$

where $\widetilde{\boldsymbol{M}}$ is the noisy squared distance matrix.

As described in Section 2, there are two kinds of missing entries. The first group is consisted of the structured missing entries corresponding to the distances greater than d_{max}. We denote this group by S defined as

$$S = \{(i, j) : \ d_{ij} \geq d_{max}\} \ , \tag{6}$$

where $d_{ij} = \left\|\mathbf{x}_i - \mathbf{x}_j\right\|$. These structured missing entries are denoted by a matrix

$$\boldsymbol{D}^s_{ij} = \begin{cases} \boldsymbol{D}_{ij} & \text{if } (i, j) \in S \\ 0 & \text{otherwise} \end{cases}$$

Thus, the noiseless recognized pairwise distance matrix is given by

$$\boldsymbol{D}^{\bar{s}} = \boldsymbol{D} - \boldsymbol{D}^s \ ,$$

and we obtain the known squared distance matrix as

$$\begin{aligned} \boldsymbol{M}^s &= \boldsymbol{D}^s \circ \boldsymbol{D}^s \\ \boldsymbol{M}^{\bar{s}} &= \boldsymbol{D}^{\bar{s}} \circ \boldsymbol{D}^{\bar{s}} = \boldsymbol{M} - \boldsymbol{M}^s \ . \end{aligned} \tag{7}$$

Considering the noise on the known entries, we obtain

$$\widetilde{\boldsymbol{M}}^{\bar{s}} = \boldsymbol{M}^{\bar{s}} + \boldsymbol{Z}^{\bar{s}} \ , \tag{8}$$

where $\boldsymbol{Z}^{\bar{s}}$ denotes the noise on the known entries in the squared distance matrix.

For modeling the random missing entries, we assume that each entry is sampled with probability p. Sampling can be introduced by a projection operator on an arbitrary matrix $\boldsymbol{Q}_{N\times N}$, given by

$$\Psi_E(\boldsymbol{Q})_{ij} = \begin{cases} \boldsymbol{Q}_{ij} & \text{if } (i, j) \in E \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

where $E \subseteq [N] \times [N]$ denotes the known entries after random erasing process and has cardinality $|E| \approx pN^2$. Therefore, the final recognized squared distance matrix is given by

$$\boldsymbol{M}^E = \Psi_E(\widetilde{\boldsymbol{M}}^{\bar{s}}) \ . \tag{10}$$

The goal of the matrix recovery algorithm is to find the missing entries and remove the noise, given matrix $\boldsymbol{M}^E$.

3.3. Noise Model

The level of noise in extracting the pairwise distances, w_{ij} in (4), increases as the distances grow [16]. We model this effect through

$$\boldsymbol{W} = \boldsymbol{\Upsilon} \circ \boldsymbol{D} \,,$$

where the normalized noise matrix $\boldsymbol{\Upsilon}_{N\times N}$ is consisted of entries with sub-Gaussian distribution with variance ς^2, thus [14]

$$\mathbb{P}(|\boldsymbol{\Upsilon}_{ij}| \geq \beta) \leq 2\,\mathrm{e}^{-\frac{\beta^2}{2\varsigma^2}} \,. \tag{11}$$

Based on (8), $\boldsymbol{Z}^{\bar{s}}_{ij} = 2d^2_{ij}\boldsymbol{\Upsilon}_{ij} + d^2_{ij}\boldsymbol{\Upsilon}^2_{ij}$; the physical setup confines $|\boldsymbol{Z}^{\bar{s}}_{ij}| \leq 2a$ where a is the radius of the table.

3.4. Evaluation Measure

Extracting the absolute position of the microphones deployed in ζ dimensional space requires at least $\zeta + 1$ anchor points in addition to the distance matrix. Therefore, in a scenario that the only available information are pairwise distances, the evaluation measure must quantify the error in estimation of the *relative* position of the microphones thus robust to the rigid transformations (translation, rotation and reflection). Hence, we quantify the distance between the actual locations $\boldsymbol{X}$ and estimated locations $\hat{\boldsymbol{X}}$ as [19]

$$\begin{aligned} \mathrm{dist}(\boldsymbol{X}, \hat{\boldsymbol{X}}) &= \frac{1}{N} \left\| \boldsymbol{J}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{J} - \boldsymbol{J}\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^T\boldsymbol{J} \right\|_{\mathrm{F}} \,, \\ \boldsymbol{J} &= \mathbb{I}_N - (1/N)\mathbf{1}_N\mathbf{1}_N^T \end{aligned} \tag{12}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm and $\mathbb{I}_N$ is the $N \times N$ identity matrix. The distance measure stated in (12) is useful to compare the performance of different methods in terms of microphone array geometry estimation.

Table 1 summarizes the set of important notations.

Table 1: Summary of notations.

Symbol	Meaning	Symbol	Meaning
N	number of microphones	$\boldsymbol{D}$	complete noiseless distance matrix
a	radius of the circular table on which microphones are distributed	$\boldsymbol{M}$	squared distance matrix
ς	normalized standard deviation of noise	$\widetilde{\boldsymbol{M}}$	noisy squared distance matrix
Ψ_E	projection into matrices with entries on index set E	$\hat{\boldsymbol{M}}$	estimated squared distance matrix
$\mathcal{P}_e$	projection to EDM cone	$\boldsymbol{Z}$	noise matrix
p	probability of having random missing entries	$\boldsymbol{M}^E$	observed matrix
d_{max}	radius of the circle defining structured observed entries	$\boldsymbol{X}$	positions matrix
$\boldsymbol{M}^s$	distance matrix with observed entries on index set S	$\hat{\boldsymbol{X}}$	estimated positions matrix

4. Euclidean Distance Matrix Completion Algorithm

The approach proposed in this paper exploits low-rank matrix completion and incorporates the EDM properties for recovering the distance matrix.

4.1. Matrix Completion

We recall our problem of having N microphones distributed on a space of dimension ζ. Hence, the squared distance matrix $\boldsymbol{M}$ has rank $\eta = \zeta + 2$, but it is only partially known. The objective is to recover $\boldsymbol{M}_{N\times N}$ of rank $\eta \ll N$ from a sampling of its entries without having to ascertain all the N^2 entries, or collect N^2 or more measurements about $\boldsymbol{M}$. The approach proposed through *matrix completion* relies on the fact that a low-rank data matrix carries much less information than its ambient dimension implies. Intuitively, as the matrix $\boldsymbol{M}$ has $(2N-\eta)\eta$ degrees of freedom[1], we need to know at least ηN of the row entries as well as ηN of the column entries reduced by η^2 number of the repeated values to recover the entire elements of $\boldsymbol{M}$.

Given $\boldsymbol{M}^E$ defined in (10), the matrix completion recovers an estimate of the distance matrix $\hat{\boldsymbol{M}}$ through the following optimization

$$\begin{aligned} &\text{Minimize} && \operatorname{rank}(\hat{\boldsymbol{M}}) \\ &\text{subject to} && \hat{\boldsymbol{M}}_{ij} = \boldsymbol{M}_{ij}, \qquad (i,j) \in E \end{aligned} \tag{13}$$

[1] The degrees of freedom can be estimated by counting the parameters in the singular value decomposition (the number of degrees of freedom associated with the description of the singular values and of the left and right singular vectors). When the rank is small, this is considerably smaller than N^2 [20].

In this paper, we use the procedure of OptSpace proposed by Keshavan et al. [14] for estimating a matrix given the desired rank η. This algorithm is implemented in three steps: (1) Trimming, (2) Projection and (3) Minimizing the cost function.

In the trimming step, a row or a column is considered to be over-represented if it contains more samples than twice the average number of samples per row or column. These rows or columns can dominate the spectral characteristics of the observed matrix $\boldsymbol{M}^E$. Thus, some of their entries are removed uniformly at random from the observed matrix. Let $\widetilde{\boldsymbol{M}}^E$ be the resulting matrix of this trimming step.

In the projection step, we first compute the singular value decomposition (SVD) of $\widetilde{\boldsymbol{M}}^E$ thus

$$\widetilde{\boldsymbol{M}}^E = \sum_{i=1}^{N} \sigma_i(\widetilde{\boldsymbol{M}}^E)\boldsymbol{U}_{.i}\boldsymbol{V}_{.i}^T \ , \tag{14}$$

where $\sigma_i(\cdot)$ denotes the i^{th} singular value of the matrix and $\boldsymbol{U}_{.i}$ and $\boldsymbol{V}_{.i}$ designate the i^{th} column of the corresponding SVD matrices. Then, the rank-η projection, $\mathcal{P}_\eta(\cdot)$ returns the matrix obtained by setting to 0 all but the η largest singular values as

$$\mathcal{P}_\eta(\widetilde{\boldsymbol{M}}^E) = (N^2/|E|) \sum_{i=1}^{\eta} \sigma_i(\widetilde{\boldsymbol{M}}^E)\boldsymbol{U}_{.i}\boldsymbol{V}_{.i}^T = \boldsymbol{U}_0\boldsymbol{S}_0\boldsymbol{V}_0^T \ . \tag{15}$$

Starting from the initial guess provided by the rank-η projection $\mathcal{P}_\eta(\widetilde{\boldsymbol{M}}^E)$, $\boldsymbol{U} = \boldsymbol{U}_0$, $\boldsymbol{V} = \boldsymbol{V}_0$ and $\boldsymbol{S} = \boldsymbol{S}_0$, the final step solves a minimization problem stated as follows: Given $\boldsymbol{U} \in \mathbb{R}^{N\times\eta}, \boldsymbol{V} \in \mathbb{R}^{N\times\eta}$, find

$$\begin{aligned} F(\boldsymbol{U},\boldsymbol{V}) &= \min_{\boldsymbol{S}\in\mathbb{R}^{\eta\times\eta}} \mathcal{F}(\boldsymbol{U},\boldsymbol{V},\boldsymbol{S}) \,, \\ \mathcal{F}(\boldsymbol{U},\boldsymbol{V},\boldsymbol{S}) &= \frac{1}{2} \sum_{(i,j)\in E} (\boldsymbol{M}_{ij} - (\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T)_{i,j})^2 \end{aligned} \tag{16}$$

$F(\boldsymbol{U},\boldsymbol{V})$ is determined by minimizing the quadratic function $\mathcal{F}$ over $\boldsymbol{S}, \boldsymbol{U}, \boldsymbol{V}$ estimated by gradient decent with line search in each iteration. This last step tries to get us as close as possible to the correct low-rank matrix $\boldsymbol{M}$.

4.2. Cadzow Projection to the Set of EDM Properties

The classic matrix completion algorithm as described above recovers a low-rank matrix with elements as close as possible to the known entries. However, the recovered matrix does not necessarily correspond to a Euclidean distance matrix; for example, EDMs are symmetric with zero diagonal elements. These properties are not incorporated in the matrix completion algorithm. Hence, we modify the aforementioned procedure to have, as output, matrices that are closer to EDMs [16].

To this end, we apply a Cadzow-like method. The Cadzow algorithm [21] (also known as Papoulis-Gershberg) is a method for finding a signal which satisfies a composite of properties by iteratively projecting the signal into the property sets. We modify the matrix completion algorithm by inserting an extra step at each iteration. In the classic version of this algorithm a simple rank-η approximation is used as the starting point for the iterations using gradient descent on (16). After each iteration of the gradient descent, we apply the transformation $\mathcal{P}_c : \mathbb{R}^{N\times N} \longmapsto \mathbb{S}_h^N$ on the obtained matrix where $\mathbb{S}_h^N$ is the space of symmetric, positive hollow matrices, to make sure that the output satisfies the following properties

$$\hat{\boldsymbol{M}} \in \mathbb{S}_h^N \Longleftrightarrow \begin{cases} d_{ij} = 0 \Leftrightarrow \boldsymbol{x}_i = \boldsymbol{x}_j \\ d_{ij} > 0,\ i \neq j \\ d_{ij} = d_{ji} \end{cases} \tag{17}$$

for $i, j \in [N]$; nonnegativity and symmetry are achieved by setting all the negative elements to zero and averaging the symmetric elements.

4.3. Matrix Completion with Projection onto the EDM cone

In section 4.2, three characteristics of EDMs are employed through the Cadzow projection to reduce the reconstruction error of the distance matrix. In order to increase the accuracy even further, we propose to project to the cone of Euclidean distance matrix, $\mathbb{EDM}^N$, at each iteration of the algorithm. In other words, after one step of the gradient descent method on the Cartesian product of two Grassmannian manifolds $\mathcal{G}$, we apply a projection, $\mathcal{P}_e : \mathbb{R}^{N\times N} \longmapsto \mathbb{EDM}^N$ to

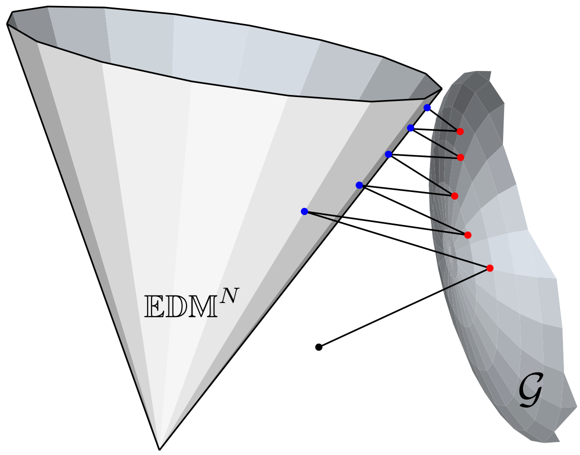

Figure 2: Matrix completion with projection onto the EDM cone.

decrease the distance between the estimated matrix and the EDM cone. This is visualized in Figure 2. Note that the demonstration of the cone and the manifold are not mathematically accurate and only serve as visualizations (The dimension of the cone and the manifold are too large to be demonstrated graphically). The projected matrix must satisfy the following EDM properties [22]

$$\hat{\boldsymbol{M}} \in \mathbb{EDM}^N \iff \begin{cases} -z^T \hat{\boldsymbol{M}} z \geq 0 \\ \mathbf{1}^T z = 0 \\ (\forall \|z\| = 1) \\ \hat{\boldsymbol{M}} \in \mathbb{S}_h^N \end{cases} \tag{18}$$

The EDM properties include the triangle inequality, thus

$$d_{ij} \leq d_{ik} + d_{kj}, \quad i \neq j \neq k \ , \tag{19}$$

as well as the relative-angle inequality; $\forall i, j, l \neq k \in [N], i < j < l$, and for $N \geq 4$ distinct points $\{\boldsymbol{x}_k\}$, the inequalities

$$\begin{gathered} \cos(\tau_{jkl}+\tau_{lkj}) \leq \cos\tau_{ikj} \leq \cos(\tau_{ikl} - \tau_{lkj}) \\ 0 \leq \tau_{ikl},\ \tau_{lkj},\ \tau_{ikj} \leq \pi \end{gathered} \tag{20}$$

where τ_{ikj} denotes the angle between vectors at $\boldsymbol{x}_k$ and it is satisfied at each position $\boldsymbol{x}_k$.

The projection $\mathcal{P}_e$ must map the output of matrix completion to the closest matrix on $\mathbb{EDM}^N$ with the properties listed in (18). The projection onto $\mathbb{S}_h^N$ is achieved by $\mathcal{P}_c$ implemented via Cadzow; thereby, we define $(\boldsymbol{U}_c, \boldsymbol{V}_c, \boldsymbol{S}_c) = \mathcal{P}_c(\boldsymbol{U}^{k+1/2}, \boldsymbol{V}^{k+1/2}, \boldsymbol{S}^{k+1/2})$. To achieve the full EDM properties, we search in the EDM cone using a cost function defined as

$$\mathcal{H}(\boldsymbol{X}) = \left\| \mathbf{1}_N \boldsymbol{\Lambda}^T + \boldsymbol{\Lambda} \mathbf{1}_N{}^T - 2\boldsymbol{X}\boldsymbol{X}^T - \boldsymbol{U}_c \boldsymbol{S}_c \boldsymbol{V}_c^T \right\|_{\mathrm{F}}^2 . \tag{21}$$

To minimize the cost function, we start from the vertex of the $\mathbb{EDM}^N$ thus assume that all microphones are located in the origin of the space $\mathbb{R}^\zeta$. Denoting the location of microphone i with $\boldsymbol{x}_i = [x_{i1}, x_{i2}, ...x_{i\zeta}]^T$, $\mathcal{H}(\boldsymbol{X})$ is a polynomial function of x_{i1} of degree 4. The minimum of $\mathcal{H}(\boldsymbol{X})$ with respect to x_{i1} can be computed by equating the partial derivation of equation (21) to zero to obtain the new estimates, thus

$$\begin{gathered} \hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \mathcal{H}(\boldsymbol{X}) \\ (\boldsymbol{U}^{k+1}, \boldsymbol{V}^{k+1}, \boldsymbol{S}^{k+1}) = \text{SVD } (\mathbf{1}_N \hat{\boldsymbol{\Lambda}}^T + \hat{\boldsymbol{\Lambda}} \mathbf{1}_N{}^T - 2\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^T) \end{gathered} \tag{22}$$

where $\hat{\boldsymbol{\Lambda}} = (\hat{\boldsymbol{X}} \circ \hat{\boldsymbol{X}})\mathbf{1}_\zeta$. The stopping criteria is satisfied when the new estimates differ from the old ones by less than a threshold.

The modified iterations can be summarized in two steps:

- iteration $k + 1/2$:

$$\begin{aligned}
\boldsymbol{U}^{k+1/2} &= \boldsymbol{U}^k + \vartheta \frac{\partial F(\boldsymbol{U}^k, \boldsymbol{V}^k)}{\partial \boldsymbol{U}} \\
\boldsymbol{V}^{k+1/2} &= \boldsymbol{V}^k + \vartheta \frac{\partial F(\boldsymbol{U}^k, \boldsymbol{V}^k)}{\partial \boldsymbol{V}} \\
\boldsymbol{S}^{k+1/2} &= \arg\min_{\boldsymbol{S}} \mathcal{F}(\boldsymbol{U}^k, \boldsymbol{V}^k, \boldsymbol{S})
\end{aligned}$$

- iteration $k + 1$:

$$(\boldsymbol{U}^{k+1}, \boldsymbol{V}^{k+1}, \boldsymbol{S}^{k+1}) = \mathcal{P}_e(\boldsymbol{U}^{k+1/2}, \boldsymbol{V}^{k+1/2}, \boldsymbol{S}^{k+1/2}) \tag{23}$$

where ϑ is the step-size found using line search.

Once the distance matrix is recovered by either classic or Cadzow matrix completion algorithms, MDS is used to find the coordinates of the microphones, $\hat{\boldsymbol{X}}$, whereas the proposed Euclidean distance matrix completion algorithm directly yields the coordinates.

5. Theoretical Guarantees for Microphone Calibration

In this section, we derive the error bounds on the reconstruction of the positions of N microphones distributed randomly on a circular table of radius a using the matrix completion algorithm and considering the locality constraint on the known entries, i.e. $d_{ij} \leq d_{max}$, as well as the noise model with the standard deviation ς as stated in (11). Based on the following theorem we guarantee that there is an upper bound on the calibration error which decreases by the number of microphones.

Theorem 1. There exist constants C_1 and $C_2(\varsigma)$, such that the output $\hat{\boldsymbol{X}}$ satisfies

$$\operatorname{dist}(\boldsymbol{X}, \hat{\boldsymbol{X}}) \leq C_1 \frac{a^2 \log_2 N}{pN} + C_2(\varsigma) \frac{d_{max}^4}{a^2} \tag{24}$$

with probability greater than $1 - N^{-3}$, provided that the right-hand side is less than $\sigma_\eta(\boldsymbol{M})/N$.

5.1. Proof of Theorem 1

The squared distance matrix $\boldsymbol{M} \in \mathbb{R}^{N\times N}$ with rank$-\eta$, singular values $\sigma_k(\boldsymbol{M})$, $k \in [\eta]$ and singular value decomposition $\boldsymbol{U\Sigma U}^T$ is (μ_1, μ_2)-*incoherent* if the following conditions hold.

$\mathcal{A}_1$. For all $i \in [N]$: $\sum_{k=1}^{\eta} \boldsymbol{U}_{ik}^2 \le \eta\,\mu_1$.

$\mathcal{A}_2$. For all $i, j \in [N]$: $\left|\sum_{k=1}^{\eta} \boldsymbol{U}_{ik}(\sigma_k(\boldsymbol{M})/\sigma_1(\boldsymbol{M}))\boldsymbol{U}_{jk}\right| \le \sqrt{\eta}\,\mu_2$.

where without loss of generality, $\boldsymbol{U}^T\boldsymbol{U} = N\mathbb{I}$.

For a (μ_1, μ_2)-incoherent matrix $\boldsymbol{M}$, (25) is correct with probability greater than $1 - N^{-3}$; cf. [14]-Theorem 1.2.

$$\frac{1}{N}\|\boldsymbol{M} - \hat{\boldsymbol{M}}\|_{\mathrm{F}} \le \frac{C_1' \,\|\Psi_E(\boldsymbol{M}^s)\|_2 + C_2' \,\left\|\Psi_E(\boldsymbol{Z}^{\bar{s}})\right\|_2}{p\,N} \ , \tag{25}$$

provided that

$$|E| \ge C_1' N \kappa_\eta^2(\boldsymbol{M}) \max\left\{\mu_1\eta\log N \,;\, \mu_1^2\eta^2\kappa_\eta^4(\boldsymbol{M}) \,;\, \mu_2^2\eta^2\kappa_\eta^4(\boldsymbol{M})\right\} \ , \tag{26}$$

and

$$\frac{C_1' \,\|\Psi_E(\boldsymbol{M}^s)\|_2 + C_2' \,\left\|\Psi_E(\boldsymbol{Z}^{\bar{s}})\right\|_2}{p\,N} \le \sigma_\eta(\boldsymbol{M})/N \ , \tag{27}$$

where the condition number $\kappa_\eta(\boldsymbol{M}) = \sigma_1(\boldsymbol{M})/\sigma_\eta(\boldsymbol{M})$.

To prove Theorem 1, in the first step, we show the correctness of the upper bound stated in (24) based on the following Theorems 2 and 3. In the second step, conditions (26) and (27) are shown to hold along with the (μ_1, μ_2)-incoherence property.

Theorem 2. There exists a constant C_1'', such that with probability greater than $1 - N^{-3}$,

$$\|\Psi_E(\boldsymbol{M}^s)\|_2 \le C_1''\, a^2\, \log_2 N \ . \tag{28}$$

The proof of this theorem is explained in Appendix .1.

Theorem 3. There exists a constant $C_2''(\varsigma)$, such that with probability greater than $1 - N^{-3}$,

$$\left\|\Psi_E(\boldsymbol{Z}^{\bar{s}})\right\| \leq C_2''(\varsigma)\frac{d_{\max}^4}{a^2} p\, N \ . \tag{29}$$

The proof of this theorem is explained in Appendix .2.

On the other hand, the following condition holds for any arbitrary network of microphones [23]

$$\mathrm{dist}(\boldsymbol{X}, \hat{\boldsymbol{X}}) \leq \frac{1}{N}\|\boldsymbol{M} - \hat{\boldsymbol{M}}\|_{\mathrm{F}} \ . \tag{30}$$

Therefore, based on Theorem 2, Theorem 3 and the relations (25) and (30), the upper bound stated in (24) is correct where $C_1 = C_1'C_1''$ and $C_2(\varsigma) = C_2'C_2''(\varsigma)$; it is enough to investigate conditions (26) and (27) and (μ_1, μ_2)-incoherency of $\boldsymbol{M}$ to prove Theorem 1.

To show the inequality stated in (26), we can equivalently show that

$$Np \geq C_1'\mu^2\kappa_\eta^6(\boldsymbol{M}) \log N \ , \tag{31}$$

where $\mu = \max(\mu_1, \mu_2)$. In order to show that (31) holds with high probability for $N \geq C \log N/p$ and some constant C, we show that $\kappa_\eta(\boldsymbol{M})$ and μ are bounded with high probability independent of N.

The squared distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j \in \mathbb{R}^\zeta$ is given by

$$\boldsymbol{M}_{ij} = \rho_i^2 + \rho_j^2 - 2\boldsymbol{x}_i^T\boldsymbol{x}_j \ , \tag{32}$$

where ρ_i is the distance of microphone i from the center of the table. The squared distance matrix can be expressed as

$$\boldsymbol{M} = \boldsymbol{A}\mathcal{S}\boldsymbol{A}^T \ , \tag{33}$$

where for a planar deployment of microphones, i.e., $\zeta = 2$, $\eta = 4$, and $\boldsymbol{x}_i^T = [x_i, y_i] \in \mathbb{R}^2$, we have

$$\boldsymbol{A} = \begin{bmatrix} a/2 & x_1 & y_1 & -a^2/4 + \rho_1^2 \\ \vdots & \vdots & \vdots & \vdots \\ a/2 & x_N & y_N & -a^2/4 + \rho_N^2 \end{bmatrix},$$

and

$$\mathcal{S} = \begin{bmatrix} 2 & 0 & 0 & 2/a \\ 0 & -2 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 2/a & 0 & 0 & 0 \end{bmatrix}.$$

Since $\mathcal{S}$ is nondefective, using eigendecomposition, there is a non-singular matrix $\mathcal{W}$ and diagonal matrix $\boldsymbol{\Gamma}$ such that

$$\mathcal{S} = \mathcal{W}\boldsymbol{\Gamma}\mathcal{W}^{-1}\,, \tag{34}$$

where

$$\boldsymbol{\Gamma} = \text{diag}\left(-2, -2, \frac{a + \sqrt{4 + a^2}}{a}, \frac{a - \sqrt{4 + a^2}}{a}\right).$$

The largest and smallest singular values of $\mathcal{S}$ are $\sigma_1(\mathcal{S}) = \frac{a+\sqrt{4+a^2}}{a}$ and $\sigma_4(\mathcal{S}) = \min\left(2, \frac{\sqrt{4+a^2}-a}{a}\right)$ respectively. Based on (33), we have

$$\sigma_1(\boldsymbol{M}) \leq \sigma_1(\mathcal{S})\,\sigma_1(\boldsymbol{A}\boldsymbol{A}^T)\,, \tag{35}$$

$$\sigma_4(\boldsymbol{M}) \geq \sigma_4(\mathcal{S})\,\sigma_4(\boldsymbol{A}\boldsymbol{A}^T)\,. \tag{36}$$

Therefore, to bound $\kappa_4(\boldsymbol{M}) = \sigma_1(\boldsymbol{M})/\sigma_4(\boldsymbol{M})$, we need to derive the bound for $\sigma_1(\boldsymbol{A}\boldsymbol{A}^T)$ and $\sigma_4(\boldsymbol{A}\boldsymbol{A}^T)$. Assuming a uniform distribution of the microphones on the circular table, we have the following distribution for ρ

$$P_\rho(\rho) = \frac{2\rho}{a^2} \qquad \text{for} \quad 0 \leq \rho \leq a\,. \tag{37}$$

Therefore, the expectation of the matrix $\boldsymbol{A}^T\boldsymbol{A}$ is

$$\mathbb{E}[\boldsymbol{A}^T\boldsymbol{A}] = \begin{bmatrix} Na^2/4 & 0 & 0 & Na^3/4 \\ 0 & Na^4/4 & 0 & 0 \\ 0 & 0 & Na^4/4 & 0 \\ Na^3/8 & 0 & 0 & 7Na^4/48 \end{bmatrix} . \tag{38}$$

Hence, the largest and smallest singular values of $\mathbb{E}[\boldsymbol{A}^T\boldsymbol{A}]$ are $N\sigma_{\max}(a)$ and $N\sigma_{\min}(a)$ respectively with $\sigma_{\max}(a)$ and $\sigma_{\min}(a)$ independent of N. Moreover, $\sigma_i(\cdot)$ is a Lipschitz continuous function of its arguments and based on the Chernoff bound [24], we get

$$\mathbb{P}(\sigma_1(\boldsymbol{A}\boldsymbol{A}^T) > 2N\sigma_{\max}(a)) \leq e^{-C'N} , \tag{39}$$

$$\mathbb{P}(\sigma_1(\boldsymbol{A}\boldsymbol{A}^T) < (1/2)N\sigma_{\max}(a)) \leq e^{-C'N} , \tag{40}$$

$$\mathbb{P}(\sigma_4(\boldsymbol{A}\boldsymbol{A}^T) < (1/2)N\sigma_{\min}(a)) \leq e^{-C'N} , \tag{41}$$

for a constant C'. Hence, with high probability, based on relations (35), (36), (39) and (41), we have

$$\kappa_4(\boldsymbol{M}) \leq \frac{4\sigma_{\max}(a)\,\sigma_1(\mathcal{S})}{\sigma_{\min}(a)\,\sigma_4(\mathcal{S})} = f_{\kappa_4}(a) . \tag{42}$$

This bound is independent of N.

In the next step, we have to bound μ_1 and μ_2. The rank of matrix $\boldsymbol{A}$ is η, therefore there are matrices $\boldsymbol{B} \in \mathbb{R}^{\eta\times\eta}$ and $\boldsymbol{V} \in \mathbb{R}^{N\times\eta}$ such that $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{B}^T$ and $\boldsymbol{V}^T\boldsymbol{V} = N\mathbb{I}$. Given $\boldsymbol{M} = \boldsymbol{U}\Sigma\boldsymbol{U}^T$ and (33), we have $\Sigma = \boldsymbol{Q}^T\boldsymbol{B}^T\mathcal{S}\boldsymbol{B}\boldsymbol{Q}$ and $\boldsymbol{U} = \boldsymbol{V}\boldsymbol{Q}$ for an orthogonal matrix $\boldsymbol{Q}$. To show the incoherence property $\mathcal{A}_1$, we show that

$$||\boldsymbol{V}_{i.}||^2 \leq \eta\,\mu_1 \qquad \forall\, i \in [N] , \tag{43}$$

where $\boldsymbol{V}_{i.}$ denotes the transpose of i^{th} row of the corresponding matrix. For $\eta = 4$, since $\boldsymbol{V}_{i.} = \boldsymbol{B}^{-1}\boldsymbol{A}_{i.}$, we have $\|\boldsymbol{V}_{i.}\|^2 \leq \sigma_4(\boldsymbol{B})^{-2}\|\boldsymbol{A}_{i.}\|^2$ and $\sigma_4(\boldsymbol{A}) = \sqrt{N}\,\sigma_4(\boldsymbol{B})$, therefore

$$\|\boldsymbol{V}_{i.}\|^2 \leq \sigma_4(\boldsymbol{A})^{-2}\|\boldsymbol{A}_{i.}\|^2\, N\ . \tag{44}$$

Moreover, $\|\boldsymbol{A}_{i.}\|^2 = a^2/4 + \rho_i^2 + (-a^2/4 + \rho_i^2)^2 \leq 5a^2/4 + 9a^4/16$. Defining

$$f_{\mu_1}(a) = \frac{5a^2/2 + 9a^4/8}{\sigma_{\min}(a)}\ , \tag{45}$$

and based on (41) and (44), with high probability we have

$$\|\boldsymbol{U}_{i.}\|^2 \leq f_{\mu_1}(a) \qquad\qquad \forall\ i \in [N]\ . \tag{46}$$

Therefore, the incoherence property $\mathcal{A}_1$ for $\mu_1 = f_{\mu_1}(a)/\eta$ is correct; that is independent of N.

To prove the incoherence property $\mathcal{A}_2$, it is enough to prove that $\left|\boldsymbol{M}_{ij}/\sigma_1(\boldsymbol{M})\right| \leq \sqrt{\eta}\ \mu_2/N$ for all $i, j\ \in [N]$. The maximum value of $\boldsymbol{M}_{ij}$ is $4a^2$ and based on (36) and (41) we have

$$\sigma_1(\boldsymbol{M}) \geq \frac{1}{2}N\,\sigma_{\min}(a)\,\sigma_4(\mathcal{S})\ , \tag{47}$$

Defining $f_{\mu_2}(a) = 8a^2/\sigma_{\min}(a)\,\sigma_4(\mathcal{S})$, we have

$$\left|\boldsymbol{M}_{ij}/\sigma_1(\boldsymbol{M})\right| \leq \frac{f_{\mu_2}(a)}{N} \qquad\qquad \forall\, i, j\ \in [N]\ . \tag{48}$$

Therefore, the incoherence property $\mathcal{A}_2$ for $\mu_2 = f_{\mu_2}(a)/\sqrt{\eta}$ is correct; that is independent of N. Since $\kappa_4(\boldsymbol{M})$, μ_1 and μ_2 are bounded independent of N, matrix $\boldsymbol{M}$ is (μ_1, μ_2)-incoherent and the inequalities (26) and (31) are correct.

Further, (27) holds with high probability, if the right-hand side of (24) is less than $C_3\,\sigma_{\min}(a)\,\sigma_4(\mathcal{S})$, since based on (41), $\frac{\sigma_\eta(\boldsymbol{M})}{N} \geq \frac{1}{2}\sigma_{\min}(a)\sigma_4(\mathcal{S})$. This finishes the proof of Theorem 1.

■

The theoretical error bounds of ad hoc microphone calibration established above corresponds to the classic matrix completion algorithm. We will extend the mathematical results to the completion of Euclidean distance matrices incorporating the Cadzow and EDM projections through the experiments. As we will see in Section 7, this bound is not tight for the Cadzow projection and the Euclidean distance matrix completion algorithm as we achieve better results than matrix completion for microphone array calibration.

6. Related Methods

The objective is to extract the position of N microphones denoted as $\boldsymbol{x}_i$, $i \in \{1, \ldots, N\}$ up to a rigid transformation. Some of the state-of-the-art methods to achieve this goal are (1) Multi-Dimensional Scaling (MDS) [25], (2) Semi-Definite Programming (SDP) [26] and S-Stress (SS) [19] discussed briefly in the following sections. We refer the reader to the references for further details.

6.1. Classic Multi-Dimensional Scaling Algorithm

MDS refers to a set of statistical techniques used in finding the configuration of objects in a low dimensional space such that the measured pairwise distances are preserved [27]. Given a distance matrix, finding the relative microphone positions is achieved by MDSLocalize [19]. In the ideal case where matrix $\boldsymbol{M}$ is complete and noiseless, this algorithm outputs the relative positions of the microphones in the desired dimension through the steps summarized as follows

- ⋄ Double centering $\boldsymbol{M}$ via $\Xi(\boldsymbol{M}) = \frac{-1}{2} \boldsymbol{J} \boldsymbol{M} \boldsymbol{J}$.
- ⋄ Eigenvalue decomposition of $\Xi(\boldsymbol{M})$ as $\boldsymbol{U} \boldsymbol{\Pi} \boldsymbol{U}^T$.
- ⋄ Extracting the ζ largest eigenvalues and the corresponding eigenvectors denoted by $\boldsymbol{\Pi}_+$ and $\boldsymbol{U}_+$.
- ⋄ The microphone positions are obtained as $\boldsymbol{X} = \boldsymbol{U}_+ \boldsymbol{\Pi}_+$.

In a real scenario of missing distances, a modification called MDS-MAP [25] computes the shortest paths between all pairs of nodes in the region of consideration. The shortest path between microphones i and j is defined as the path between two nodes such that the sum of the estimated distance measures of its constituent edges is minimized. By approximating the missing distances with the shortest path and constructing the distance matrix, classical MDS is applied to estimate the microphone array geometry.

6.2. *Semidefinite Programming*

Another efficient method that can be used for calibration is the semidefinite programming approach formulated as

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \sum_{(i,j)\in E} w_{ij} \left| \left\| \boldsymbol{x}_i - \boldsymbol{x}_j \right\|^2 - \tilde{d}_{ij}^2 \right| , \tag{49}$$

where w_{ij} shows the reliability measure on the estimated pairwise distances. The basis vectors in Euclidean space $\mathbb{R}^N$ are denoted by $\{u_1, u_2, \cdots, u_N\}$. The optimization expressed in equation (49) is not convex but can be relaxed as a convex minimization via

$$\begin{aligned} &\min_{\boldsymbol{X},\boldsymbol{Y}} \sum_{(i,j)\in E} w_{ij} \left| (u_i - u_j)^T [\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{X}^T, \mathbb{I}_\zeta](u_i - u_j)^T - \tilde{d}_{ij}^2 \right| \\ &\text{subject to} \quad [\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{X}^T, \mathbb{I}_\zeta] \succeq 0, \quad \left\| \boldsymbol{X}^T \mathbf{1}_N \right\| = 0 \end{aligned} \tag{50}$$

where $\boldsymbol{Y}_{N\times N}$ is a positive semidefinite matrix and $\succeq$ is a generalized matrix inequality on the positive semidefinite cone [28]. To further increase the accuracy, a gradient decent is applied on the output of SDP minimization [26].

6.3. *Algebraic S-Stress Method*

The s-stress method for calibration extracts the topology of the ad hoc network by optimizing the cost function stated as

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \sum_{(i,j)\in E} w_{ij} \left(\left\| \boldsymbol{x}_i - \boldsymbol{x}_j \right\|^2 - \tilde{d}_{ij}^2 \right)^2 . \tag{51}$$

The reliability measure w_{ij} controls the least square regression stated in equation (51) which can be set according to the measure of $\tilde{d}_{ij}$. If $w_{ij} = \tilde{d}_{ij}^{-2}$, we have *elastic scaling* that gives importance to large and small distances. If $w_{ij} = 1$, large distances are given more importance than the small distances. In general, incorporation of $w_{ij} = \tilde{d}_{ij}^{\alpha}$, $\alpha \in \{..., -2, -1, 0, 1, 2, ...\}$ yields different loss functions and depending on the structure of the problem, one of them may work better than the other [29].

7. Experimental Analysis

7.1. A-priori Expectations

The simplest method that we discussed is the classical MDS algorithm. This method assumes that all the pairwise distances are known and in the case of missing entries and noise, it does not minimize a meaningful utility function. An extension of this method is MDS-MAP which replaces the missing distances with the shortest path. In many scenarios, this is considered as a coarse approximation of the true distances.

The SDP-based method on the other hand performs fairly well with missing distance information. Together with its final gradient descent phase, it can find good estimates of the location. However, since each distance information translates into a constraint in the semi-definite program, this approach is not scalable and becomes intractable for large sensor networks.

The alternative approach is to minimize the non-convex s-stress function. Although performing well in many cases, in the case of missing distances, one cannot eliminate the possibility of falling into local minima using this approach.

The approach that we proposed in this paper exploits matrix completion algorithm to recover the missing distances considering the low-rank as well as Euclidean properties of the distance matrix. The classic matrix completion does not take into account the EDM properties. By integrating the Cadzow projection, the estimated matrix have partial EDM properties, and hence we expect better reconstruction results. Further, by incorporating the full EDM structure, we achieve a Euclidean distance matrix completion algorithm and expect more fidelity in the reconstruction results.

In this section, we present the evaluation results of the microphone calibration methods using real data recordings collected at Idiap's instrumented meeting room.

7.2. *Recording Set-up*

We consider a scenario in which eleven microphones are located on a planar area: Eight of them are located on a circle with diameter 20cm and one microphone is at the center. There are two additional microphones with 70cm distance from the central microphone. The microphones are Sennheiser MKE-2-5-C omnidirectional miniature lapel microphones.

The floor of the room is covered with carpet and surrounded with plaster walls having two big windows. The enclosure is a $8 \times 5.5 \times 3.5\,\mathrm{m}^3$ rectangular room and it is moderately reverberant. It contains a centrally located $4.8 \times 1.2\,\mathrm{m}^2$ rectangular table. This scenario mimics the MONC database [30]. The sampling rate is 48k while the processing applied for microphone calibration is based on down-sampled signal of rate 16k to reduce the computational cost of pairwise distance estimation.

7.3. *Pairwise Distance Estimation*

In order to estimate the pairwise distances, we take two microphone signals of length 2.14 s and frame them into short windows of length 1024 samples using a Tukey function (parameter = 0.25) and apply Fourier transform. For each frame, we compute the coherence function. The average of the coherence functions over 1000 frames are computed and used for estimation of the pairwise distance by fitting a sinc function as stated in (1) using the algorithm described in [16]. This algorithm is an improved version of the distance estimation using diffuse noise coherence model which enables a reasonable estimate up to 73 cm. We empirically confirm that the distances beyond that are not reliably estimated so they are regarded as *missing*. Thereby, the following entries of the Euclidean distance matrix are missing, $d_{10,11}, d_{1,10}, d_{7,10}, d_{8,10}, d_{5,11}, d_{6,11}$, $d_{7,11}$ (see Figure 5).

7.4. *Geometry Estimation*

In the scenario described above, microphone calibration is achieved in two steps. First, all methods are used to find the nine close microphones in order to evaluate them for geometry estimation when we have all distances. The geometry of these microphones are fixed and used to calibrate the rest of the network. Figure 3 demonstrates the results of MDS-MAP, SDP, s-stress and the proposed Euclidean distance matrix completion algorithm. The calibration error is quantified based on (12). The best results are achieved by the proposed algorithm with error 5.85. The second place belongs to s-stress with error 6.1 followed by MDS-MAP and SDP with errors 8.13 and 8.63 respectively. Figure 4 provides a comparative illustration of the results of

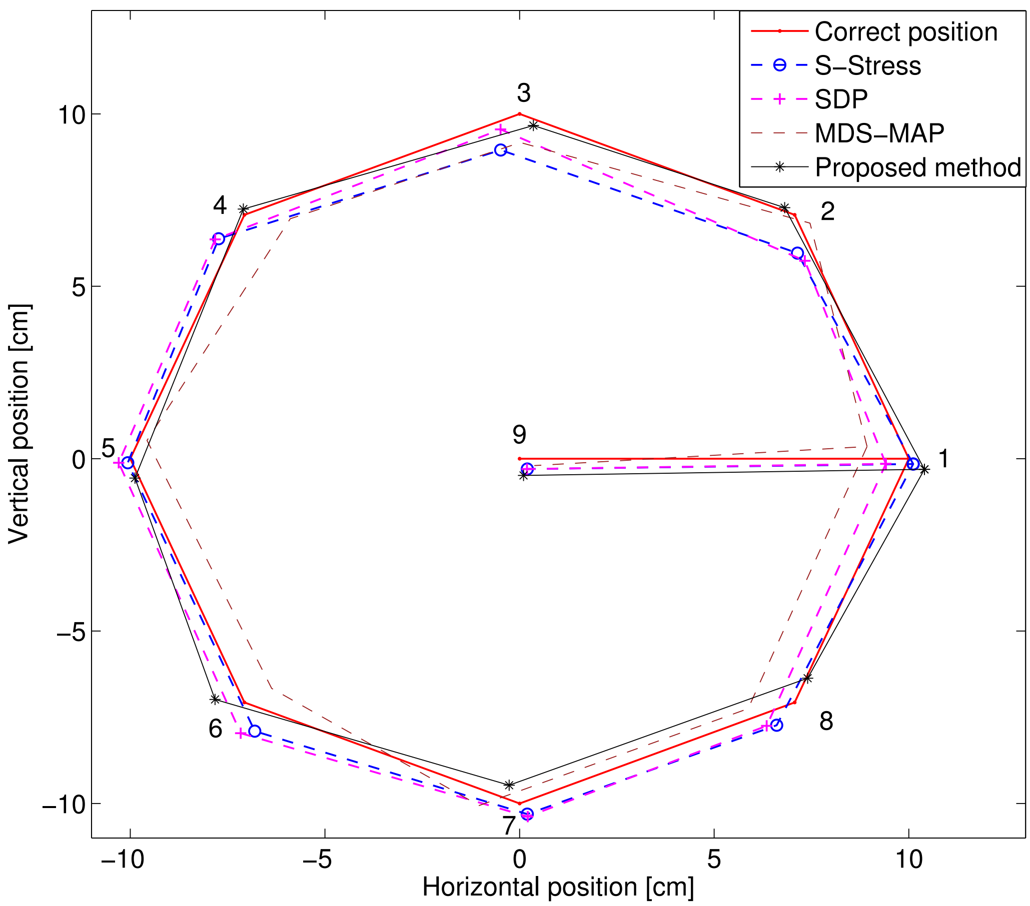

Figure 3: Calibration of the nine-element microphone array. The geometries are estimated using MDS, S-stress, SDP and the proposed method.

matrix completion (MC), MC+Cadzow and the proposed Euclidean distance matrix completion algorithm. We can see that MC+Cadzow yields better result with error 7.68 compared to MDS-MAP, but worse than s-stress. Table 2 summarizes all the results.

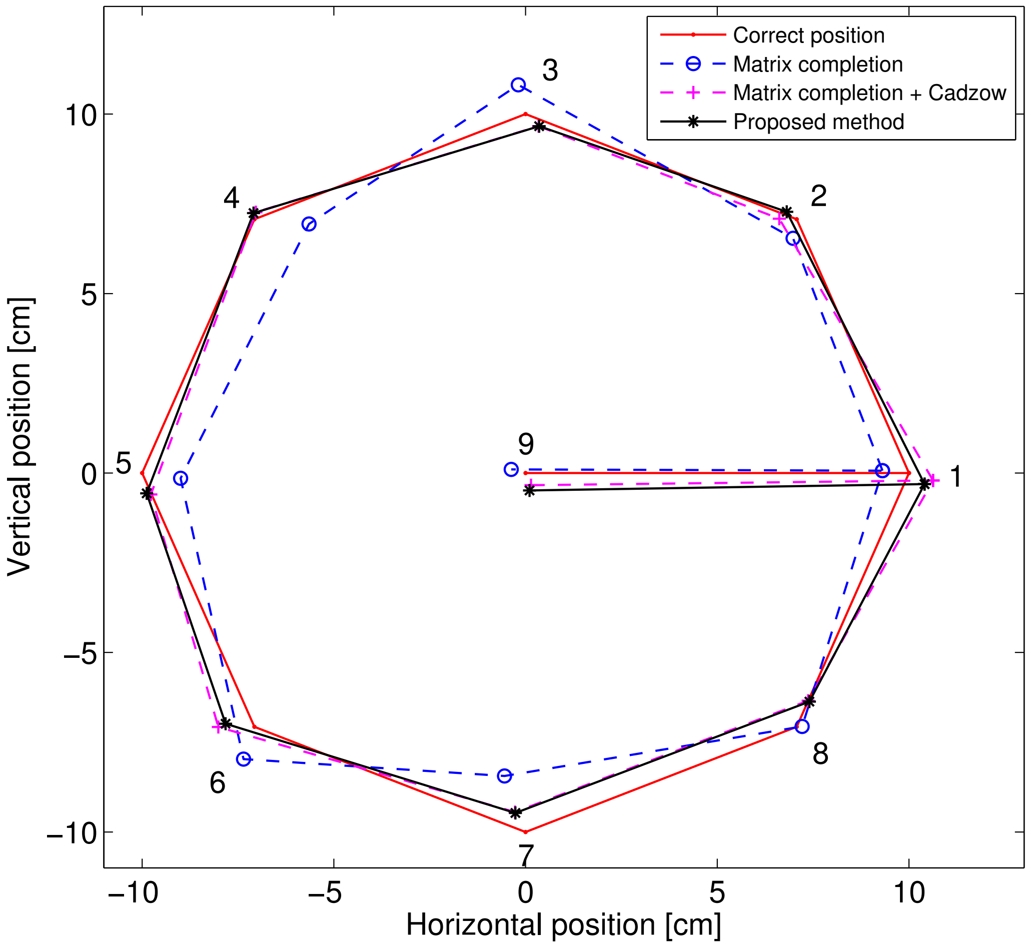

Figure 4: Calibration of the nine-element microphone array. The geometries are estimated using MC, MC+Cadzow and the proposed method.

Methods	Errors
MDS	8.13
SDP	8.63
S-stress	6.1
Matrix completion	9.75
Matrix completion+Cadzow	7.68
Proposed method	5.85

Table 2: Performance comparison of different methods for calibration of the nine-element microphone array. The error is quantified based on (12).

The scenario using eleven channels of microphones addresses the problem of having partial distance estimation for calibration of an ad hoc microphone array. The experiments show that the proposed method offers the best estimation of the geometry as illustrated in Figure 5 and 6 with error 49.6. As we can see, the proposed Euclidean distance matrix completion algorithm achieves more than twice less error than the best state-of-the-art alternative.

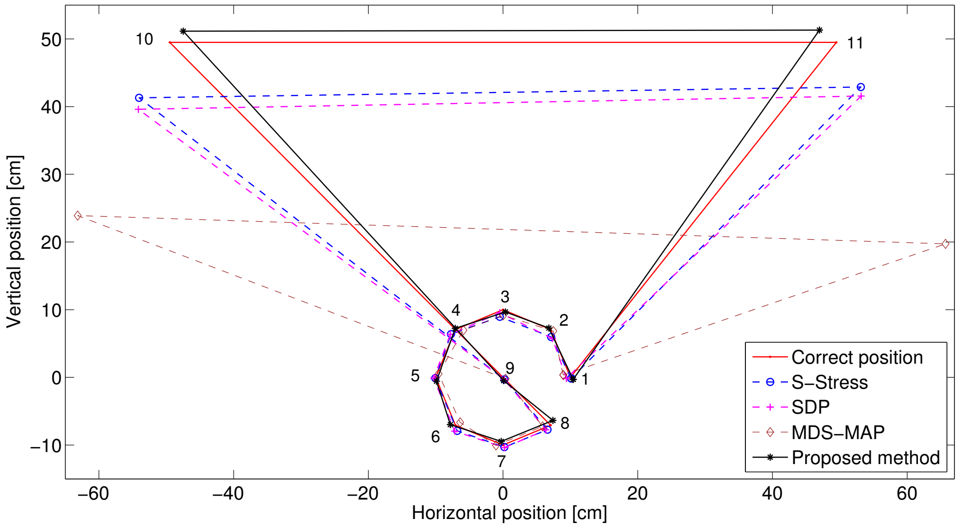

Figure 5: Calibration of the eleven-element microphone array while several pairwise distances are missing. The geometries are estimated using MDS-MAP, SDP, S-stress and the proposed method.

The worst result belongs to MDS-MAP with error 438 because the shortest path is a poor estimation of missing entries. The s-stress and SDP search the Euclidean space corresponding to the feasible positions hence, their performance are more reasonable with errors 141 and 125. The advantage of being constrained to a physically possible search space or close to it is considered in extensions of matrix completion in MC+Cadzow and the proposed method and achieves the best performance. The results are summarized in Table 3. These experimental evaluations confirm the effectiveness of the proposed algorithm and demonstrate the hypothesis that incorporating the EDM properties in matrix completion algorithm enables calibration of ad hoc microphone arrays from partial measurements of the pairwise distances.

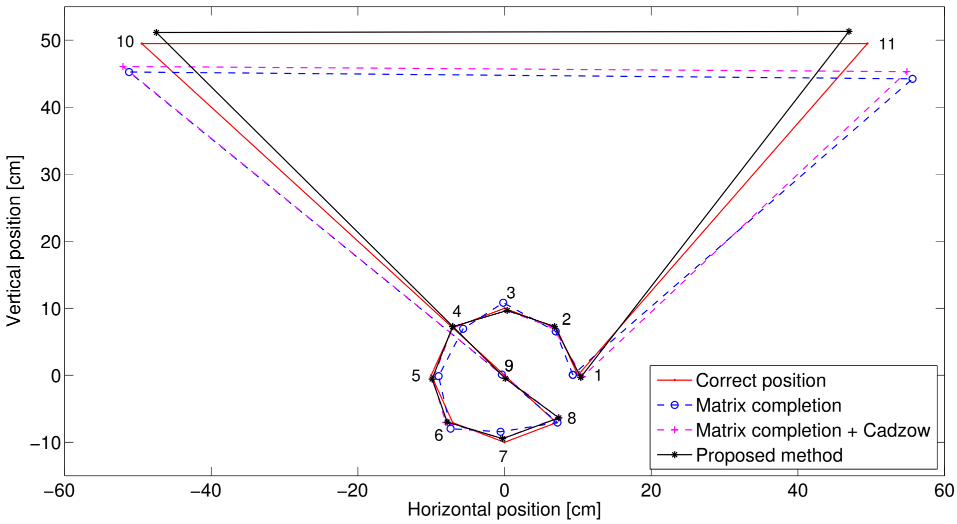

Figure 6: Calibration of the eleven-element microphone array while several pairwise distances are missing. The geometries are estimated using MC, MC+Cadzow, and the proposed method.

Methods	Errors
MDS-MAP	438
SDP	141
S-stress	125
Matrix completion	133
Matrix completion+Cadzow	119
Proposed method	49.6

Table 3: Performance comparison of different methods for calibration of the eleven-element microphone array while several pairwise distances are missing. The error is quantified based on (12).

8. Conclusions

We proposed a Euclidean distance matrix completion algorithm for calibration of ad hoc microphone arrays from partially known pairwise distances. This approach exploits the low-rank property of the distance matrix and recovers the missing entries based on a matrix completion optimization scheme. To incorporate for the properties of a Euclidean distance matrix, the estimated matrix at each iteration of the matrix completion is projected onto the EDM cone. Furthermore, we derived the theoretical bounds on the calibration error using matrix completion algorithm. The experimental evaluations conducted on real data recordings demonstrate that the proposed method outperforms the state-of-the-art techniques for ad hoc array calibration in particular in the scenarios of missing distances. This study confirmed that exploiting the combination of the rank condition of EDMs, similarity in the measured distances, and iterative projection on the EDM cone leads to the best position reconstruction results. The proposed algorithm and the theoretical guarantees are applicable to the general framework of ad hoc sensor networks calibration.

Appendix .1. Proof of Theorem 2

The goal is to find the bound of the norm of the squared distance matrix with missing entries according to structures indicated by E and S. Based on (6) and (9), we define matrix $\mathcal{E}$ as

$$\mathcal{E}_{ij} = \begin{cases} 1 & \text{if } (i, j) \in E \cap S \\ 0 & \text{otherwise} \end{cases}$$

Both E and S are symmetric matrices, hence $\mathcal{E}$ is also symmetric. Due to the physical setup, we know that $\Psi_E(\boldsymbol{M})_{ij} \leq 4a^2$ for all $i, j \in [N]$ and from the norm definition we have

$$\|\Psi_E(\boldsymbol{M}^s)\|_2 \quad \leq \quad 4a^2 \max_{\|\mathbf{h}\|=\|\boldsymbol{\hbar}\|=1} \sum_{i,j} |h_i|\,|\hbar_j|\,\mathcal{E}_{ij} = 4a^2\|\mathcal{E}\|_2 \ ,$$

where $\mathbf{h} = [h_1, h_2, ..., h_N]^T$ and $\boldsymbol{\hbar} = [\hbar_1, \hbar_2, ..., \hbar_N]^T$ are right and left eigenvectors of matrix $\mathcal{E}$. In order to bound $\|\mathcal{E}\|_2$, we first define a binomial random variable vector $\boldsymbol{\nu} = [\nu_1, \nu_2, ..., \nu_N]^T$ where

$$\nu_i = \sum_{j\in[N]} |\mathcal{E}_{ij}| \ . \tag{.1}$$

Based on the Gershgorin circle theorem we have $\|\mathcal{E}\|_2 \leq \|\boldsymbol{\nu}\|_\infty$. Each entry in matrix $\mathcal{E}$ is one with probability $p\,q$ where q is the probability that the entry is included in structured missing entries or

$$q = \mathbb{P}\{|\boldsymbol{x}_i - \boldsymbol{x}_j| \geq d_{\max}\} \ .$$

Hence, we have

$$\mathbb{E}[\nu_i] = Npq \ , \tag{.2}$$

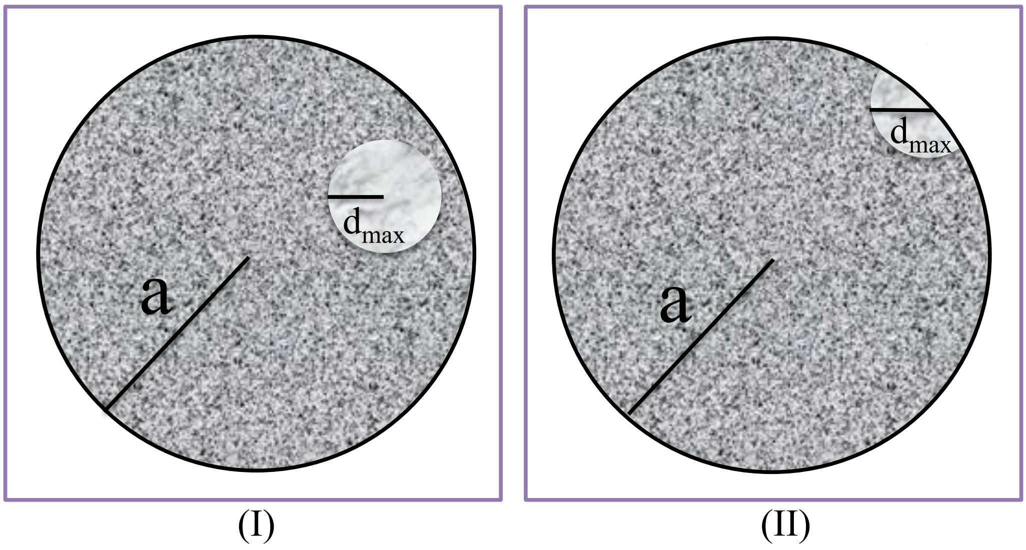

Figure .7: Scenario corresponding to the (I) lower bound and (II) upper bound of the probability q of structured missing distances.

For bounding $\mathbb{E}[v_i]$, it is necessary to bound q. Figure .7.I depicts the lowest probability of missing distances if the microphone location with respect to the edge of the circular table has a distance more than $d_{\max}$ and Figure .7.II depicts the highest probability if the microphone is located right at the edge of the table. The maximum of $d_{\max}$ is a. We denote the upper bound and lower bound with $q_{\max}(a, d_{\max})$ and $q_{\min}(a, d_{\max})$ respectively, therefore

$$q_{\min}(a, d_{\max}) \leq q \leq q_{\max}(a, d_{\max}) \ . \tag{.3}$$

As illustrated in Figure .7. $q_{\min}(a, d_{\max}) = \max\{1 - (\frac{d_{\max}}{a})^2, 0\}$ and $q_{\max}(a, d_{\max}) = 1 - \frac{B}{\pi a^2}$ where B is the intersection area between the two circles. By computing B, we obtain

$$q_{\max} = 1 - \frac{2\gamma}{\pi} + \frac{1}{2\pi} \sin 4\gamma + \frac{2\xi^2}{\pi}[2\gamma + \sin 2\gamma] - 2\xi^2 \ , \tag{.4}$$

where $\xi = d_{\max}/2a$ and $\gamma = \sin^{-1}\xi$. Based on (.2) and (.3) we have

$$Npq_{\min}(a, d_{\max}) \leq \mathbb{E}[\nu_i] \leq Npq_{\max}(a, d_{\max}) \; . \tag{.5}$$

By applying the Chernoff bound to ν_i we have

$$\mathbb{P}(\nu_i > (1+\epsilon)\mathbb{E}[\nu_i]) \leq 2^{-(1+\epsilon)\mathbb{E}[\nu_i]} \; ,$$

where ϵ is an arbitrary positive constant. Therefore, based on (.5) we have

$$\mathbb{P}(\nu_i > (1+\epsilon)Np\, q_{\max}) \leq 2^{-(1+\epsilon)Np\, q_{\min}} \; . \tag{.6}$$

By applying the union bound we have

$$\mathbb{P}(\max_{i\in[N]} \nu_i > (1+\epsilon)Np\, q_{\max}) \leq 2^{-(1+\epsilon)Np\, q_{\min}+\log_2 N} \; . \tag{.7}$$

We assume that $q_{\min}$ and $q_{\max}$ grow as $O(\frac{\log_2 N}{N})$; this assumption indicates that the ratio of the structured missing entries with respect to N decreases as N grows or in other words, $d_{\max}$ increases as the size of the network N grows. Therefore, we have

$$\mathbb{P}(\max_{i\in[N]} \nu_i > (1+\epsilon)Np\, q_{\max}) \leq N^{-\theta} \; , \tag{.8}$$

where the positive parameter $\theta = (1+\epsilon)p - 1$; by choosing $\epsilon \geq 4/p - 1$, with probability greater than $1 - N^{-3}$, we have

$$\|\Psi_E(\boldsymbol{M}^s)\|_2 \leq 4a^2 \max_{i\in[N]} \nu_i \; ,$$

and based on (.8)

$$\|\Psi_E(\boldsymbol{M}^s)\|_2 \leq 4a^2(1+\theta)q_{\max}N \; .$$

Therefore, we achieve

$$\|\Psi_E(\boldsymbol{M}^s)\|_2 \leq C_1'' a^2 \log_2 N \; .$$

■

Appendix .2. Proof of Theorem 3

Based on the noise model described in Section 3.3, the maximum entry of the matrix $\mathbf{Z}^{\bar{s}}_{ij}$ is obtained as

$$\max \mathbf{Z}^{\bar{s}}_{ij} = \max_{i,j} d^2_{ij} \Upsilon_{ij} \left(2 + \Upsilon_{ij}\right) \,, \tag{.9}$$

where $d_{ij} \leq d_{\max}$ and based on (11), the value of $|\Upsilon_{ij}|\left(2 + |\Upsilon_{ij}|\right)$ with probability greater than $1 - N^{-3}$ is less than $16\varsigma + 64\varsigma^2$ for a typical network of less than $N < 10^4$ microphones. Assuming a uniform distribution of the microphones, $\mathbf{Z}^{\bar{s}}$ has $N\,(d_{\max}/a)^2$ non zero entries. Based on the Gershgorin circle theorem [31] we have

$$\left\|\mathbf{Z}^{\bar{s}}\right\| \leq \max_i \sum_j |\mathbf{Z}^{\bar{s}}_{ij}| \,. \tag{.10}$$

By applying Ψ_E projection, with $C''_2(\varsigma) = 16\varsigma + 64\varsigma^2$ [2] we have

$$\left\|\Psi_E(\mathbf{Z}^{\bar{s}})\right\| \leq C''_2(\varsigma) \frac{d^4_{\max}}{a^2} p\,N \,.$$

■

Acknowledgment

This work was supported by the Swiss National Science Foundation under the National Center of Competence in Research (NCCR) on "Interactive Multi-modal Information Management" (IM2).

[2]The value of ς can be approximated with a constant for many types of rooms and estimated as less than 0.05 through simulations.

References

[1] A. Asaei, M. J. Taghizadeh, M. Bahrololum, M. Ghanbari, Verified speaker localization utilizing voicing level in split-bands, Signal Processing 89(6).

[2] A. Asaei, P. N. Garner, H. Bourlard, Sparse component analysis for speech recognition in mullti-speaker environment, in: Proceedings of INTERSPEECH, 2010.

[3] C.-T. Do, M. J. Taghizadeh, P. N. Garner, Combining cepstral normalization and cohlear implant-like speech processing for microphone array-based speech recognition, in: IEEE workshop on Spoken Language Technology (SLT), 2012.

[4] J. M. Sachar, H. F. Silverman, W. R. Patterson, Microphone position and gain calibration for a large-aperture microphone array, IEEE Transactions on Speech and Audio Processing 13(1).

[5] V. C. Raykar, I. V. Kozintsev, R. Lienhart, Position calibration of microphones and loudspeakers in distributed computing platforms, IEEE Transactions on Speech and Audio Processing 13(1).

[6] B. P. Flanagan, K. L. Bell, Array self-calibration with large sensor position errors, Signal Processing 81.

[7] M. Chen, Z. Liu, L. He, P. Chou, Z. Zhang, Energy-based position estimation of microphones and speakers for ad-hoc microphone arrays, in: IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 2007.

[8] I. McCowan, M. Lincoln, I. Himawan, Microphone array shape calibration in diffuse noise fields, IEEE Transactions on Audio,Speech and Language Processing 16(3).

[9] M. J. Taghizadeh, P. N. Garner, H. Bourlard, H. R. Abutalebi, A. Asaei, An integrated framework for multi-channel multi-source localization and voice activity detection, in: IEEE Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA), 2011.

[10] M. J. Taghizadeh, P. N. Garner, H. Bourlard, Microphone array beampattern characterization for hands-free speech applications, in: IEEE, 7th Sensor Array and Multichannel Signal Processing Workshop (SAM), 2012.

[11] M. J. Taghizadeh, P. N. Garner, H. Bourlard, Enhanced diffuse field model for ad hoc microphone array calibration, in: Elsevier, Signal Processing (in press), available online at `http://www.sciencedirect.com/science/article/pii/S0165168414000826`.

[12] E. J. Candes, Y. Plan, Matrix completion with noise, IEEE Signal Processing Magazine 98(6).

[13] R. H. Keshavan, S. Oh, OptSpace: A gradient descent algorithm on the grassman manifold for matrix completion`arXiv:0910.5260`.

[14] R. H. Keshavan, A. Montanari, S. Oh, Matrix completion from noisy entries, Journal of Machine Learning Research 11.

[15] P. Drineas, M. Javed, M. Magdon-Ismail, G. Pandurangant, R. Virrankoski, A. Savvides, Distance matrix reconstruction from incomplete distance information for sensor network localization, Sensor and Ad Hoc Communications and Networks 2.

[16] M. J. Taghizadeh, R. Parhizkar, P. N. Garner, H. Bourlard, Euclidean distance matrix completion for ad-hoc microphone array calibration, in: IEEE 18th International Conference in Digital Signal Processing, 2013.

[17] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, M. C. Thompson, Measurement of correlations coefficients in reverberant sound fields, Journal of the Acoustical Society of America 27.

[18] Y. Shang, W. Ruml, Y. Zhang, M. P. J. Fromherz, Localization from mere connectivity, in: MobiHoc, 2003, pp. 201–212.

[19] I. Borg, P. J. Groenen, Modern multidimensional scaling theory and applications, in: Springer, 2005.

[20] E. Candes, B. Recht, Exact matrix completion via convex optimization, Magazine Communications of the ACM 55.

[21] J. Cadzow, Signal enhancement-a composite property mapping algorithm, IEEE Transactions on Acoustics, Speech, and Signal Processing 36.

[22] J. Dattorro, Convex Optimization and Euclidean Distance Geometry, Meboo Publishing, USA, 2012.

[23] R. Parhizkar, Euclidean distance matrices: Properties, algorithms and applications, Ph.D. thesis, École cole polytechnique fédérale de Lausanne EPFL (2014).

[24] J. A. Tropp, User-friendly tail bounds for sums of random matrices, Found. Comput. Math. 12(3).

[25] Y. Shang, W. Ruml, Y. Zhang, M. P. J. Fromherz, Localization from connectivity in sensor networks, IEEE Transactions on Parallel Distribute Systems 15(11).

[26] P. Biswas, T. C. Liang, K. C. Toh, T. C. Wang, Y. Ye, Semidefinite programming approaches for sensor network localization with noisy distnce measurments, IEEE Transactions on Automation Science and Engineering 3.

[27] T. F. Cox, M. A. A. Cox, Multidimensional scaling, Chapman-Hall.

[28] S. Boyd, L.Vandenberghe, Convex optimization, in: Cambridge, 2012.

[29] A. Buja, D. F. Swayne, Visualization methodology for multidimensional scaling, Journal of Classification 19.

[30] D. Moore, I. McCowan, The multichannel overlapping numbers corpus, Idiap resources available online:`http://www.cslu.ogi.edu/corpora/monc.pdf`.

[31] D. S. Scott, On the accuracy of the gerschgorin circle theorem for bounding the spread of a real symmetric matrix, Lin. Algebra Appl. 65.