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], Afsaneh Asaei[a]

Emails: {mohammad.taghizadeh, phil.garner, herve.bourlard, afsaneh.asaei}@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 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. This study elucidates the links between the calibration error and the number of microphones along with the noise level and the ratio of missing distances. Thorough experiments on real data recordings and simulated setups are conducted to demonstrate these theoretical insights. A significant improvement is achieved by the proposed Euclidean distance matrix completion algorithm over the state-of-the-art techniques for ad hoc microphone array calibration.

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

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 assume that the recordings are synchronized and address the problem of finding the microphone positions; this problem is referred to as *microphone calibration*. The precise knowledge of the microphones positions is required for a plethora of multi-channel audio processing applications such as distant speech recognition [1, 2, 3, 4, 5] and source localization and separation [6, 7, 8, 9, 10].

Previous studies often consider activation of a (known) source signal in a specific configuration to estimate the distances between the source and microphones. The pairwise distances are then used to reconstruct the array geometry. This approach is referred to as self-calibration.

It may be noted that the knowledge of the source signal simplifies the estimation problems. If the signal is known beforehand, the time of arrival (ToA) of the source signal for each individual microphone is obtained through cross-correlation with the given signal. Hence, the negative effects of noise and reverberation are reduced as only one of the signals is noisy. On the other hand, if the source signal is unknown, the time difference of arrival (TDoA) for a pair of microphones is estimated through cross-correlation of the two microphone signals. Although TDoA-based methods can alleviate the need for activating a specific source signal or prior knowledge on the original signal, they may be more sensitive to noise and reverberation.

Sachar et al. [11] presented a set-up using a pulsed acoustic excitation generated by five domed tweeters for measuring the transmit times and distances between speakers and microphones. Raykar et al. [12] exploited a maximum length sequence or chirp signal in a distributed computing platform. The time difference of arrival of microphone signals were 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.

Gaubitch et al. [13] proposed an auto-localization method exploiting the asynchronous time-of-arrival measurements obtained from spatially distributed acoustic events. He developed an

iterative rank reduction algorithm to correct for the time offsets without imposing any geometrical constraint on the placement of the microphones and sound sources. These aligned time-of-arrival measurements are then used to estimate the location of source and microphones using the bilinear optimization approach proposed in [14]. This approach requires a minimum of five microphones and thirteen sound source events.

In an alternative approach to alleviate the requirement for a specific source signal, Chen et al. [15] formulated an energy-based method for maximum likelihood estimation of joint source-sensor positions. These methods require several active sources and the pairwise distances are used for a nonlinear optimization to extract the array geometry.

Pollefeys and Nistre proposed a method for direct joint source and microphone localization which requires matrix factorization and solving linear equations [16]. Along that line, Kuang et al. [17] exploited the rank constraints to determine the unknown time offset for time-of-arrival measurements. The problem is then reduced to solving a system of polynomials for extracting the location of source and microphones. The estimates are further refined using a non-linear least squares optimization to find the correct position and time delays matching the measurements of source-microphones distances. It has been shown that exploiting the structure underlying the problem through rank and polynomial constraints enables direct recovery of source and microphone positions using as few as three microphones and six sources, thus achieves a minimal case for the self-calibration problem.

Recently, McCowan and Lincoln [18] exploited properties of a diffuse noise field model for microphone calibration; this approach alleviates the need for activating several sources. 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 squares sense. Once the pairwise distances are estimated, the classic multi-dimensional scaling method is used to reconstruct the microphone array geometry [19]. Along similar lines, Hennecke et al. [20] proposed a hierarchical approach where the compact sub-arrays are calibrated using the coherence model of a diffuse sound field. A sound signal is activated and the

relative positions of the distributed arrays are determined using steered response power based source localization.

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 in reverberant enclosures [7] and no requirement for activating a specific source signal [1]. Estimation of the pairwise distances becomes unreliable as the distances between the microphones are increased. Hence, the goal of this paper is to enable microphone calibration when some of the pairwise distances are missing.

The problem of missing data arises when the pairwise distance of only a subset of the sensors can be measured. If a source event is activated, device malfunctioning or architectural barriers (e.g. indoor calibration) may cause the signal of the emitted sounds to reach, or be acquired, by only a subset of the sensors. Furthermore, some of sensors deployed far apart may fail to capture the source energy leading to a locality constraint in distance estimation in ad hoc microphone arrays [21]. In this paper, as an example use case, the local pairwise distances are measured based on the diffuse sound field coherence model. However, the proposed algorithm and theoretical results are applicable for calibration of a general ad hoc microphone array network. The approach proposed in this paper imposes no constraint on the geometrical set up.

To address the problem of missing distances, we rely on the characteristics of a Euclidean distance matrix. The matrix consisting 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. [22] 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 [23]. Furthermore, they proved that their algorithm is robust against noise [24]. Drineas et al. [25] 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.

[1] In fact, the ambient noise typical of many enclosures provides a sound field which enables microphone calibration and no additional source signal is required to be played.

In the present study, 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 [26]; 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 accurately 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 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 extraction of 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 analysis are presented in section 7. Finally, 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 with the relation expressed as

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

where ω is the frequency, d_{ij} is the distance between the two microphones i and j, and c is the speed of sound [27]. Figure 1 represents an example of the coherence and the fitted sinc function.

In practice, if the distance between the sensors is large (e.g. greater than 73 cm [26, 28]) 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 consisting of the pairwise distances. 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 the distance matrix; the random missing entries intend to model the distances which can not be measured due to mismatch or violations of the underlying pairwise distance estimation model pertained to the acoustic ambiguities. Furthermore, the known entries are noisy due to measurement inaccuracies and variations of diffuseness [28].

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 $\boldsymbol{\Lambda} = (\boldsymbol{X} \circ \boldsymbol{X})\mathbf{1}_\zeta$; $\circ$ 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 $\zeta + 2$ [25]. 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 algorithm in this paper.

3.2. Objective

The noisy estimates of the pairwise distances are 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 consists of the structured missing entries corresponding to the distances greater than d_{max}. We denote this group by S defined as

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

These structured missing entries are represented by a matrix

$$\boldsymbol{D}_{ij}^{s} = \begin{cases} \boldsymbol{D}_{ij} & \text{if } (i,j) \in S \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Hence, the noiseless recognized pairwise distance matrix is given by

$$\boldsymbol{D}^{\bar{s}} = \boldsymbol{D} - \boldsymbol{D}^{s} \,, \tag{8}$$

and we obtain the corresponding known squared distance matrix as

$$\begin{aligned} \boldsymbol{M}^{s} &= \boldsymbol{D}^{s} \circ \boldsymbol{D}^{s} \\ \boldsymbol{M}^{\bar{s}} &= \boldsymbol{D}^{\bar{s}} \circ \boldsymbol{D}^{\bar{s}} = \boldsymbol{M} - \boldsymbol{M}^{s} \,. \end{aligned} \tag{9}$$

Considering the noise on the known entries, we obtain

$$\widetilde{\boldsymbol{M}}^{\bar{s}} = \boldsymbol{M}^{\bar{s}} + \boldsymbol{Z}^{\bar{s}} \,, \tag{10}$$

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

To model 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

$$\Psi_{E}(\boldsymbol{Q})_{ij} = \begin{cases} \boldsymbol{Q}_{ij} & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

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

$$\boldsymbol{M}^{E} = \Psi_{E}(\widetilde{\boldsymbol{M}}^{\bar{s}}) \,. \tag{12}$$

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 become larger [28]. We model this effect through

$$\boldsymbol{W} = \boldsymbol{\Upsilon} \circ \boldsymbol{D} \,, \tag{13}$$

where the normalized noise matrix $\boldsymbol{\Upsilon}_{N\times N}$ consists of entries with sub-Gaussian distribution with variance ς^2, thus [24]

$$\mathbb{P}(|\boldsymbol{\Upsilon}_{ij}| \geq \beta) \leq 2\,\mathrm{e}^{-\frac{\beta^2}{2\varsigma^2}} \,. \tag{14}$$

Based on (10), $\boldsymbol{Z}^{\bar{s}}_{ij} = 2d^2_{ij}\boldsymbol{\Upsilon}_{ij} + d^2_{ij}\boldsymbol{\Upsilon}^2_{ij}$; thereby $\boldsymbol{Z}^{\bar{s}}_{ij}$ is also a sub-Gaussian random variable with a bounded constant $2\varsigma d^2_{ij}$. The physical setup confines $|\boldsymbol{Z}^{\bar{s}}_{ij}| \leq 4a^2$ where a is the radius of the table[2].

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 where the only available information is 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 [29]

$$\begin{aligned} \mathrm{dist}(\boldsymbol{X}, \hat{\boldsymbol{X}}) =& \frac{1}{N} \left\| \boldsymbol{J}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{J} - \boldsymbol{J}\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^T\boldsymbol{J} \right\|_{\mathrm{F}} \,, \\ & \boldsymbol{J} = \mathbb{I}_N - (1/N)\mathbf{1}_N\mathbf{1}_N^T \end{aligned} \tag{15}$$

[2]The sub-Gaussian assumption is exploited for the proof of Theorem 3 stated in Section 5. This model is not restrictive in practice and a Gaussian noise is considered for the simulations conducted in Section 7.

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm and $\mathbb{I}_N$ is the $N \times N$ identity matrix. The distance measure stated in (15) is useful to compare the performance of different methods in terms of microphone array geometry estimation.

Table 1 summarizes the set of important notation.

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 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[3], we need to know at least ηN of the row entries as well as ηN of the column entries reduced by η^2 of the repeated values to recover the entire elements of $\boldsymbol{M}$.

Given $\boldsymbol{M}^E$ defined in (12), the matrix completion recovers an estimate of the distance matrix $\hat{\boldsymbol{M}}$ through the following optimization

$$\begin{aligned} &\text{Minimize} && \operatorname{rank}(\hat{\boldsymbol{M}}) \\ &\text{subject to} && \hat{\boldsymbol{M}}_{ij} = \boldsymbol{M}_{ij}, \qquad (i,j) \in E \end{aligned} \tag{16}$$

In this paper, we use the procedure of OptSpace proposed by Keshavan et al. [24] for estimating a matrix given the desired rank η. This algorithm is implemented in three steps: (1)

[3]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 [30].

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 non-zero 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

$$\widetilde{\boldsymbol{M}}^E = \sum_{i=1}^{N} \sigma_i(\widetilde{\boldsymbol{M}}^E)\boldsymbol{U}_{.i}\boldsymbol{V}_{.i}^T \ , \tag{17}$$

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

$$\mathcal{P}_\eta(\widetilde{\boldsymbol{M}}^E) = (N^2/|E|)\sum_{i=1}^{\eta} \sigma_i(\widetilde{\boldsymbol{M}}^E)\boldsymbol{U}_{.i}\boldsymbol{V}_{.i}^T = \boldsymbol{U}_0\boldsymbol{S}_0\boldsymbol{V}_0^T \ . \tag{18}$$

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

$$\begin{aligned} F(\boldsymbol{U},\boldsymbol{V}) &= \min_{\boldsymbol{S}\in\mathbb{R}^{\eta\times\eta}} \mathcal{F}(\boldsymbol{U},\boldsymbol{V},\boldsymbol{S}), \\ \mathcal{F}(\boldsymbol{U},\boldsymbol{V},\boldsymbol{S}) &= \frac{1}{2}\sum_{(i,j)\in E} (\boldsymbol{M}_{ij} - (\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T)_{i,j})^2 \end{aligned} \tag{19}$$

$F(\boldsymbol{U},\boldsymbol{V})$ 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 [26].

To this end, we apply a Cadzow-like method. The Cadzow algorithm [31] (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 (19). 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

$$\hat{\boldsymbol{M}} \in \mathbb{S}_h^N \Longleftrightarrow \begin{cases} d_{ij} = 0 \Leftrightarrow \boldsymbol{x}_i = \boldsymbol{x}_j \\ d_{ij} > 0, \ i \neq j \\ d_{ij} = d_{ji} \end{cases} \tag{20}$$

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

decrease the distance between the estimated matrix and the EDM cone. This is visualized in Figure 2. Note that the illustration 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 illustrated graphically).

The projected matrix must satisfy the following EDM properties [32]

$$\hat{\boldsymbol{M}} \in \mathbb{EDM}^N \Longleftrightarrow \begin{cases} -z^T \hat{\boldsymbol{M}} z \geq 0 \\ \mathbf{1}^T z = 0 \\ (\forall \|z\| = 1) \\ \hat{\boldsymbol{M}} \in \mathbb{S}_h^N \end{cases} \tag{21}$$

The EDM properties include the triangle inequality, thus

$$d_{ij} \leq d_{ik} + d_{kj}, \quad i \neq j \neq k \,, \tag{22}$$

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

$$\begin{gathered} \cos(\tau_{jkl}+\tau_{lkj}) \leq \cos \tau_{ikj} \leq \cos(\tau_{ikl} - \tau_{lkj}) \\ 0 \leq \tau_{ikl},\, \tau_{lkj},\, \tau_{ikj} \leq \pi \end{gathered} \tag{23}$$

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 (21). 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

$$\mathcal{H}(\boldsymbol{X}) = \left\| \mathbf{1}_N \boldsymbol{\Lambda}^T + \boldsymbol{\Lambda} \mathbf{1}_N{}^T - 2\boldsymbol{X}\boldsymbol{X}^T - \boldsymbol{U}_c \boldsymbol{S}_c \boldsymbol{V}_c^T \right\|_{\mathrm{F}}^2 . \tag{24}$$

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 $\boldsymbol{x}_i = [x_{i1}, ..., x_{i\zeta}]^T$, $\mathcal{H}(\boldsymbol{X})$ 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 (24) to zero to obtain the new estimates, thus

$$\begin{aligned} \hat{\boldsymbol{X}} &= \arg\min_{\boldsymbol{X}} \mathcal{H}(\boldsymbol{X}) \\ (\boldsymbol{U}^{k+1}, \boldsymbol{V}^{k+1}, \boldsymbol{S}^{k+1}) &= \text{SVD}\ (\mathbf{1}_N \hat{\boldsymbol{\Lambda}}^T + \hat{\boldsymbol{\Lambda}} \mathbf{1}_N{}^T - 2\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^T) \end{aligned} \tag{25}$$

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 $k + 1/2$:

$$\begin{aligned} \boldsymbol{U}^{k+1/2} &= \boldsymbol{U}^k + \vartheta \frac{\partial F(\boldsymbol{U}^k, \boldsymbol{V}^k)}{\partial \boldsymbol{U}} \\ \boldsymbol{V}^{k+1/2} &= \boldsymbol{V}^k + \vartheta \frac{\partial F(\boldsymbol{U}^k, \boldsymbol{V}^k)}{\partial \boldsymbol{V}} \\ \boldsymbol{S}^{k+1/2} &= \arg\min_{\boldsymbol{S}} \mathcal{F}(\boldsymbol{U}^k, \boldsymbol{V}^k, \boldsymbol{S}) \end{aligned} \tag{26}$$

- iteration $k + 1$:

$$(\boldsymbol{U}^{k+1}, \boldsymbol{V}^{k+1}, \boldsymbol{S}^{k+1}) = \mathcal{P}_e(\boldsymbol{U}^{k+1/2}, \boldsymbol{V}^{k+1/2}, \boldsymbol{S}^{k+1/2}) \tag{27}$$

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 $\varsigma\, d_{ij}$ as stated in (14). 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, such that the output $\hat{\boldsymbol{X}}$ satisfies

$$\text{dist}(\boldsymbol{X}, \hat{\boldsymbol{X}}) \leq C_1 \frac{a^2 \log_2 N}{pN} + C_2 \varsigma \frac{d_{max}^2}{\sqrt{pN}} \tag{28}$$

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 $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^T$ is (μ_1, μ_2)-*incoherent* if the following conditions hold.

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

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

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

For a (μ_1, μ_2)-incoherent matrix $\boldsymbol{M}$, (29) is correct with probability greater than $1 - N^{-3}$; cf. [24]-Theorem 1.2.

$$\frac{1}{N} \|\boldsymbol{M} - \hat{\boldsymbol{M}}\|_{\text{F}} \leq \frac{C_1' \, \|\Psi_E(\boldsymbol{M}^s)\|_2 + C_2' \, \left\|\Psi_E(\boldsymbol{Z}^{\bar{s}})\right\|_2}{p\, N} \,, \tag{29}$$

provided that

$$|E| \geq C_1' N \kappa_\eta^2(\boldsymbol{M}) \max\left\{\mu_1 \eta \log N \,;\, \mu_1^2 \eta^2 \kappa_\eta^4(\boldsymbol{M}) \,;\, \mu_2^2 \eta^2 \kappa_\eta^4(\boldsymbol{M})\right\} \,, \tag{30}$$

and

$$\frac{C_1' \, \|\Psi_E(\boldsymbol{M}^s)\|_2 + C_2' \, \left\|\Psi_E(\boldsymbol{Z}^{\bar{s}})\right\|_2}{p\,N} \leq \sigma_\eta(\boldsymbol{M})/N \; , \tag{31}$$

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 (28) based on the following Theorems 2 and 3. In the second step, conditions (30) and (31) 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 $1 - N^{-3}$,

$$\|\Psi_E(\boldsymbol{M}^s)\|_2 \leq C_1'' \, a^2 \, \log_2 N \; . \tag{32}$$

The proof of this theorem is explained in Appendix 1.

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

$$\left\|\Psi_E(\boldsymbol{Z}^{\bar{s}})\right\| \leq C_2'' d_{\max}^2 \boldsymbol{\varsigma} \sqrt{pN} \; . \tag{33}$$

The proof of this theorem is explained in Appendix 2.

On the other hand, the following condition holds for any arbitrary network of microphones [33]

$$\mathrm{dist}(\boldsymbol{X}, \hat{\boldsymbol{X}}) \leq \frac{1}{N} \|\boldsymbol{M} - \hat{\boldsymbol{M}}\|_{\mathrm{F}} \; . \tag{34}$$

Therefore, based on Theorem 2, Theorem 3 and the relations (29) and (34), the upper bound stated in (28) is correct where $C_1 = C_1' C_1''$ and $C_2 = C_2' C_2''$; it is enough to investigate conditions (30) and (31) and (μ_1, μ_2)-incoherency of $\boldsymbol{M}$ to prove Theorem 1.

To show the inequality stated in (30), we can equivalently show that

$$Np \geq C_1' \mu^2 \eta^2 \kappa_\eta^6(\boldsymbol{M}) \log N \; , \tag{35}$$

where $\mu = \max(\mu_1, \mu_2)$. In order to show that (35) 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

$$\boldsymbol{M}_{ij} = \rho_i^2 + \rho_j^2 - 2\boldsymbol{x}_i^T \boldsymbol{x}_j \,, \tag{36}$$

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

$$\boldsymbol{M} = \boldsymbol{A}\mathcal{S}\boldsymbol{A}^T \,, \tag{37}$$

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

$$\boldsymbol{A} = \begin{bmatrix} a/2 & x_1 & y_1 & -a^2/4 + \rho_1^2 \\ \vdots & \vdots & \vdots & \vdots \\ a/2 & x_N & y_N & -a^2/4 + \rho_N^2 \end{bmatrix}, \tag{38}$$

and

$$\mathcal{S} = \begin{bmatrix} 2 & 0 & 0 & 2/a \\ 0 & -2 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 2/a & 0 & 0 & 0 \end{bmatrix}. \tag{39}$$

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

$$\mathcal{S} = \mathcal{W}\boldsymbol{\Gamma}\mathcal{W}^{-1} \,, \tag{40}$$

where

$$\boldsymbol{\Gamma} = \operatorname{diag}\left(-2, -2, \frac{a + \sqrt{4 + a^2}}{a}, \frac{a - \sqrt{4 + a^2}}{a}\right). \tag{41}$$

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 (37), we have

$$\sigma_1(\boldsymbol{M}) \leq \sigma_1(\mathcal{S})\,\sigma_1(\boldsymbol{A}\boldsymbol{A}^T)\,, \tag{42}$$

$$\sigma_4(\boldsymbol{M}) \geq \sigma_4(\mathcal{S})\,\sigma_4(\boldsymbol{A}\boldsymbol{A}^T)\,. \tag{43}$$

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

$$P_\rho(\rho) = \frac{2\rho}{a^2} \qquad \text{for} \quad 0 \leq \rho \leq a\,. \tag{44}$$

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

$$\mathbb{E}[\boldsymbol{A}^T\boldsymbol{A}] = \begin{bmatrix} Na^2/4 & 0 & 0 & Na^3/8 \\ 0 & Na^4/4 & 0 & 0 \\ 0 & 0 & Na^4/4 & 0 \\ Na^3/8 & 0 & 0 & 7Na^4/48 \end{bmatrix}. \tag{45}$$

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 [34], we get

$$\mathbb{P}(\sigma_1(\boldsymbol{A}\boldsymbol{A}^T) > 2N\sigma_{\max}(a)) \leq e^{-C'N}\,, \tag{46}$$

$$\mathbb{P}(\sigma_1(\boldsymbol{A}\boldsymbol{A}^T) < (1/2)N\sigma_{\max}(a)) \leq e^{-C'N}\,, \tag{47}$$

$$\mathbb{P}(\sigma_4(\boldsymbol{A}\boldsymbol{A}^T) < (1/2)N\sigma_{\min}(a)) \leq e^{-C'N}\,, \tag{48}$$

for a constant C'. Hence, with high probability, based on relations (42), (43), (46) and (48), we have

$$\kappa_4(\boldsymbol{M}) \leq \frac{4\sigma_{\max}(a)\,\sigma_1(\mathcal{S})}{\sigma_{\min}(a)\,\sigma_4(\mathcal{S})} = f_{\kappa_4}(a)\ . \tag{49}$$

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 (37), 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

$$\|\boldsymbol{V}_{i.}\|^2 \leq \eta\,\mu_1 \qquad \forall\ i \in [N]\ , \tag{50}$$

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

$$\|\boldsymbol{V}_{i.}\|^2 \leq \sigma_4(\boldsymbol{A})^{-2}\|\boldsymbol{A}_{i.}\|^2\,N\ . \tag{51}$$

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

$$f_{\mu_1}(a) = \frac{5a^2/2 + 9a^4/8}{\sigma_{\min}(a)}\ , \tag{52}$$

and based on (48) and (51), with high probability we have

$$\|\boldsymbol{U}_{i.}\|^2 \leq f_{\mu_1}(a) \qquad\qquad \forall\ i \in [N]\ . \tag{53}$$

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 (43) and (48) we have

$$\sigma_1(\boldsymbol{M}) \geq \sigma_4(\boldsymbol{M}) \geq \frac{1}{2}N\,\sigma_{\min}(a)\,\sigma_4(\mathcal{S})\ , \tag{54}$$

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

$$\left|\boldsymbol{M}_{ij}/\sigma_1(\boldsymbol{M})\right| \leq \frac{f_{\mu_2}(a)}{N} \qquad \forall\, i, j \in [N]\,. \tag{55}$$

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 (30) and (35) are correct.

Further, (31) holds with high probability, if the right-hand side of (28) is less than $C_3\,\sigma_{\min}(a)\,\sigma_4(\mathcal{S})$, since based on (48), $\frac{\sigma_\eta(\boldsymbol{M})}{N} \geq \frac{1}{2}\sigma_{\min}(a)\sigma_4(\mathcal{S})$. This finishes the proof of Theorem 1.

■

The theoretical analysis elaborated in this section, elucidates a link between the performance of microphone array calibration and the number of microphones, noise level and the ratio of missing pairwise distances. In Section 7, thorough evaluations are conducted that demonstrate these theoretical insights. Furthermore, 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 relative (up to a rigid transformation) microphone positions $\boldsymbol{x}_i$, $i \in \{1, \ldots, N\}$ from the measurements of pairwise distances. Some of the state-of-the-art methods to achieve this goal are (1) Multi-Dimensional Scaling (MDS) [35], (2) Semi-Definite Programming (SDP) [36] and S-Stress [29] 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 [19]. Given a distance matrix, finding the relative microphone positions is achieved by MDSLocalize [29]. In the ideal case where matrix $\boldsymbol{M}$ is complete and noiseless, this algorithm outputs the relative positions of the microphones. At the first step, a double centering transformation is applied to $\boldsymbol{M}$ to subtract the row and column means of the distance matrix via $\Xi(\boldsymbol{M}) = \frac{-1}{2}\boldsymbol{J}\,\boldsymbol{M}\,\boldsymbol{J}$ where $\boldsymbol{J} = \mathbb{I}_N - 1/N\mathbf{1}_N\mathbf{1}_N^T$. The ζ largest eigenvalues and the corresponding eigenvectors of $\Xi(\boldsymbol{M})$ denoted by $\boldsymbol{\Pi}_+$ and $\boldsymbol{U}_+$ are calculated and the microphone positions are obtained as $\boldsymbol{X} = \boldsymbol{U}_+\sqrt{\boldsymbol{\Pi}_+}$.

In a real scenario of missing distances, a modification called MDS-MAP [35] 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

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \sum_{(i,j)\in E} w_{ij}\left|\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2 - \tilde{d}_{ij}^2\right| , \tag{56}$$

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 (56) is not convex but can be relaxed as a convex minimization via

$$\begin{aligned} &\min_{\boldsymbol{X},\boldsymbol{Y}} \sum_{(i,j)\in E} w_{ij}\left|(u_i - u_j)^T[\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{X}^T, \mathbb{I}_\zeta](u_i - u_j)^T - \tilde{d}_{ij}^2\right| \\ &\text{subject to} \quad [\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{X}^T, \mathbb{I}_\zeta] \succeq 0, \quad \left\|\boldsymbol{X}^T\mathbf{1}_N\right\| = 0 \end{aligned} \tag{57}$$

where $\boldsymbol{Y}_{N\times N}$ is a positive semidefinite matrix and $\succeq$ is a generalized matrix inequality on the positive semidefinite cone [37]. To further increase the accuracy, a gradient decent is applied on the output of SDP minimization [36].

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

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \sum_{(i,j)\in E} w_{ij}\left(\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2 - \tilde{d}_{ij}^2\right)^2 . \tag{58}$$

The reliability measure w_{ij} controls the least square regression stated in equation (58) 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 [38].

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 is known to perform fairly well with missing distance information. Together with its final gradient descent phase, has been shown to 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 it is known to perform well in many conditions, 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 a 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 has 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 performance. In this section, we present thorough evaluation of ad hoc microphone array calibration on simulated setups and real data recordings.

7.2. Simulated Data Evaluations

The simulated experiments are conducted to evaluate the performance of the proposed method and compare and contrast it against the state-of-the-art alternative approaches in different scenarios with varying number of microphones, magnitude of the pairwise distance measurements errors, percentage of missing distances as well as jitter.

The presented evaluation relies on a local connectivity assumption in pairwise distance measurements. We do not assume a particular (e.g. diffuse noise) model for pairwise distance estimation and the conclusions of this section hold for a general ad hoc array calibration framework where the pairwise distances may be provided by any other means meeting the local connectivity assumption.

7.2.1. Performance for Different Numbers of Microphones

In this section, we present the performance of ad hoc array calibration when the number of microphones varies from 15 to 200. The microphones are uniformly distributed on a disc of diameter 19 m. The maximum pairwise distance that can be measured is 7.5 m. In addition, 5% of the distances are assumed to be randomly missing. Hence, the total missing entries vary from

42% to 60%. The standard deviation of the noise on measured distances (expressed through (13)-(14)) between two microphones i and j is ςd_{ij} where $\varsigma = 0.0167$; the dependency of the noise level on the distance is due to the limitation of the diffuse noise coherence model for pairwise distance estimation as elaborated in [28].

The results for each number of microphones are averaged over 500 random configurations. The calibration error is quantified using the metric defined in (15). Furthermore, the absolute position of the microphones is estimated using the nonlinear optimization method [39] and the mean position error as defined in (59) over all configurations is evaluated. Figures 3 and 4 illustrate the results; the error bars are shown for one standard deviation from the mean estimates. The Cramér rao bound (CRB) is quantified using the method elaborated in [12, 14].

The results show that the performance improves as the number of microphones increases. This observation is inline with the theoretical analysis provided in Section 5. The best results are achieved by the proposed E-MC^2 algorithm as it confines the search space to the Euclidean space through iterative EDM projections. We can see that for the number of microphones above 45, the error in position estimation is less than 6.2 cm and it reduces to 2.2 cm for 200 microphones. Although the mathematical proof of the unbiasedness of the proposed estimator is not achieved in this paper, we empirically found no evidence of bias. Therefore, CRB provides a reasonable benchmark for our evaluation.

7.2.2. Performance for Different Noise Levels

To evaluate the effect of noise on calibration performance, similar (500) configurations of 45 microphones as generated in Section 7.2.1 are simulated. The level of white Gaussian noise added to the measured pairwise distance d_{ij} are varying as ςd_{ij} where $\varsigma = \{0.0056, \ldots, 0.1\}$. Figures 5 and 6 illustrate the results. We can see that the performance improves as the noise level gets smaller.

Based on the theoretical analysis of Section 5 as expressed in (28), a linear relationship between the calibration error of matrix completion and ς is expected. The empirical observations are in line with this theoretical insight. As depicted in Figure 5, for $\varsigma < 0.0167$, the second term in (28) is getting too small so the first term becomes dominant as the slope of the error reduction

is reduced.

7.2.3. Performance for Different Missing Ratios

To study the sensitivity of the proposed algorithm to different levels of missing distances, a cubic room of unit dimensions ($1 \times 1 \times 1$ m^3) is simulated and 60 microphones are distributed uniformly at random positions. 300 random configurations are generated and the average mean position error is evaluated. As an alternative approach, the self-calibration method proposed by Crocco et al. [14] is implemented considering 30 sources and 30 sensors (thus 60 nodes in total). It may be noted that the number of nodes for calibration is equal for both approaches. The distances between all source and microphone pairs are known. Some of the distances are assumed to be missing at random. In addition, white Gaussian noise with standard deviation 0.02 m is added to the known distances. The simulated scenario mimics the evaluation setups of [14] and requires fixing the position of two microphones to derive the network position.

Figure 7 illustrates the errors in position estimation for different ratios of missing distances. We can see that up to 50% missing are effectively handled by the proposed algorithm. The rigorous analysis provided in Section 5 requires that $Np \gg \log N$ for the calibration error to be bounded; when the ratio of random missing entries is 60% (i.e. $p = 0.4$), we have $Np/\log N = 5.85$ (violating the condition $\gg$) so the error in calibration is expected to increase significantly. The theoretical analysis is confirmed by this empirical observation.

7.2.4. Effect of Jitter on Calibration Performance

The study presented in this paper assumes that the microphones are synchronized prior to calibration. If a pilot signal at sampling frequency f =16 kHz is used for synchronization, the effect of jitter can be modeled by a uniform error in distance measures as $[\frac{-c}{2f}, \frac{c}{2f}]$ where c is the speed of sound and set to 340 m/s. Hence, we can model the jitter as an additional uniform noise on the distance measures within the range of $[-1.065, 1.065]$ cm.

The effect of jitter is evaluated for different levels of noise on the distances. The number of microphones is 45 distributed on a disc of diameter 19 m. 60% of distances are missing consisting of 5% random and 55% structured. The experiments are repeated for 300 random configurations

and the average calibration error and position estimation error are quantified. Figure 8 illustrates the results. We can see that the effect of jitter on position estimation increases from 3 mm to 8 mm and its effect on calibration error increases from 0.01 m^2 to 0.09 m^2 as the distances are measured more accurately (smaller ς).

7.2.5. Distributed Array Calibration

To further study the performance of the proposed approach for distributed array calibration, two scenarios are simulated. In the first scenario, a room of dimensions $11\times8\times5$ m^3 is considered which yields d_{max} = 101 cm [28] [4]. The reverberation time is about 430 ms. Two sets of 9-channel circular uniform microphone array of diameter 20 cm are simulated where the center of both compact arrays are 1 m apart. In the second scenario, a room of dimensions $8\times5.5\times3.5$ m^3 is considered which yields d_{max} = 73 cm [26]. The reverberation time is about 300 ms. A circular 9-channel microphone array of diameter 20 cm located inside another 6-channel circular array of diameter 70 cm is simulated.

The standard deviation of the noise on distance measures is $\varsigma\, d_{ij}$ where ς = 0.06. There are no random missing entries and all of the missing distances are due to the limitation of the diffuse noise model in distance estimation thus around 25% of the distances are missing in the first scenario (18-mic) and around 30% of the distances are missing in the second scenario (15-mic). The results are listed in Table 2. We can see that the positions are estimated with less than 1.6 cm error. Furthermore, we repeated the same experiment 25 times and averaged the estimates of the positions. We can see that the error after averaging is noticeably reduced.

7.3. Real Data Evaluation

The real data recordings are collected at Idiap's smart meeting room [40].

[4] The maximum distance that can be estimated using the diffuse noise model depends on the size of the room and acoustic parameters. A linear relation between the maximum measurable distance and the room dimension has been shown rigorously [28]. Nevertheless, application of the diffuse noise method for pairwise distance estimation is just an example use case of the proposed algorithm and many alternative approaches can be exploited.

7.3.1. 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 20 cm and one microphone is at the center. There are two additional microphones at 70 cm distance from the central microphone. The microphones are Sennheiser MKE-2-5-C omnidirectional miniature lapel type. Although the recording setup for collecting data is regular, the uniform geometry of the microphone array provides no particular constraint. Hence, without loss of generality, we rely on this available setup to evaluate the performance of the proposed approach.

The floor of the room is covered with carpet and surrounded with plaster walls having two big windows. The enclosure is a $8 \times 5.5 \times 3.5$ m^3 rectangular room and it is moderately reverberant; the reverberation time is about 300 ms. It contains a centrally located 4.8×1.2 m^2 rectangular table. This scenario mimics the MONC database [40]. The sampling rate is 48 kHz while the processing applied for microphone calibration is based on down-sampled signal of rate 16 kHz to reduce the computational cost of pairwise distance estimation.

7.3.2. Pairwise Distance Estimation

In order to estimate the pairwise distances, we take two microphone signals of length 2.14 s, 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 [26]. 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 9).

7.3.3. 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 is fixed and used to calibrate the rest of the network. Figure 11 demonstrates the results of MDS-MAP, SDP, s-stress and the proposed Euclidean distance matrix completion algorithm. The calibration error is quantified based on (15). The best results are achieved by the proposed algorithm with error 5.85 cm^2. The second place belongs to s-stress with error 6.14 cm^2 followed by MDS-MAP and SDP with errors 8.13 cm^2 and 8.63 cm^2 respectively.

Figure 12 provides a comparative illustration of the results of matrix completion (MC), MC+Cadzow (MC^2) and the proposed Euclidean distance matrix completion (E-MC^2) algorithm. We can see that MC^2 yields better result with error 7.68 compared to MDS-MAP, SDP and MC, but worse than s-stress. The proposed E-MC^2 algorithm achieves the best performance.

The scenario using eleven channels of microphones addresses the problem of having partial estimates of the distances for calibration of a microphone array. The experiments show that the proposed method offers the best estimation of the geometry as illustrated in Figure 9 and 10 with an error of 49.6 cm^2. As we can see, the proposed Euclidean distance matrix completion algorithm achieves less than half the error of the best state-of-the-art alternative.

The worst result belongs to MDS-MAP with error 434.4 cm^2 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 cm^2 and 125 cm^2. 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 (MC^2) and the proposed method (E-MC^2) and achieves the best performance. These experimental evaluation confirm the effectiveness of the proposed algorithm and demonstrate the hypothesis that incorporating the EDM properties in matrix completion algorithm enables calibration of microphone arrays from partial measurements of the pairwise distances.

The theoretical analysis provided in Section 5 elucidates a link between the calibration er-

ror and the number of microphones. To demonstrate this relation, a calibration of a 8-channel circular array when the distances are all measured is performed. In addition, an extra microphone (#12) is also included which is located with a symmetry to microphone 10. Hence, $d_{12,11}, d_{10,12}, d_{3,12}, d_{4,12}, d_{5,12}$ are also missing. The calibration errors are listed in Table 3.

Furthermore, in addition to the calibration error expressed in (15), we apply the nonlinear optimization proposed in [39] to find the best match between $\hat{\boldsymbol{X}}$ and $\boldsymbol{X}$ by considering various rigid transformations and quantify the position error as

$$\frac{1}{N}\sum_{n=1}^{N}\|\hat{\boldsymbol{x}}_n - \boldsymbol{x}_n\|_2 \,. \tag{59}$$

The position errors are listed in Table 4. The results show that considering further microphone improves the calibration performance which is in line with the theoretical analysis of Section 5.

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 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. 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 (11), we define matrix $\mathcal{E}$ as

$$\mathcal{E}_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E \cap S \\ 0 & \text{otherwise} \end{cases} \tag{60}$$

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

$$\nu_i = \sum_{j\in[N]} |\mathcal{E}_{ij}| \ . \tag{61}$$

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

$$q = \mathbb{P}\{|\boldsymbol{x}_i - \boldsymbol{x}_j| \geq d_{\max}\} \ . \tag{62}$$

Hence, we have

$$\mathbb{E}[\nu_i] = Npq \ , \tag{63}$$

For bounding $\mathbb{E}[\nu_i]$, it is necessary to bound q. Figure 13.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 13.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

$$q_{\min}(a, d_{\max}) \leq q \leq q_{\max}(a, d_{\max}) \ . \tag{64}$$

As illustrated in Figure 13. $q_{\min}(a, d_{\max}) = \max\{1 - (\frac{d_{\max}}{a})^2, 0\}$ 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

$$q_{\max} = 1 - \frac{2\gamma}{\pi} + \frac{1}{2\pi} \sin 4\gamma + \frac{2\xi^2}{\pi}[2\gamma + \sin 2\gamma] - 2\xi^2 \ , \tag{65}$$

where $\xi = d_{\max}/2a$ and $\gamma = \sin^{-1} \xi$. Based on (63) and (64) we have

$$Npq_{\min}(a, d_{\max}) \leq \mathbb{E}[\nu_i] \leq Npq_{\max}(a, d_{\max}) \ . \tag{66}$$

By applying the Chernoff bound to ν_i we have

$$\mathbb{P}(\nu_i > (1 + \epsilon)\mathbb{E}[\nu_i]) \leq 2^{-(1+\epsilon)\mathbb{E}[\nu_i]} \ , \tag{67}$$

where ϵ is an arbitrary positive constant. Therefore, based on (66) we have

$$\mathbb{P}(\nu_i > (1 + \epsilon)Np\, q_{\max}) \leq 2^{-(1+\epsilon)Np\, q_{\min}} \ . \tag{68}$$

By applying the union bound we have

$$\mathbb{P}(\max_{i\in[N]} \nu_i > (1 + \epsilon)Np\, q_{\max}) \leq 2^{-(1+\epsilon)Np\, q_{\min} + \log_2 N} \ . \tag{69}$$

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[5] or in other words, $d_{\max}$ in-

[5]This assumption can be dropped to achieve a tighter bound, but it increases the complexity of the proof.

creases as the size of the network N grows. Therefore, we have

$$\mathbb{P}(\max_{i\in[N]} \nu_i > (1+\epsilon)Np\, q_{\max}) \leq N^{-\theta}\ , \tag{70}$$

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

$$\|\Psi_E(\boldsymbol{M}^s)\|_2 \leq 4a^2 \max_{i\in[N]} \nu_i\ , \tag{71}$$

and based on (70)

$$\|\Psi_E(\boldsymbol{M}^s)\|_2 \leq 4a^2(1+\theta)q_{\max}N\ . \tag{72}$$

Therefore, we achieve

$$\|\Psi_E(\boldsymbol{M}^s)\|_2 \leq C_1'' a^2 \log_2 N\ . \tag{73}$$

■

Appendix 2. Proof of Theorem 3

Based on the noise model described in Section 3.3, $\boldsymbol{Z}_{ij}^{\bar{s}}$ is obtained as

$$\boldsymbol{Z}_{ij}^{\bar{s}} = d_{ij}^2 \boldsymbol{\Upsilon}_{ij}\left(2 + \boldsymbol{\Upsilon}_{ij}\right) \approx 2d_{ij}^2 \boldsymbol{\Upsilon}_{ij}, \tag{74}$$

where $d_{ij} \leq d_{\max}$ and based on concentration inequality for 1-Lipschitz function $\|.\|$ on i.i.d random variables $\Psi_E(\boldsymbol{Z}^{\bar{s}})$ with zero mean and sub-Gaussian tail with parameter $4\varsigma^2 d_{\max}^4$(14), (74) [41]

$$\mathbb{P}\left(\left|\ \left\|\Psi_E(\boldsymbol{Z}^{\bar{s}})\right\| - \ \mathbb{E}\left(\left\|\Psi_E(\boldsymbol{Z}^{\bar{s}})\right\|\right)\ \right| \ > t\ \right) \leq \exp\left(\frac{-t^2}{8\ \varsigma^2 d_{\max}^4}\right). \tag{75}$$

By setting $t = 2d_{\max}^2\sqrt{6\varsigma^2 \log N}$ we have

$$\left\|\Psi_E(\boldsymbol{Z}^{\bar{s}})\right\| \leq \ \mathbb{E}\left(\left\|\Psi_E(\boldsymbol{Z}^{\bar{s}})\right\|\right) + 2d_{\max}^2\sqrt{6\varsigma^2 \log N} \tag{76}$$

with probability bigger than $1 - N^{-3}$. So we need to extract bound for expectation of $\Psi_E(\boldsymbol{Z}^{\bar{s}})$ that

has symmetric random enties. By using Theorem 1.1 from [42],

$$\mathbb{E}\left(\left\|\Psi_E(\boldsymbol{Z}^{\bar{s}})\right\|\right) \leq C_4\ \mathbb{E}\left(\max_{j\in[N]}\left\|\Psi_E(\boldsymbol{Z}^{\bar{s}}_{.j})\right\|\right) \tag{77}$$

Furthermore by using union bound and with apply Chernoff bound on the sum of independent random variables [24]

$$\mathbb{E}\left(\max_{j\in[N]}\left\|\Psi_E(\boldsymbol{Z}^{\bar{s}}_{.j})\right\|^2\right) \leq C_5 d^4_{max}\ \varsigma^2 pN \tag{78}$$

Since

$$\mathbb{E}\left(\max_{j\in[N]}\left\|\Psi_E(\boldsymbol{Z}^{\bar{s}}_{.j})\right\|\right) \leq \sqrt{\mathbb{E}\left(\max_{j\in[N]}\left\|\Psi_E(\boldsymbol{Z}^{\bar{s}}_{.j})\right\|^2\right)} \tag{79}$$

Base on relations (77), (78) and (79)

$$\mathbb{E}\left(\left\|\Psi_E(\boldsymbol{Z}^{\bar{s}})\right\|\right) \leq C_6 d^2_{\max}\varsigma\sqrt{pN} \tag{80}$$

By using (80) and (76) for $pN \gg \log N$ we have

$$\left\|\Psi_E(\boldsymbol{Z}^{\bar{s}})\right\| \leq\ C''_2 d^2_{\max}\varsigma\sqrt{pN} \tag{81}$$

■

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). The authors would like to acknowledge the anonymous reviewers for their comments and remarks to improve the quality and clarity of the manuscript.

[1] K. Kumatani, J. McDonough, B. Raj, Microphone Array Processing for Distant Speech Recognition: From Close-talking Microphones to Far-field Sensors, IEEE Signal Processing Magazine, Special Issue on Fundamental Technologies in Modern Speech Recognition .

[2] C.-T. Do, M. J. Taghizadeh, P. N. Garner, Combining Cepstral Normalization and Cochlear Implant-Like Speech

Processing for Microphone Array-based Speech Recognition, in: IEEE workshop on Spoken Language Technology (SLT), 2012.

[3] A. Asaei, Model-based Sparse Component Analysis for Multiparty Distant Speech Recognition, Ph.D. thesis, École Polytechnique Fédéral de Lausanne (EPFL), 2013.

[4] M. Wolfel, J. McDonough, Distant Speech Recognition, New York: John Wiley & Sons .

[5] 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.

[6] M. Brandstein, D. Ward (Eds.), Microphone Arrays: Signal Processing Techniques and Applications, Springer, 2001.

[7] 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.

[8] H. T. Do, Robust cross-correlation-based methods for sound-source localization and separation using a large-aperture microphone array, Ph.D. thesis, Brown University, 2011.

[9] A. Asaei, M. J. Taghizadeh, H. Bourlard, V. Cevher, Multi-party Speech Recovery Exploiting Structured Sparsity Models, in: 12th Annual Conference of the International Speech Communication Association (INTERSPEECH), 2011.

[10] A. Asaei, M. Davies, H. Bourlard, V. Cevher, Computational Methods for structured Sparse Component Analysis of Convolutive Speech Mixtures, in: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2012.

[11] 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).

[12] 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).

[13] N. D. Gaubitch, W. B. Kleijn, R. Heusdens, Auto-localization in ad-hoc microphone arrays, in: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2013.

[14] M. Crocco, A. Del Bue, V. Murino, A bilinear approach to the position self-calibration of multiple sensors, IEEE Transactions on Signal Processing 60 (2) (2012) 660–673.

[15] 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.

[16] M. Pollefeys, D. Nister, Direct computation of sound and microphone locations from time-difference-of-arrival data., in: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2445–2448, 2008.

[17] Y. Kuang, K. Åström, Stratified sensor network self-calibration from TDOA measurements, EUSIPCO, 2013.

[18] I. McCowan, M. Lincoln, I. Himawan, Microphone Array Shape Calibration in Diffuse Noise Fields, IEEE Transactions on Audio,Speech and Language Processing 16(3).

[19] T. F. Cox, M. A. A. Cox, Multidimensional Scaling, Chapman-Hall .

[20] M. Hennecke, T. Plotz, G. Fink, J. Schmalenstroeer, R. Haeb-Umbach, A hierarchical approach to unsupervised shape calibration of microphone array networks, in: IEEE/SP 15th Workshop on Statistical Signal Processing (SSP), 257–260, 2009.

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

[22] E. J. Candes, Y. Plan, Matrix Completion with Noise, IEEE Signal Processing Magazine 98(6).

[23] R. H. Keshavan, S. Oh, OptSpace: A Gradient Descent Algorithm on the Grassman Manifold for Matrix Completion `arXiv:0910.5260`.

[24] R. H. Keshavan, A. Montanari, S. Oh, Matrix Completion from Noisy Entries, Journal of Machine Learning Research 11.

[25] 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.

[26] 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.

[27] 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.

[28] M. J. Taghizadeh, P. N. Garner, H. Bourlard, Enhanced diffuse field model for ad hoc microphone array calibration, Signal Processing 101 (0) (2014) 242 – 255.

[29] I. Borg, P. J. Groenen, Modern Multidimensional Scaling Theory and Applications, in: Springer, 2005.

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

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

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

[33] R. Parhizkar, Euclidean Distance Matrices: Properties, Algorithms and Applications, Ph.D. thesis, École cole polytechnique fédérale de Lausanne EPFL, 2014.

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

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

[36] 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.

[37] S. Boyd, L.Vandenberghe, Convex Optimization, in: Cambridge, 2012.

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

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

[40] D. Moore, I. McCowan, The Multichannel Overlapping Numbers Corpus, Idiap resources available online at `www.cslu.ogi.edu/corpora/monc.pdf`, 2003.

[41] M. Talagrand, A new look at independence, The Annals of probability (1996) 1–34.

[42] Y. Seginer, The Expected Norm of Random Matrices, Combinatorics, Probability and Computing 9.

Table 1: Summary of the notation.

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

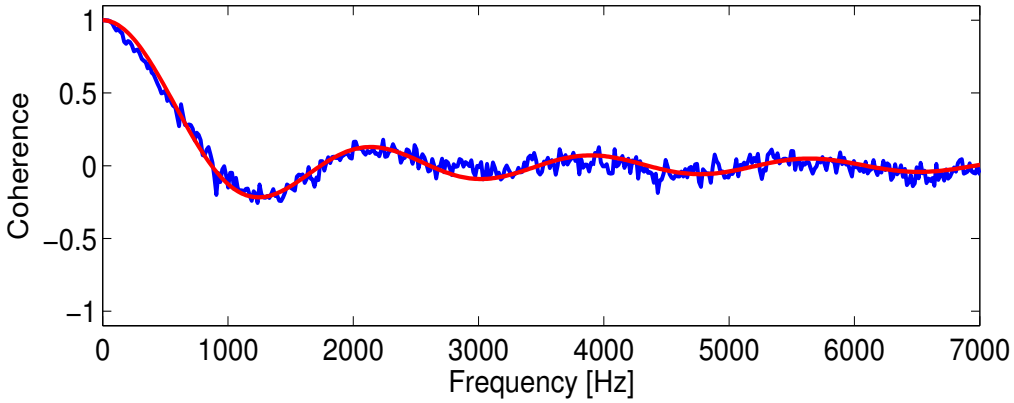

Figure 1: Coherence of the signal of two microphones at $d_{ij} = 20$ cm and the fitted sinc function using real data recordings.

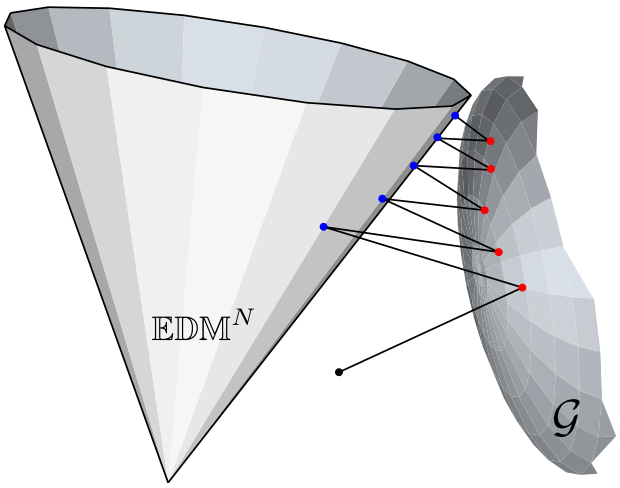

Figure 2: Matrix completion with projection onto the EDM cone.

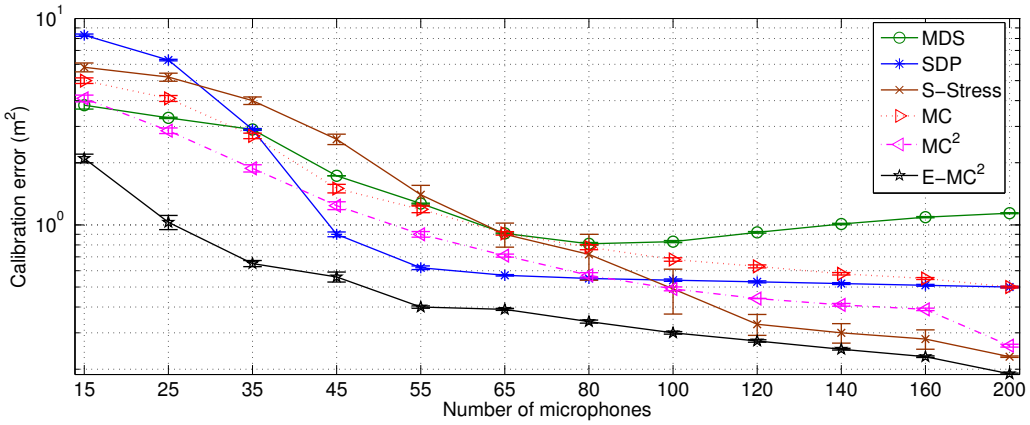

Figure 3: Calibration error (logarithmic scale) as defined in (15) versus the number of microphones. The standard deviation of noise on measured distances is ςd_{ij} where $\varsigma = 0.0167$. The error bars correspond to one standard deviation from the mean estimates.

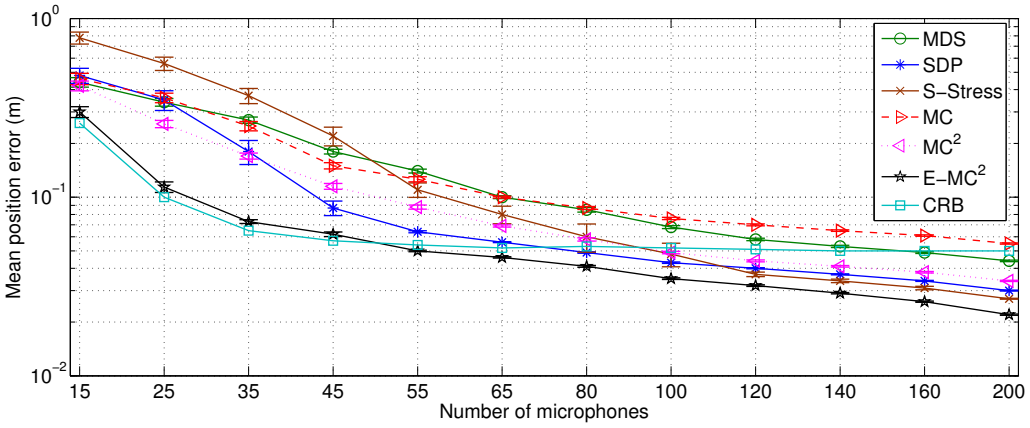

Figure 4: Mean position error (logarithmic scale) as defined in (59) versus the number of microphones. The standard deviation of the noise on measured distances is ςd_{ij} where $\varsigma = 0.0167$. The error bars correspond to one standard deviation from the mean estimates.

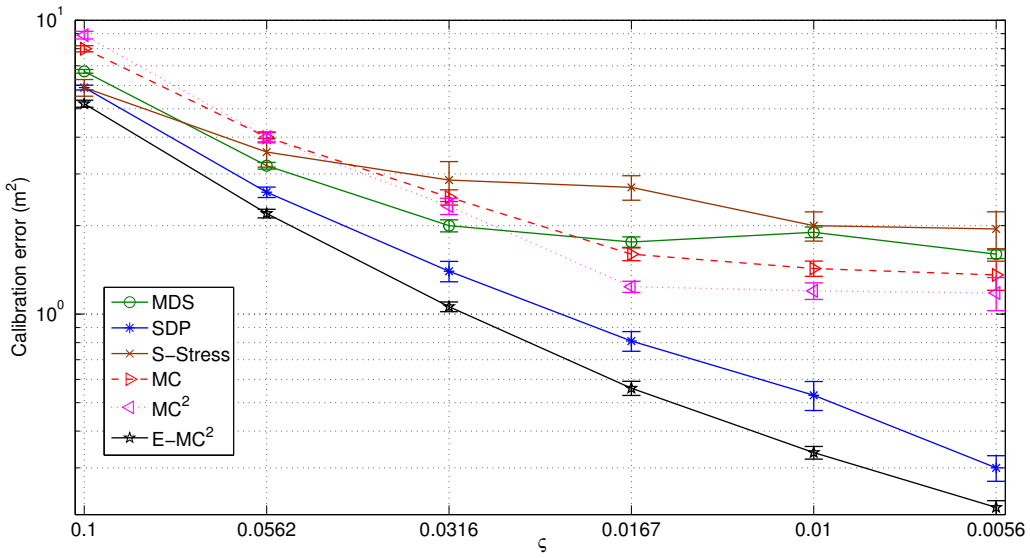

Figure 5: Calibration error (logarithmic scale) as quantified in (15) versus ς. The error bars correspond to one standard deviation from the mean estimates.

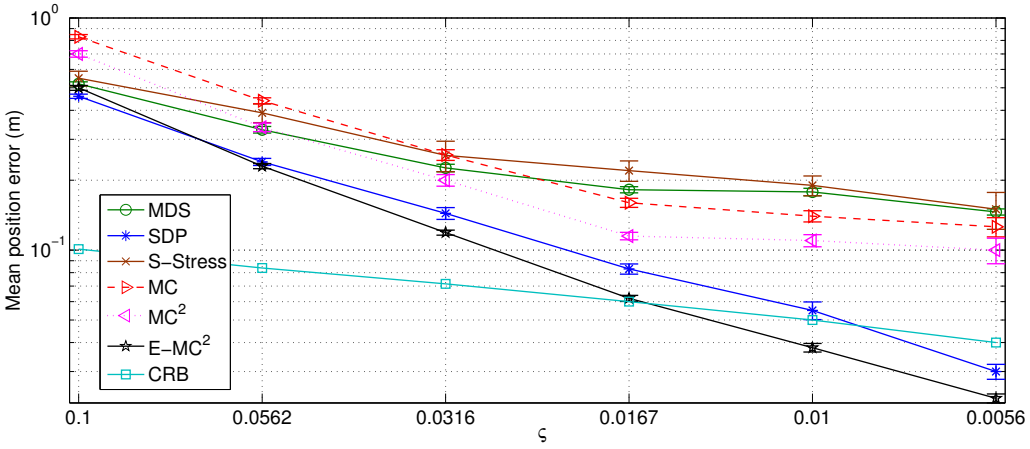

Figure 6: Mean position error (logarithmic scale) as defined in (59) versus ς. The error bars correspond to one standard deviation from the mean estimates.

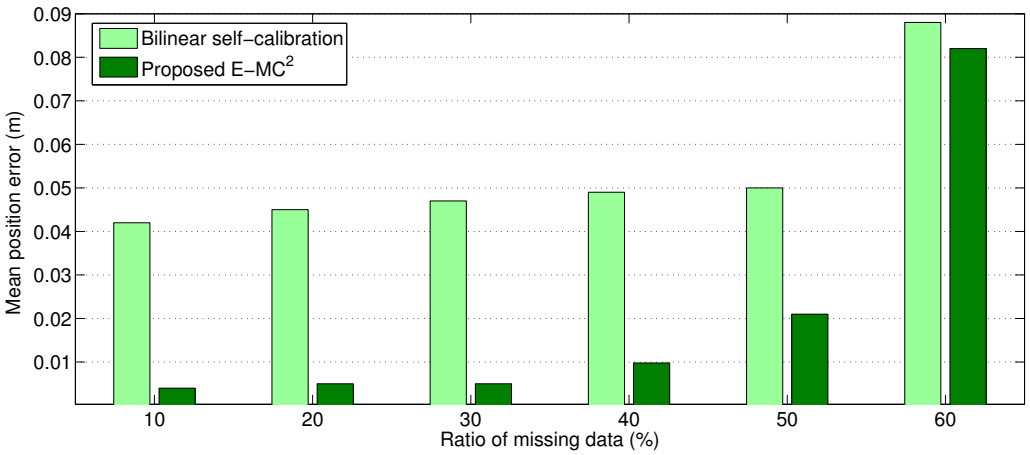

Figure 7: Mean position calibration error versus the ratio of missing pairwise distances for 30 sources and 30 microphones (60 nodes in total) considered in the self-calibration method [14] and 60 microphones used for the proposed E-MC^2 algorithm. The standard deviation of noise in pairwise distance estimation is 0.02.

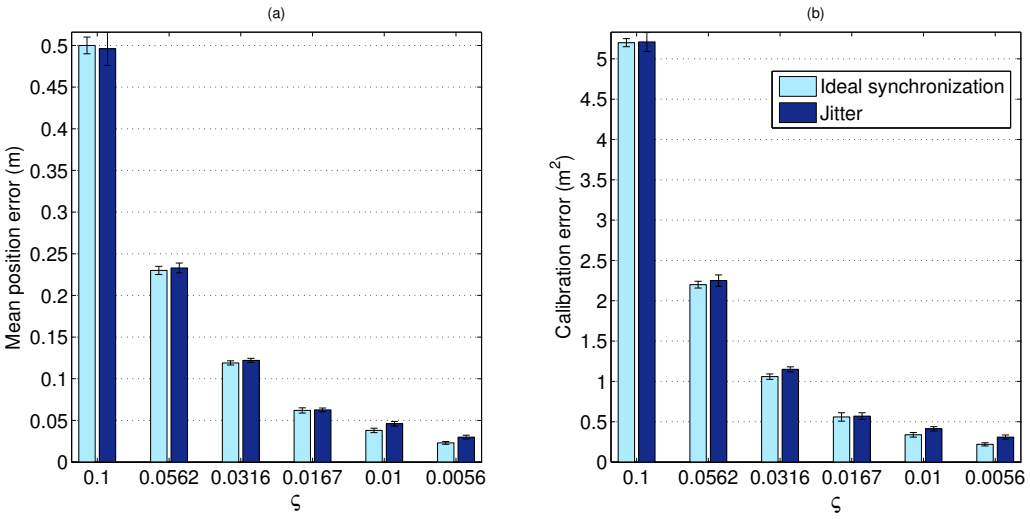

Figure 8: Effect of jitter on E-MC^2 algorithm quantified in terms of (a) mean position error as defined in (59) as well as (b) calibration error as defined in (15) versus ς. The error bars correspond to one standard deviation from the mean estimates. The number of microphones is 45 and 60% of the pairwise distances are missing.

Table 2: Performance of microphone array calibration in two scenarios. (1) Scenario 18-mic: two sets of 9-channel circular microphone array of diameter 20 cm; the center of both compact arrays are 1 m apart, and (2) Scenario 15-mic: a circular 9-channel microphone array of diameter 20 cm is located inside another 6-channel circular array of diameter 70 cm. The mean *position* error (cm) and the *calibration* error (cm^2) as defined in (15) are evaluated for different methods . The numbers in parenthesis corresponds to the error in position estimation if the experiments are repeated and averaged over 25 trials.

	Scenario 18-mic		**Scenario 15-mic**	
	Position (cm)	**Calibration (cm^2)**	**Position (cm)**	**Calibration (cm^2)**
MDS-MAP	3.3 (0.72)	175.8	3.18 (0.73)	170.5
SDP	2.1 (0.3)	96.3	4.64 (0.65)	258.8
S-Stress	6.8 (0.96)	265	7.05 (0.92)	281.5
MC	6.9 (1.35)	272	7.5 (1.55)	305
MC^2	6.56 (0.91)	225.1	6.8 (0.94)	274
E-MC^2	1.58 (0.37)	95.5	1.71 (0.41)	105.83

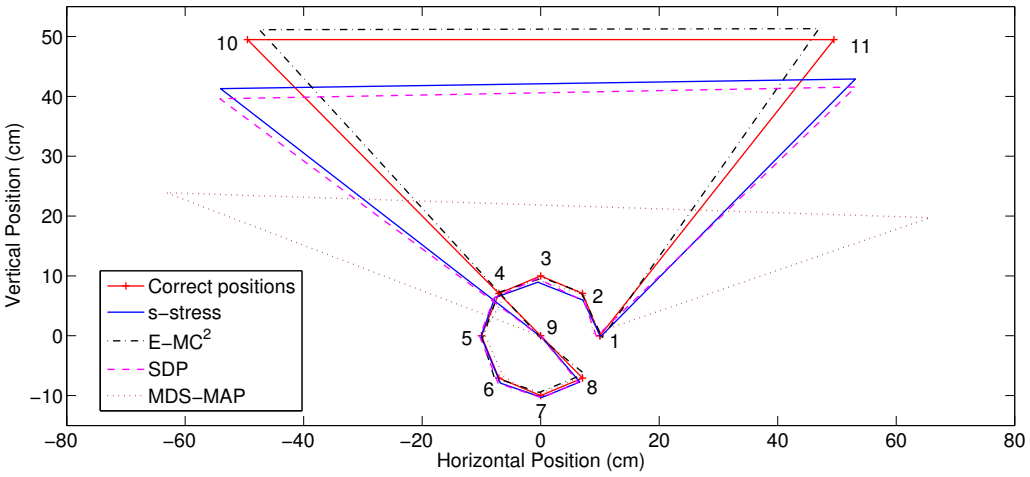

Figure 9: Calibration of the eleven-element microphone array while several pairwise distances are missing. The geometries are estimated using MDS-MAP, SDP, S-stress and the proposed proposed algorithm E-MC^2.

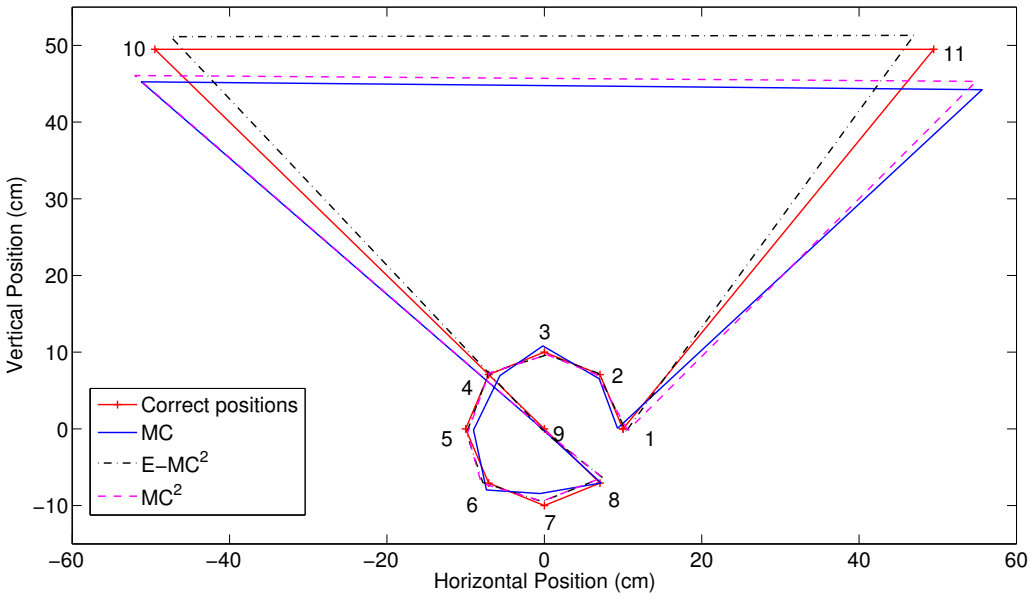

Figure 10: Calibration of the eleven-element microphone array while several pairwise distances are missing. The geometries are estimated using MC, MC+Cadzow (MC^2), and the proposed algorithm E-MC^2.

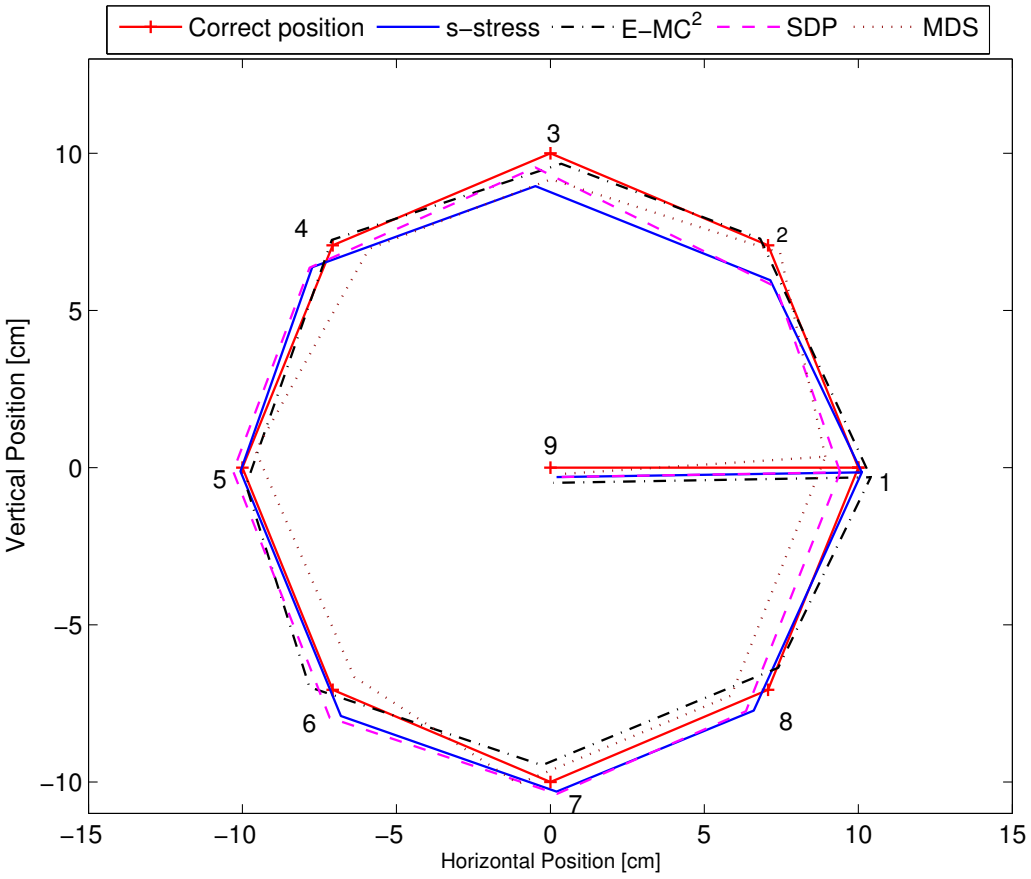

Figure 11: Calibration of the nine-element microphone array. The geometries are estimated using MDS-MAP, S-stress, SDP and the proposed Euclidean distance matrix completion algorithm, E-MC^2.

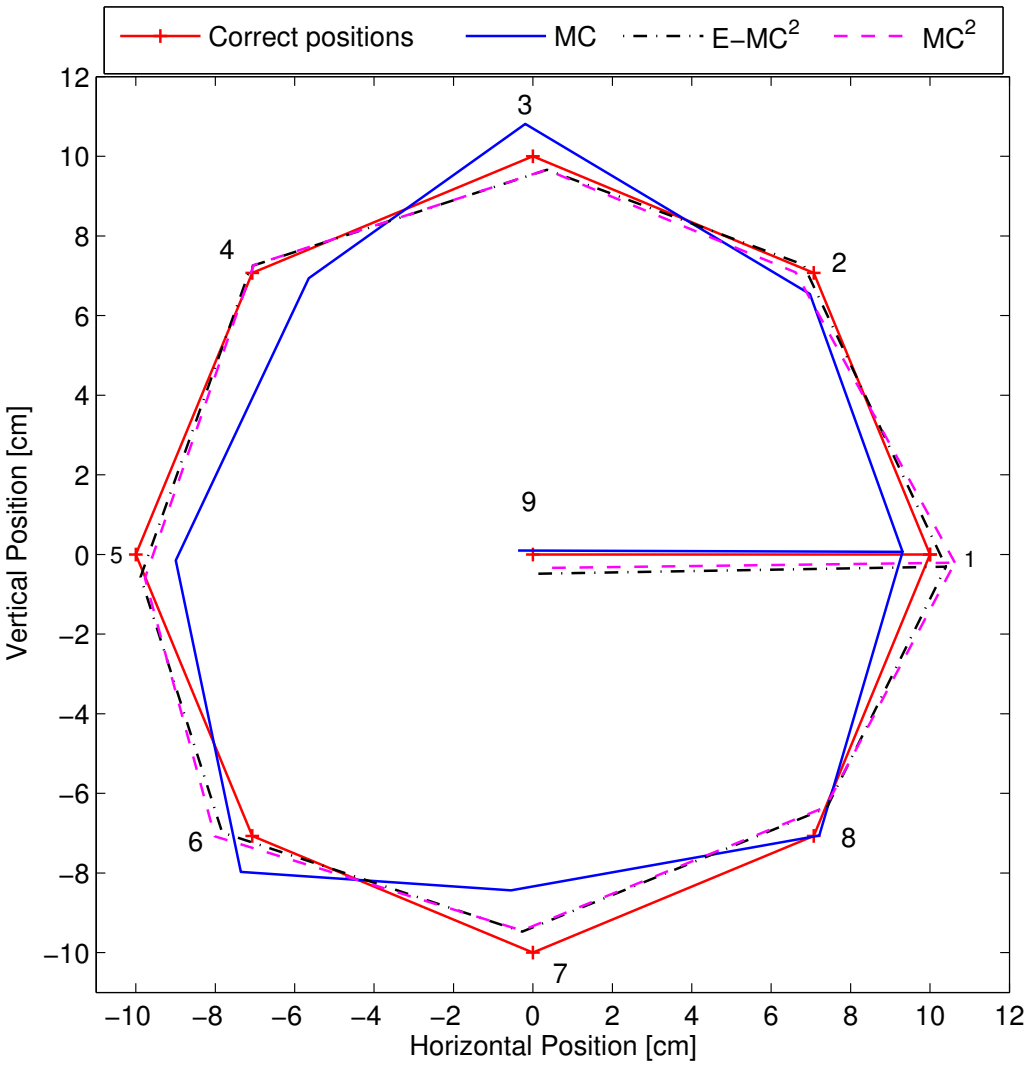

Figure 12: Calibration of the nine-element microphone array. The geometries are estimated using MC, MC+Cadzow (MC^2) and the proposed algorithm E-MC^2.

Table 3: Calibration errors (cm^2) as defined in (15) for different methods of microphone array calibration.

	Known		**Missing**	
	8-mic	**9-mic**	**11-mic**	**12-mic**
MDS-MAP	9	8.13	434.4	472
SDP	9.09	8.63	141	135
S-Stress	6.86	6.14	125	95
MC	10.6	9.75	133	115
MC^2	9.2	7.68	119	52
E-MC^2	6.5	5.85	49.6	46

Table 4: Position estimation errors (cm) as defined in (59) for different methods of microphone array calibration.

	Known		**Missing**	
	8-mic	**9-mic**	**11-mic**	**12-mic**
MDS-MAP	0.83	0.78	6.34	7.23
SDP	0.86	0.81	2.88	2.35
S-Stress	0.69	0.61	2.5	1.9
MC	1.1	0.97	2.6	2.1
MC^2	0.91	0.74	2.16	1.7
E-MC^2	0.64	0.97	1.06	1

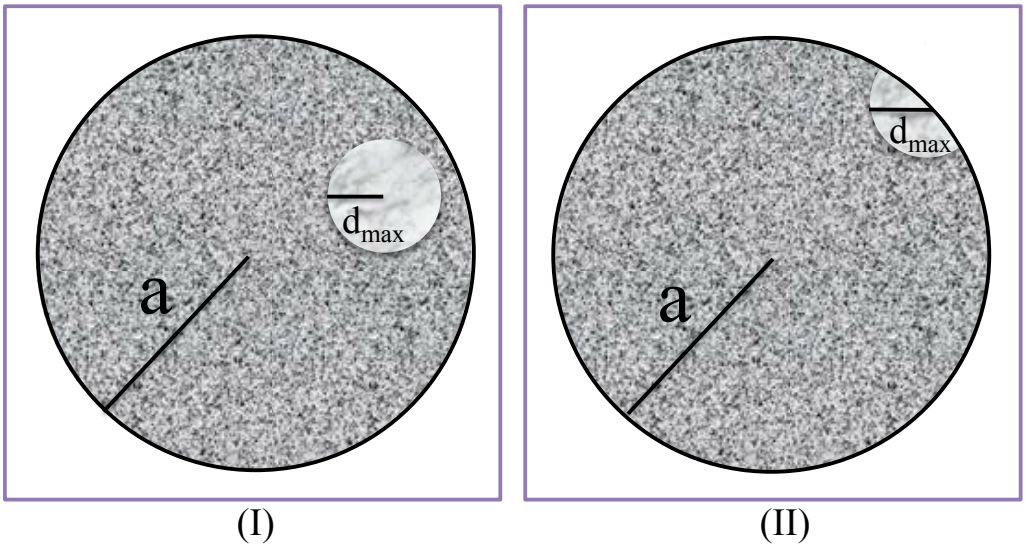

Figure 13: Scenario corresponding to the (I) lower bound and (II) upper bound of the probability q of structured missing distances.