Localizing Unsynchronized Sensors With Unknown Sources

Dalia El Badawy , Viktor Larsson , Marc Pollefeys, *Fellow, IEEE*, and Ivan Dokmanić

Abstract—We propose a method for sensor array self-localization using a set of sources at unknown locations. The sources produce signals whose times of arrival are registered at the sensors. We look at the general case where neither the emission times of the sources nor the reference time frames of the receivers are known. Unlike previous work, our method directly recovers the array geometry, instead of first estimating the timing information. The key component is a new loss function which is insensitive to the unknown timings. We cast the problem as a minimization of a non-convex functional of the Euclidean distance matrix of microphones and sources subject to certain non-convex constraints. After convexification, we obtain a semidefinite relaxation which gives an approximate solution; subsequent refinement on the proposed loss via the Levenberg-Marquardt scheme gives the final locations. Our method achieves state-of-the-art performance in terms of reconstruction accuracy, speed, and the ability to work with a small number of sources and receivers. It can also handle missing measurements and exploit prior geometric and temporal knowledge, for example if either the receiver offsets or the emission times are known, or if the array contains compact subarrays with known geometry.

Index Terms—Localization, calibration, array signal processing, wireless sensor networks, convex optimzation.

I. INTRODUCTION

IN MIMO radar [1], ultrasound imaging [2], underwater acoustics [3], time-reversal [4], [5], and room acoustics [6], [7] a collection of sources emit signals that are then captured by the receivers. In these applications, we often need an accurate knowledge of the geometry of the source and receiver positions to proceed with common array processing algorithms such as beamforming and source localization. Knowing the sensor array geometry similarly enables the reconstruction of physical fields in environmental monitoring using ad hoc sensor networks [8], [9]. Further, knowing device locations enables a host of location-based services in the context of the Internet of Things [10]. The recurring quintessential problem is thus to efficiently estimate the geometry of a set of nodes.

We study estimating the geometry of the nodes from the times of arrival (TOAs) of the source signals. The setup is illustrated in Fig. 1 where receivers register TOAs of source events. Some events can be (near-)collocated with the nodes, such is the case with transceivers. If the sources are anchors with known positions, locating the nodes becomes an exercise in geometry: intersecting spheres or hyperboloids depending on whether or not the devices are synchronized. However, the most general and practically appealing setup is when the sources are at arbitrary, unknown locations. This enables one to use signals of opportunity such as speech, transient sounds, or radio signals [11]. But it also means that receivers are no longer synchronized with the sources. To complicate things further, receivers themselves do not have to be mutually synchronized. While in many audio applications, microphones are connected to a common interface, we are also surrounded by ad hoc networks of smartphones and voice-based assistants. Thus, not only do we measure times of arrival as opposed to absolute times of flight, but those registered times of arrival further depend on the unknown time reference frame of each receiver. In this paper, we introduce an algorithm for jointly localizing a set of receivers and a set of sources from measured TOAs in the general case when all nodes are not synchronized: sources go off at unknown times and the reference time frame of each receiver can be different and unknown.

A. Notation

We denote matrices with capital letters ($\boldsymbol{S}$, $\boldsymbol{R}$) and vectors with lowercase letters ($\boldsymbol{s}$, $\boldsymbol{r}$). Uppercase subscripts indicate the size, e.g., $\boldsymbol{J}_L$ is an $L \times L$ matrix, and $\boldsymbol{R}_{M \times K}$ is an $M \times K$ matrix. Lowercase subscripts d_{ij} indicate the element at the ith row and j^{th} column of a matrix. A superscript T as in $\boldsymbol{J}^T$ denotes the matrix transpose. Lowercase superscripts $x^{(k)}$ denote the kth iteration. The real d-space is denoted $\mathbb{R}^d$. The vectorization operator is denoted $\text{vec}(x)$. We use the diagonalization operator where $\text{diag}(\boldsymbol{G})$ is a vector of the elements on the diagonal of the
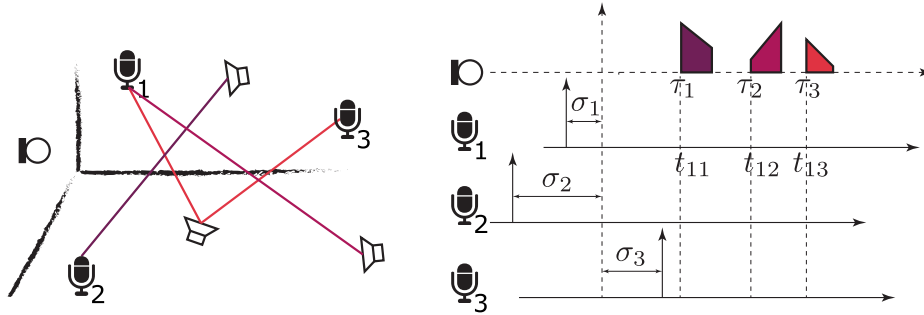
Fig. 1. The problem setup addressed in this paper. Source events (such as sounds in the environment produced by people or loudspeakers, or electromagnetic signals of opportunity) are captured by the receivers. In this illustration, there are three loudspeakers (sources) with emission times $\{\tau_k\}_{k=1}^3$. We also have three microphones (receivers) numbered 1, 2, 3 with internal times of reference $\{\sigma_m\}_{m=1}^3$ relative to a global reference (not numbered). The time of arrival t_{mk} at each microphone depends on the source's emission time τ_k and the microphone's offset σ_m. Not all source events reach all receivers, thus some entries t_{mk} might be missing.

TABLE I
OVERVIEW OF RELATED WORK WITH RESPECT TO THE SETUP

Approach	Input	Unknowns	Challenges
Multidimensional scaling (MDS) [17], [18], [19], [20]	Pairwise distances	-	Requires full synchronization
Multidimensional unfolding [21], [22], ML optimization [14]	Distances between nodes/events	-	Requires full synchronization, Bad local minima
SDP relaxations [23], [24], [25], [26], [27], [28], [29]	Distances/TDOA/FDOA	-	Requires anchor nodes or positions of the sensor nodes
Majorization [30], Two-stage [31], [32], [33], [34]	TDOA	Source or receiver offsets	Bad local minima, cannot handle near-minimal configurations
Two-stage [16]	TDOA	Source & receiver offsets	Slow, cannot handle near-minimal configurations
Proposed	TOA/TDOA	Source & receiver offsets	-

matrix. Similarly, $\mathrm{diag}(\sigma_1, \ldots, \sigma_N)$ is an $N \times N$ matrix with $\sigma_1, \ldots, \sigma_N$ on the diagonal.

B. Related Work

Among the earliest work on self-localization are the papers of Rockah and Schultheiss [12], [13] and Weiss and Friedlander [14] from the 1980 s. Plinge et al. [15] provide an in-depth overview of microphone array localization techniques. A systematization of existing approaches to self-localization is also given by Wang et al. [16]. In the following, we review some of the major points, related to localization in different setups. An overview is also shown in Table I.

In some cases, the pairwise distance between all the nodes can be estimated. This happens for example when the nodes can both send and receive [35] or by measuring the diffuse noise coherence [36]. Localization then amounts to multidimensional scaling (MDS) [17], [18], [19], [20].

A more common situation in audio applications is that the nodes can only receive or only send. The "sending" nodes need not be real devices; they can be any acoustic events or signals of opportunity. We can distinguish the case when the sources and the receivers are synchronized so we can estimate the source–receiver distances, or the various cases when sources, receivers, or both sources and receivers are not synchronized.

a) Known Source Emission Times and Receiver Offsets: If the emission times happen to be known and the nodes and events all have a common time reference (that is, they are

synchronized),[1] then the TOAs correspond to times of flight and directly give the distances between the nodes and the events. Given these distances, the joint localization problem reduces to multidimensional unfolding [21], [22]. Some methods are based on direct optimization of the maximum likelihood (ML) criterion [14], but these often fail due to the non-convexity of the objective and the existence of bad local minima [16]. Other methods are based on Euclidean distance matrices and semidefinite programming (SDP) [37].

Crocco et al. [22] proceed by constructing a certain low-rank matrix of differences of squared TOAs from which the positions can be recovered by solving a low-dimensional non-convex optimization problem as follows. The low-rank matrix of differences of squared TOAs [22] is related to the (translated) source positions S and receiver positions R as $H \approx \bar{R}^T \bar{S}$, where the latter is the matrix of inner products between *centered* receiver points and *centered* source points. The estimated H is decomposed using singular value decomposition as $H = U\Sigma V^T$. Up to a translation due to centering, the source and receiver locations are then $\bar{S} = Q^{-1}\Sigma V^T$ and $\bar{R} = (UQ)^T$ for some unknown invertible 3×3 matrix Q. Since we have that $H = \bar{R}^T QQ^{-1}\bar{S}$, finding the correct relative geometry (between R and S) amounts to identifying the right Q and the difference vector between the center of R and the center of S which is done via non-convex optimization and can suffer from local optima, especially in the presence of noise. In our

---

[1] Additionally, the internal delays of the receivers are known.

approach, we avoid the need to estimate Q to *stitch* R and S together by working with the full point set $X = [R; S]$ and the corresponding Gram matrix, instead of splitting X into R and S. The Gram matrix and its constraints then automatically encode the relative geometry and *glue* together the sources and the receivers (see Section III-A).

b) When One Set of Times is Known: A more realistic scenario is when the source emission times are unknown and different but the microphones are synchronized. The reverse case is also relevant: when the receivers are not synchronized but the source emission times are known. This is the case if we use a distributed sensor array.

The two unknowns (geometry and one set of times) can be estimated by directly optimizing an objective, for example using majorization [30]. Again, this often fails due to the nonconvexity of the problem and cannot work with near-minimal configurations.

An alternative approach is to estimate the unknowns sequentially, first recovering the unknown times, and then using them to recover the geometry [31], [32]. Early work of Pollefeys and Nister [38] exploits the low rank of a certain matrix of squared TOA differences. Their work is a near-field generalization of the work of Thrun [33], which was also adapted to work with missing measurements [34]. Heusdens and Gaubitch propose a more robust scheme based on structured total-least-squares [39] to reconstruct the times.[2]

c) Unknown Emission Times and Receiver Offsets: The closest to our work is the method of Wang et al. [16] who treat the most general case with unknown source emission times and receiver offsets. Same as the methods that address the case of one unknown set of offsets [39], Wang et al. proceed in two stages. The timing information is estimated via structured least-squares (Gauss–Newton), noting that with the correct timing estimates a certain matrix computed from the measured T(D)OA measurements must become low rank. This timing estimation procedure may require numerous random restarts. Once the timings are estimated, they proceed as Crocco et al. [22] to recover the geometry.

d) Convex Relaxations: In the context of sensor network localization (with fixed anchor nodes), Biswas et al. [23], [24] proposed to use SDP relaxations [28], [29] to handle the nonconvexity introduced by the square root in the distance measurements. Similar relaxations have since also been used for source node localization from TDOA measurements (see e.g. Yang et al. [25], Vaghefi and Beuhrer [26]) and mixed TDOA/FDOA measurements (Wang et al. [27]). Yang et al. [25] also proposed a SOCP relaxation; however, it requires that the target node lies within the convex hull of the anchors. Jiang et al. [40] propose to use Truncated Nuclear Norm Regularization (TNNR) to solve the TDOA self-calibration problem. Their optimization scheme consists of solving a sequence of convex surrogate problems based on the (convex) nuclear norm. We propose SDP relaxations for the full self-calibration problem, where both senders are receivers are unknown and not synchronized.

e) Minimal Estimation Problems: A series of work considers minimal problems in TOA, where the goal is to estimate the unknowns from as few measurements as possible. Following up on their work on minimal problems for TOA measurements [41], Kuang et al. [42] propose a stratified approach for estimating the unknown time offsets from TDOA measurements. Once the offsets are recovered, the solvers from their previous work [41] can be applied to recover the sender and receiver positions. Zhayida et al. [43] (and later Farmani et al. [44]) considered the minimal solutions to the special case of dual microphone setups. Burgess et al. [45] proposed solutions for settings where either the sender or the receivers lie in a lower dimensional space. Batstone et al. [46] consider the case of constant offset TDOA self-calibration (i.e. transmissions have a known period but unknown offset); this is also a stratified approach which first solves for the unknown time offset [42]. We show that our approach can also succeed in near-minimal configurations.

C. Contributions

Our work is motivated by the fact that the two-stage procedure is suboptimal. The (valid) reason to adopt sequential estimation are poor local minima in the loss function which involves both the times and the positions. From a statistical point of view, however, joint estimation is preferred. Further, timing estimation can fail or require many random restarts; the same holds for the non-convex optimization to recover the relative configurations from $\bar{R}^T \bar{S}$. The two-stage procedure also makes it challenging to exploit prior geometric information such us known distances or distance bounds.

In light of related work and the above discussions, we summarize our contributions as follows:

- What we are usually interested in are the sensor positions. We thus formulate a new self-localization loss which is invariant to offsets and delays. The proposed loss uses non-squared distances as opposed to the usual squared; we prove that minimizing this objective yields the maximum likelihood estimate of the positions under Gaussian noise. It enables us to recover the geometry without worrying about the times and without random restarts.
- We formulate a non-convex semidefinite optimization problem in terms of this new loss. We work with the full point set $X = [R; S]$ and the corresponding Gram matrix. The Gram matrix and its constraints intrinsically *glue* together the sources and the receivers and thus obviate the need for the non-convex optimization step of Crocco et al. [22] (see Section III-A).
- Although our method does not require synchronization, it can leverage synchronization among the receivers or the sources if it is fully or partially present [39], [47], [48]. It can further leverage geometric side information such as when the receivers contain subarrays with known relative geometry or some sources and receivers are collocated. Finally, it can handle missing measurements.

---

[2]Heusdens and Gaubitch in fact address the case when the microphone delays are unknown and the source is periodic; this is equivalent to only having unknown microphone delays; see Section IV-C.

The proposed method achieves state-of-the art results, not only in terms of localization accuracy, runtime, or exploiting side information, but also in the ability to solve near-minimal problems with few nodes. Finally, we note that as suggested by Ono [30], our method can also be interpreted as a virtual synchronization method.

II. PROBLEM STATEMENT

We wish to localize a set of M receivers with unknown positions $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_M \in \mathbb{R}^d$ using a set of K sources with unknown positions $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_K \in \mathbb{R}^d$. We assume that the sources emit pulses whose times of arrival (TOA) can be measured at the receivers.[3] Since we do not assume synchronization between the sources and the receivers, the absolute emission time τ_k of the kth source is unknown. Since the sources are not necessarily synchronized among themselves, the differences $\tau_k - \tau_{k'}$ are also unknown. Similarly, since we do not assume synchronization among receivers (nor knowing their internal delays), the temporal frames of reference of each receiver are shifted by an unknown σ_m with respect to some reference clock. The times of arrival of the kth source signal at the mth receiver thus become

$$t_{mk} = v^{-1} \|\boldsymbol{r}_m - \boldsymbol{s}_k\| + \sigma_m + \tau_k, \qquad (1)$$

where for simplicity we let both σ_m and τ_k have arbitrary sign. We assume the transmission speed v to be known and w.l.o.g. let $v = 1$ for the remainder of the paper.

Note that if we instead use the time-difference-of-arrival (TDOA), then with the first sensor as the reference, the TDOA is

$$\begin{aligned}
\tilde{t}_{mk} &= t_{mk} - t_{1k} \\
&= \|\boldsymbol{r}_m - \boldsymbol{s}_k\| + (\sigma_m - \sigma_1) - \|\boldsymbol{r}_1 - \boldsymbol{s}_k\| \\
&= \|\boldsymbol{r}_m - \boldsymbol{s}_k\| + \tilde{\sigma}_m + \tilde{\tau}_k.
\end{aligned}$$

And so we can still think of the TDOA $\tilde{t}_{mk}$ as a TOA but with modified emission times and offsets.

A. Minimal Number of Sources and Receivers

Common sensor localization scenarios are either in the horizontal plane ($d = 2$) or in 3D space ($d = 3$). The number of the degrees of freedom in the positions of sources and receivers is then $d(M + K)$. However, since rigid motions of the entire setup cannot be distinguished from TOA data, we can choose the associated $d(d + 1)/2$ degrees of freedom freely (d translational and $\binom{d}{2}$ rotational). Each source and receiver come with an additional unknown time, which gives another $M + K$ degrees of freedom. However, we can choose a global time offset arbitrarily, which is another gauge of invariance and subtracts one degree of freedom. The total number of the degrees of freedom is thus

$$\#(\text{DOF}) = (d + 1)\left(M + K - \tfrac{d}{2}\right) - 1$$

As measurements we get MK TOAs. In order to get a dimension-zero solution set (a solution set that is a finite or countable set of points, but not necessarily unique), this number should at least match the number of the degrees of freedom. Solving for the minimal number of sources as a function of the number of receivers, we obtain the following relation

$$K \geq \frac{(d + 1)M - d(d + 1)/2 - 1}{M - (d + 1)}.$$

For $d = 3$ this gives $K \geq (4M - 7)/(M - 4)$ implying that the smallest number of receivers that can be localized is $M = 5$ for which at least $K = 13$ sources are required to get a dimension-zero solution. Some other minimal cases are: $(M = 7, K = 7), (M = 13, K = 5)$.

As we demonstrate with real and numerical experiments in Section VI, our algorithms can operate in the vicinity of this minimal regime, where previous methods either fail or perform poorly.

III. A LOSS INSENSITIVE TO OFFSETS AND DELAYS

Instead of proceeding sequentially by first estimating the unknown times as in the case of the state-of-the-art methods [16], [39], we note that the primary task is usually position estimation rather than timing estimation. Besides, when the positions are known, estimating the timings $\{\sigma_m\}_{m=1}^M$ and $\{\tau_k\}_{k=1}^K$ boils down to a simple algebraic exercise (see Section III-D). We thus devise a data fidelity metric which is insensitive to the unknown timings, yet is only minimized with the correct positions.

We begin by writing (1) in matrix form as

$$\boldsymbol{T} = \boldsymbol{\Delta} + \boldsymbol{\sigma}\boldsymbol{1}_K^T + \boldsymbol{1}_M\boldsymbol{\tau}^T, \qquad (2)$$

where $\boldsymbol{T} \in \mathbb{R}^{M \times K}$, $\boldsymbol{\Delta} = (\delta_{mk})$ and $\delta_{mk} = \|\boldsymbol{r}_m - \boldsymbol{s}_k\|$, $\boldsymbol{\sigma} = [\sigma_1, \ldots, \sigma_M]^T$ and $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_K]^T$. Now we make the key observation: multiplying (2) on both sides by a matrix which has an all-ones vector in the nullspace will annihilate the two terms that depend on the unknown times.

Given an integer L, let $\boldsymbol{J}_L$ be a geometric centering matrix of size L,[4]

$$\boldsymbol{J}_L := \boldsymbol{I}_L - \tfrac{1}{L}\boldsymbol{1}_L\boldsymbol{1}_L^T. \qquad (3)$$

It is easy to verify that $\boldsymbol{J}_M\boldsymbol{1}_M = \boldsymbol{0}_M$ and $\boldsymbol{1}_K^T\boldsymbol{J}_K = \boldsymbol{0}_K^T$ so that the matrix

$$\boldsymbol{P} = \boldsymbol{J}_M\boldsymbol{T}\boldsymbol{J}_K = \boldsymbol{J}_M\boldsymbol{\Delta}\boldsymbol{J}_K \qquad (4)$$

does not depend on the unknown times. In fact, we can say more.

Proposition 1: We have $\boldsymbol{J}_M\boldsymbol{T}_1\boldsymbol{J}_K = \boldsymbol{J}_M\boldsymbol{T}_2\boldsymbol{J}_K$ if and only if there exist $\boldsymbol{\sigma}'$ and $\boldsymbol{\tau}'$ such that $\boldsymbol{T}_1 = \boldsymbol{T}_2 + \boldsymbol{\sigma}'\boldsymbol{1}_K^T + \boldsymbol{1}_M\boldsymbol{\tau}'^T$. In particular, for $\boldsymbol{T}$ as in (2), we have $\boldsymbol{J}_M\boldsymbol{T}\boldsymbol{J}_K = \boldsymbol{J}_M\boldsymbol{\Delta}\boldsymbol{J}_K$ (4).

Remark: If $\boldsymbol{\Delta}$ happened to be a Euclidean (squared) distance matrix (EDM) of a point set $\boldsymbol{X}$, for example $\boldsymbol{D} = (d_{ij}^2)$, $d_{ij} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$, then $-\frac{1}{2}\boldsymbol{J}_N\boldsymbol{D}\boldsymbol{J}_N$ would equal the Gram matrix of the (centered) $\boldsymbol{X}$. No similar statement can be made here

---

[3]Note that these measurements can be obtained without physically emitting pulses. In audio, for example, one often emits long chirps which is followed by pulse compression.

[4]We often indicate matrix and vector sizes in subscripts. While making the notation a bit clunkier, it helps keep track of dimensions.

simply because $\boldsymbol{\Delta}$ holds non-squared distances, and because it only corresponds to one off-diagonal block of $\boldsymbol{D}$.

Proof: The nullspace of the linear operator

$$\mathcal{J} : \mathbb{R}^{M \times K} \to \mathbb{R}^{M \times K}$$

$$\boldsymbol{A} \overset{\mathcal{J}}{\mapsto} \boldsymbol{J}_M \boldsymbol{A} \boldsymbol{J}_K$$

is spanned by

$$\mathcal{N}(\mathcal{J}) = \mathrm{span}\{\mathbf{1}_M \boldsymbol{e}_1^\top, \dots, \mathbf{1}_M \boldsymbol{e}_K^T, \boldsymbol{e}_1 \mathbf{1}_K^T, \dots \boldsymbol{e}_M \mathbf{1}_K^T\}.$$

Note that not all of those MK matrices are linearly independent, but the argument does not hinge on independence. Every matrix in the argument of span correspond to adding a constant offset to either a source or a receiver. As a consequence, any nullspace vector can be written as a linear combination of source / receiver offsets which proves the proposition. $\square$

What Proposition 1 says is that $\boldsymbol{J}_M \boldsymbol{\Delta} \boldsymbol{J}_K$ is sufficient to determine the positions in the sense that the only locations such that this transformation matches the measurements are the true ones (if the original problem is solvable).

We can then propose the following timing-invariant objective for localization

$$\min_{\boldsymbol{R}, \boldsymbol{S}} \frac{1}{2} \|\boldsymbol{J}_M (\boldsymbol{\Delta}(\boldsymbol{R}, \boldsymbol{S}) - \boldsymbol{T}) \boldsymbol{J}_K\|_F^2, \tag{5}$$

where $\| \cdot \|_F$ denotes the Frobenius norm, $\boldsymbol{R} \in \mathbb{R}^{d \times M}$ are the receiver positions, and $\boldsymbol{S} \in \mathbb{R}^{d \times K}$ are the source positions. This objective is justified by the following direct corollary of Proposition 1.

Corollary 1: If the unsynchronized localization problem has a unique solution (up to a rigid transformation), then vanishing loss in (5) implies correct localization.

Further, the loss (5) is in fact equivalent to the ML loss *given the offsets* $\boldsymbol{\sigma}$ *and* $\boldsymbol{\tau}$. Namely, assuming i.i.d. Gaussian noise on the TOA measurements we have

$$f_{ML}(\boldsymbol{\Delta}, \boldsymbol{\sigma}, \boldsymbol{\tau}) = \|\boldsymbol{\Delta} + \boldsymbol{\sigma} \mathbf{1}_K^T + \mathbf{1}_M \boldsymbol{\tau}^T - \boldsymbol{T}\|_F^2,$$

and the loss in (5) is simply

$$f(\boldsymbol{\Delta}) = \min_{\boldsymbol{\sigma}, \boldsymbol{\tau}} f_{ML}(\boldsymbol{\Delta}, \boldsymbol{\sigma}, \boldsymbol{\tau}).$$

To see this note that

$$\nabla_{\boldsymbol{\sigma}} f_{ML} = 2 \left(\boldsymbol{\Delta} + \boldsymbol{\sigma} \mathbf{1}_K^T + \mathbf{1}_M \boldsymbol{\tau}^T - \boldsymbol{T}\right) \mathbf{1}_K = 0 \tag{6}$$

$$\implies \boldsymbol{\sigma}^\star = -\frac{1}{\mathbf{1}_K^T \mathbf{1}_K} \left(\boldsymbol{\Delta} + \mathbf{1}_M \boldsymbol{\tau}^T - \boldsymbol{T}\right) \mathbf{1}_K. \tag{7}$$

Inserting into the original cost we get

$$f_{ML}(\boldsymbol{\Delta}, \boldsymbol{\sigma}^\star, \boldsymbol{\tau}) = \|(\boldsymbol{\Delta} + \mathbf{1}_M \boldsymbol{\tau}^T - T) \underbrace{\left(\boldsymbol{I}_K - \frac{1}{K} \mathbf{1}_K \mathbf{1}_K^T\right)}_{\boldsymbol{J}_K}\|_F^2.$$

Solving for $\boldsymbol{\tau}^\star$ similarly yields $f_{ML}(\boldsymbol{\Delta}, \boldsymbol{\sigma}^\star, \boldsymbol{\tau}^\star) = f(\boldsymbol{\Delta})$.

A. Characterization in Terms of the Full Gram Matrix

As described in Section I-B, a number of state-of-the-art methods based on low-rank matrix decompositions recover the matrix $\boldsymbol{H} = \bar{\boldsymbol{R}}^T \bar{\boldsymbol{S}}$, with $\bar{\boldsymbol{R}}$ and $\bar{\boldsymbol{S}}$ being some translated (centered)

versions of $\boldsymbol{R}$ and $\boldsymbol{S}$ [16], [22], [39]. Matrix factorizations such as singular value decomposition can only determine $\bar{\boldsymbol{R}}$ and $\bar{\boldsymbol{S}}$ up to a multiplication by an arbitrary invertible matrix $\boldsymbol{Q}$, since for any such $\boldsymbol{Q}$ we have that $\boldsymbol{H} = \widehat{\boldsymbol{R}}^T \widehat{\boldsymbol{S}}$, with $\widehat{\boldsymbol{R}} = \boldsymbol{Q}^T \boldsymbol{R}$, $\widehat{\boldsymbol{S}} = \boldsymbol{Q}^{-1} \boldsymbol{S}$. Finding the right $\boldsymbol{Q}$ (which can be assumed to be upper-triangular which fixes the rotational gauge freedom) and the (one) translation vector is then achieved via direct non-convex optimization, though over a lower dimensional space than the original problem.

We sidestep this problem of having to determine $\boldsymbol{Q}$ in order to *stitch* together $\boldsymbol{R}$ and $\boldsymbol{S}$ by writing the measurements in terms of the full point set

$$\boldsymbol{X} := [\boldsymbol{R}, \ \boldsymbol{S}] \in \mathbb{R}^{d \times N},$$

where $N = M + K$, and always working with the entire $\boldsymbol{X}$. We denote the corresponding full Gram matrix by $\boldsymbol{G} = \boldsymbol{X}^T \boldsymbol{X} \in \mathbb{R}^{N \times N}$. The off-diagonal blocks of the Gram matrix $\boldsymbol{G}$ encode the relative geometry between $\boldsymbol{R}$ and $\boldsymbol{S}$—by recovering $\boldsymbol{G}$, we know that relative geometry. Note that this only resolves the relative ambiguity between $\boldsymbol{R}$ and $\boldsymbol{S}$ which is an artifact of the previous methods. The inherent invariance of localization from distances to Euclidean motions remains.

The matrix of squared distances (the EDM) between the column vectors in $\boldsymbol{X}$ can be written as a linear function of the Gram matrix. Letting $\boldsymbol{D} = (\|\boldsymbol{x}_n - \boldsymbol{x}_{n'}\|^2)_{n,n'=1}^N$, one has

$$\boldsymbol{D} = \mathcal{D}(\boldsymbol{G}) := \mathrm{diag}(\boldsymbol{G}) \mathbf{1}_N^T - 2\boldsymbol{G} + \mathbf{1}_N \mathrm{diag}(\boldsymbol{G}), \tag{8}$$

with $\mathrm{diag}(\boldsymbol{G}) = [g_{11}, \dots, g_{NN}]^T$. This means that the matrix of distances $\boldsymbol{\Delta}$ between sources and receivers can be written as an entrywise square root of a linear function of $\boldsymbol{G}$, since

$$\boldsymbol{\Delta}^{\bullet 2} = \mathcal{L}(\boldsymbol{G}) \overset{\text{def}}{=} \boldsymbol{S}_{\text{row}} \mathcal{D}(\boldsymbol{G}) \, \boldsymbol{S}_{\text{col}}, \tag{9}$$

where $\boldsymbol{S}_{\text{row}} := [\boldsymbol{I}_M, \ \mathbf{0}_{M \times K}]$ and $\boldsymbol{S}_{\text{col}} := [\mathbf{0}_{K \times M}, \ \boldsymbol{I}_K]^T$ are the appropriate row and column selection matrices, and $(\cdot)^{\bullet 2}$ denotes an entrywise square. In view of Proposition 1, this leads us to the first reformulation of the original problem:

$$\min_{\boldsymbol{G}} \quad \left\| \boldsymbol{J}_M \left(\sqrt[\bullet]{\mathcal{L}(\boldsymbol{G})} - \boldsymbol{T} \right) \boldsymbol{J}_K \right\|_F^2 \tag{10a}$$

$$\text{subject to} \quad \boldsymbol{G} \succeq 0, \tag{10b}$$

$$\boldsymbol{G} \mathbf{1}_N = \boldsymbol{0}, \tag{10c}$$

$$\mathrm{rank}(\boldsymbol{G}) = d. \tag{10d}$$

The constraint (10c) resolves the translation ambiguity; it is equivalent to the point set being centered around the origin $\frac{1}{N} \sum_n \boldsymbol{x}_n = 0$.

Note that now the receivers and sources are being translated together, unlike in earlier work where they were split. Further, the estimated locations are directly read out from a factorization of $\boldsymbol{G}$, without the need for stitching.

B. Semidefinite Relaxation

The formulation (10a) remains nonconvex. Thus, direct approaches such as randomly-initialized first-order methods (e.g., projected gradient descent) are likely to get stuck in a local

minimum. We thus propose to instead solve a convex relaxation of the problem.

There are two sources of nonconvexity: the entrywise square root in the objective, and the rank constraint on the Gram matrix. We begin by replacing the entrywise square root $\sqrt[\bullet]{\mathcal{L}(\boldsymbol{G})}$ by a new matrix variable $\boldsymbol{B}$, and adding the constraint that the entrywise square of $\boldsymbol{B}$ must equal $\mathcal{L}(\boldsymbol{G})$.

Concretely, we write the objective as

$$\min_{\boldsymbol{G},\boldsymbol{B}} \; \|\boldsymbol{J}_M(\boldsymbol{B}-\boldsymbol{T})\boldsymbol{J}_K\|_F^2 ,$$

and ask that $\boldsymbol{B}^{\bullet 2} = \mathcal{L}(\boldsymbol{G})$ to obtain an equivalent formulation. We now proceed by reformulating the added constraint as a rank constraint on some semidefinite matrix (or matrices). Since the entrywise square does not mix the entries of $\boldsymbol{B}$, we can add such constraints on a per-entry basis. This is computationally more efficient (and empirically works better) than working with the entire vectorization of $\boldsymbol{B}$ at once.

We can equivalently write the constraint $b_{mk}^2 = \mathcal{L}(\boldsymbol{G})_{mn}$ as

$$\boldsymbol{L}_{mk} := \begin{bmatrix} \mathcal{L}(\boldsymbol{G})_{mk} & b_{mk} \\ b_{mk} & 1 \end{bmatrix} \succeq \boldsymbol{0} \tag{11}$$

$$b_{mk} \geq 0 \tag{12}$$

$$\mathrm{rank}(\boldsymbol{L}_{mk}) = 1. \tag{13}$$

All the non-convexity is now lumped into the rank constraints, $\mathrm{rank}(\boldsymbol{G}) = d$ and $\mathrm{rank}(\boldsymbol{L}_{mk}) = 1$ for $1 \leq m, k \leq M, K$. The final step is to relax all of the rank constraints to get

$$\min_{\boldsymbol{G},\boldsymbol{B}} \quad \|\boldsymbol{J}_M(\boldsymbol{B}-\boldsymbol{T})\boldsymbol{J}_K\|_F^2 \tag{14a}$$

$$\text{subject to} \quad \boldsymbol{G} \succeq 0, \tag{14b}$$

$$\boldsymbol{G}\boldsymbol{1}_N = \boldsymbol{0}, \tag{14c}$$

$$\boldsymbol{L}_{mk} \succeq \boldsymbol{0}, \text{ for all } 1 \leq m, n \leq M, K, \tag{14d}$$

$$\boldsymbol{B} \geq 0. \tag{14e}$$

We can finally reconstruct the point set from the recovered $\boldsymbol{G}$ using singular value decomposition. We have $\boldsymbol{G} = \boldsymbol{U}\Lambda\boldsymbol{V}^\top$, where $\Lambda = \mathrm{diag}(\sigma_1, \ldots, \sigma_N)$ with the singular values σ_i assumed to be sorted in decreasing order. We reconstruct the point set using $\hat{\boldsymbol{X}} = [\mathrm{diag}(\sigma_1, \ldots, \sigma_d), \boldsymbol{0}_{d \times (N-d)}]\boldsymbol{V}^\top$. If the rank of the estimated $\boldsymbol{G}$ is truly d, as it ideally should be since it describes a d-dimensional point set, then the trailing singular values would be zero anyway.

C. Refinement by Levenberg-Marquardt

In general, solving the above semidefinite program gives point configurations which are good coarse approximations to the true geometry, although they are not good enough for most applications. The reason can be tracked down to the fact that after relaxing the rank constraints, the estimated matrices have higher rank than desired. Still, the positive semidefinite constraints constrain geometry sufficiently to get decent estimates in many situations. One strategy would be to employ various rank minimization strategies. Informally, we tried this and it works decently but it is very slow.

Instead, we propose to refine the output of the SDR using the Levenberg-Marquardt (LM) algorithm. It is fast, accurate, and it works better. We have also empirically observed that simple gradient descent is *much* slower requiring more iterations to converge.

To derive the LM updates, we need to compute the Jacobian of the loss in order to linearize it. The loss

$$\mathcal{J}(\boldsymbol{R},\boldsymbol{S}) = \|\boldsymbol{J}_M(\boldsymbol{\Delta}(\boldsymbol{R},\boldsymbol{S}) - \boldsymbol{T})\boldsymbol{J}_K\|_F^2 ,$$

can be written as

$$\mathcal{J}(\boldsymbol{R},\boldsymbol{S}) = \|f(\boldsymbol{R},\boldsymbol{S})\|_2^2 ,$$

where $f(\boldsymbol{R},\boldsymbol{S}) = \boldsymbol{A}\boldsymbol{\delta}(\boldsymbol{R},\boldsymbol{S}) - \boldsymbol{t}$,

$$\boldsymbol{A} = \boldsymbol{J}_K^T \otimes \boldsymbol{J}_M, \quad \boldsymbol{t} = \boldsymbol{A}\,\mathrm{vec}(\boldsymbol{T}),$$

and $\boldsymbol{\delta} = \mathrm{vec}(\boldsymbol{\Delta})$ with $\otimes$ denoting the Kronecker product. Since $\boldsymbol{\Delta} : \mathbb{R}^{d \times (M+K)} \to \mathbb{R}^{M \times K}$, the Jacobian is a tensor in $\mathbb{R}^{M \times K \times d \times (M+K)}$. We can compute

$$[\mathsf{D}\boldsymbol{\Delta}]_{m,k,:,\ell} = \begin{cases} \dfrac{\boldsymbol{r}_m - \boldsymbol{s}_k}{\|\boldsymbol{r}_m - \boldsymbol{s}_k\|} & \ell = m, \\[2mm] -\dfrac{\boldsymbol{r}_m - \boldsymbol{s}_k}{\|\boldsymbol{r}_m - \boldsymbol{s}_k\|} & \ell = M + k, \\[2mm] 0 & \text{otherwise.} \end{cases}$$

with the understanding that the first M slices in the last index of D correspond to the receivers and the last K to the sources. Note that the Jacobian is not well defined at $\boldsymbol{r}_m = \boldsymbol{s}_k$. In that case, it would be possible to remove the corresponding variables from the optimization and adjust the Jacobian accordingly.

To make things easier to compute with, we rewrite $\boldsymbol{T}$ as a function of a single vector

$$\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{R},\boldsymbol{S}) = \begin{bmatrix} \mathrm{vec}(\boldsymbol{R}) \\ \mathrm{vec}(\boldsymbol{S}) \end{bmatrix},$$

and rearrange the output into a vector so that $\boldsymbol{\delta}(\boldsymbol{x}) = \mathrm{vec}(\boldsymbol{\Delta}(\boldsymbol{R},\boldsymbol{S}))$. The Jacobian $\mathsf{D}\boldsymbol{\delta}$ is obtained from $\mathsf{D}\boldsymbol{T}$ by collapsing the first two dimensions and the last two dimensions so that $\mathsf{D}\boldsymbol{\delta} \in \mathbb{R}^{MK \times d(M+K)}$. Finally the Jacobian of f is computed as

$$\mathsf{D}f = \boldsymbol{A}\,\mathsf{D}\boldsymbol{\delta},$$

and we have an affine approximation of f as (with some abuse of notation)

$$f(\boldsymbol{x};\boldsymbol{x}_0) = f(\boldsymbol{x}_0) + \mathsf{D}f(\boldsymbol{x})(\boldsymbol{x} - \boldsymbol{x}_0).$$

This leads to the LM update,

$$\boldsymbol{x}^{(k+1)} = \underset{\boldsymbol{z}}{\mathrm{argmin}} \; \left\|f(\boldsymbol{z};\boldsymbol{x}^{(k)})\right\|^2 + \lambda \left\|\boldsymbol{z} - \boldsymbol{x}^{(k)}\right\|^2$$

which is solved by

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - (\mathsf{D}f(\boldsymbol{x}^k)^T \mathsf{D}f(\boldsymbol{x}^k) + \lambda\boldsymbol{I})^{-1}\mathsf{D}f(\boldsymbol{x}^k)^T f(\boldsymbol{x}^{(k)}).$$

The full localization algorithm is summarized in Algorithm 1.

We use the SeDuMi solver [49] for the optimization problem (14). It has a $O(\sqrt{n}\log\frac{1}{\epsilon})$ worst-case time complexity per iteration where $n = d(M + K)$ is the number of variables. The LM refinement is $O(d^2(M + K)^2 MK)$ per iteration.

Algorithm 1: Unsynchronized Sensor Localization.

Require: TOA $T \in \mathbb{R}^{M \times K}$

Ensure: Positions of receivers $R \in \mathbb{R}^{d \times M}$ and sources $S \in \mathbb{R}^{d \times K}$

Initialize R and S using SDR (14).

Update R and S using LM.

D. Estimating Offsets and Delays

Once an estimate $\widehat{\boldsymbol{\Delta}}$ of $\boldsymbol{\Delta}$ is computed, we can also compute the unknown offsets $\boldsymbol{\tau}$ and delays $\boldsymbol{\sigma}$. We start by noting from (2) that

$$T - \boldsymbol{\Delta} = \boldsymbol{\sigma} \mathbf{1}_K^T + \mathbf{1}_M \boldsymbol{\tau}^T,$$

where the right hand side is linear in $(\boldsymbol{\sigma}, \boldsymbol{\tau})$. We exploit it by vectorizing both sides:

$$
\begin{aligned}
e := \mathsf{vec}(T - \boldsymbol{\Delta}) &= \mathbf{1}_K \otimes \boldsymbol{\sigma} + \boldsymbol{\tau} \otimes \mathbf{1}_M \\
&= V\boldsymbol{\sigma} + W\boldsymbol{\tau} \\
&= [V \quad W] \begin{bmatrix} \boldsymbol{\sigma} \\ \boldsymbol{\tau} \end{bmatrix},
\end{aligned}
$$

with $V := \mathbf{1}_K \otimes I_M$, $W := I_K \otimes \mathbf{1}_M$. From here, estimating the times vector amounts to solving an overdetermined system of linear equations. Note that the system matrix has a nullspace of dimension one (spanned by $\alpha \mathbf{1}$) due to the global offset ambiguity. To eliminate the offset ambiguity we arbitrarily set $\sigma_1 = 0$, and let V' be V with the first column removed. The least-squares estimate of the unknown timings is then given as

$$
\begin{bmatrix} \widehat{\boldsymbol{\sigma}}' \\ \widehat{\boldsymbol{\tau}} \end{bmatrix} = [V' \quad W]^\dagger e, \tag{15}
$$

with $\widehat{\boldsymbol{\sigma}}' = [\widehat{\sigma}_2, \ldots, \widehat{\sigma}_M]^T$ and $(\,\cdot\,)^\dagger$ denoting the Moore-Penrose pseudoinverse.

IV. VARIANTS

In many cases, we have some prior information about the distances or offset times. The proposed method makes it straightforward to leverage this prior knowledge. In the following, we explain how by describing several typical scenarios.

A. Localization of Subarrays

Suppose that some distances are known. It is often the case that the set of receivers consists of several compact subarrays with known geometry (voice-based assistants, smart phones). These compact subarrays are then distributed in the space of interest at unknown positions and with random orientations, together with discrete microphones.

Since the EDM is a linear function of the Gram matrix, this prior knowledge corresponds to simple linear constraints on G. We thus augment the original program with a set of *known* inter-sensor distances for microphone pairs

$$\mathcal{M} = \{(m_1, m_2) \,:\, m_1 < m_2,$$

$$\text{distance } d_{m_1 m_2} \text{ between } \boldsymbol{r}_{m_1} \text{ and } \boldsymbol{r}_{m_2} \text{ is known}\},$$

Then we add the following set of constraints to our optimization:

$$\mathcal{D}(G)_{m_1 m_2} = d_{m_1 m_2}^2, \text{ for all } (m_1, m_2) \in \mathcal{M}, m_1 \neq m_2. \tag{16}$$

If the set of receivers is partitioned as

$$R = [R_1, R_2, \ldots, R_J],$$

with the distances within the jth subarray R_j known, these constraints correspond to knowing the diagonal subblocks of the upper-left block of $\mathcal{D}$.

We can again proceed with an LM refinement step. In this case, we consider the following objective

$$\min_\theta \quad \|f(\theta)\|^2 \tag{17a}$$

$$\text{subject to} \quad g(\theta) = 0, \tag{17b}$$

$$\tag{17c}$$

where $g : \mathbb{R}^{d(M+K)} \mapsto \mathbb{R}^{|\mathcal{M}|}$ is the known-distances residual defined as

$$g_{m_1 m_2}(X) = \mathsf{edm}(X, X)_{m_1 m_2} - d_{m_1 m_2}^2$$

with $\mathsf{edm}(X, X)_{m_1 m_2} = \|\boldsymbol{x}_{m_1} - \boldsymbol{x}_{m_2}\|^2$ being the entries of the EDM of the point set X. We solve (17) using the Augmented Lagrangian algorithm [50]. The corresponding update is given in Appendix A.

In some cases the distances may not be known exactly, but we might have access to good bounds. We once more leverage the linearity of $\mathcal{D}$ in the Gramian by adding the following linear constraints to the program,

$$\underline{d}_{m_1 m_2}^2 \leq \mathcal{D}(G)_{m_1 m_2} \leq \bar{d}_{m_1 m_2}^2. \tag{18}$$

where $\bar{d}_{m_1 m_2}$ and $\underline{d}_{m_1, m_2}$ are upper and lower bounds on the distance d_{m_1, m_2}.

Finally, we note that priors might be available not only for inter-receiver distances but also for inter-source distances and the distances between the sources and the receivers (we often have a coarse idea about how far the source events are from the microphones). These can be added to the constraints completely analogously.

B. When One Set of Times is Known

When one set of timings is known (either the source emission times or the receiver offsets), it would still be possible to use the general formulation with the derived loss (10a) designed for the case when both sets of times are unknown (2). Though that would require more measurements than necessary given such prior knowledge as we show next.

Both cases (unknown emission times or unknown offsets) are captured by the simplified measurement model

$$T = \boldsymbol{\Delta} + \mathbf{1}_M \boldsymbol{\tau}^T, \tag{19}$$

up to a transposition as necessary (exchanging the role of sources and receivers). From here, it follows that the influence of the timings can be eliminated simply by one left multiplication by J_M so that

$$P' = J_M T = J_M \boldsymbol{\Delta}$$

does not depend on $\boldsymbol{\tau}$. We can thus replace the loss (10a) by

$$\|\boldsymbol{J}_M(\boldsymbol{B}-\boldsymbol{T})\|_F^2. \tag{20}$$

To see how this reduces the minimal required number of measurements (sources or sensors), recall the degrees-of-freedom count from Section II-A. Concretely,

$$\#(\mathrm{DOF}) = d\left(M + K - \tfrac{d+1}{2}\right) + M$$

gives

$$K \geq \frac{(d+1)(2\,M-d)}{2(M-d)}.$$

For $d = 3$, this becomes $K \geq \frac{4M-6}{M-3}$, so that the minimal number of receivers is now $M = 4$ and the minimal cases for the dimension-zero solution are $M = 4, K = 10$, $M = 5, K = 7$, $M = 6, K = 6$, and $M = 9, K = 5$. Another way to see this is by noting that the operator $\boldsymbol{A} \mapsto \boldsymbol{J}_M \boldsymbol{A}$ has a smaller nullspace than $\boldsymbol{A} \mapsto \boldsymbol{J}_M \boldsymbol{A} \boldsymbol{J}_K$.

C. Sources With Known (Constant) Offset

As a last example of prior information that is easily handled, we mention the practically relevant case where the source is for example a smartphone emitting periodic pulses. More generally, the source is emitting signals at known (not necessarily regular) intervals. The phone is not synchronized with the receivers, so the emitting times are strictly speaking unknown, but only up to one unknown offset—the start time of the emissions. This case has been studied in previous work [39], [46].

We show how to incorporate this using our proposed method. We have that

$$\tau_k = \tau_o + \delta_k,$$

with δ_k is the known offset of the kth emission with respect to the unknown time τ_o. When the emissions are periodic, we have $\delta_k = k \cdot \delta$, but for simplicity we keep the notation more general.

Note that in general, the vectors $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ can never be recovered uniquely, unless one of the times is known. The reason for this is that the global time reference can be decided arbitrarily (we used this fact when counting the degrees of freedom in Section II-A).

It then follows that the case of sources with known offsets is mathematically equivalent to the case when only the receivers' offsets are unknown. Concretely, first write

$$\boldsymbol{\tau} = \tau_o \mathbf{1}_K + \boldsymbol{\delta},$$

where $\boldsymbol{\delta}$ is a known vector. Then note that

$$\boldsymbol{\sigma}\mathbf{1}_K^T + \mathbf{1}_M\boldsymbol{\tau}^T = (\boldsymbol{\sigma} + \tau_o\mathbf{1}_M)\mathbf{1}_K^T + \mathbf{1}_M\boldsymbol{\delta}^\top.$$

Since $\mathbf{1}\boldsymbol{\delta}^\top$ is known it can be absorbed in the measurements; the one unknown offset τ_o can then be absorbed in receiver offsets.

D. Missing Measurements

One upshot of our formulation is that it allows localization even when some sensor–receiver measurements are unavailable.

TABLE II
PARAMETERS FOR THE SETUP AND ALGORITHMS

Dimension	$d = 3$
Volume	10 m $\times$ 10 m $\times$ 3 m
M sensors	7 to 12
K sources	7 to 12
Time offset	[-1,1] s
Speed of sound	343 m/s
Solver	SeDuMi
LM iterations	At most 1000

We denote the set of missing entries by

$$\mathcal{M} = \big\{(m,k): \text{distance between } m^{\text{th}} \text{ receiver and}$$
$$k^{\text{th}} \text{ source is missing}\big\}.$$

We then replace the objective (14a) with

$$\min_{\boldsymbol{G},\boldsymbol{B},\{\alpha_{mk}\}} \left\| \boldsymbol{J}_M\left(\boldsymbol{B}-\boldsymbol{T}+ \sum_{(m,k)\in\mathcal{M}} \alpha_{mk}\boldsymbol{e}_m\boldsymbol{e}_k^\top \right)\boldsymbol{J}_K \right\|_F^2, \tag{21}$$

where $\boldsymbol{e}_n$ denotes the canonical basis, and the coefficients α_{mk} account for the missing entries.

The LM refinement then proceeds with $\theta = \begin{bmatrix} \mathrm{vec}(\boldsymbol{R}) \\ \mathrm{vec}(\boldsymbol{S}) \\ \mathrm{vec}(\boldsymbol{\alpha}) \end{bmatrix} \in$

$\mathbb{R}^{d(M+K)+|\mathcal{M}|}$. The Jacobian now includes the derivative with respect to $\boldsymbol{\alpha}$ as well

$$[\mathrm{D}_{\boldsymbol{\alpha}}f]_{mk,ij} = \begin{cases} 1 & (m,k)\in\mathcal{M}, \ i = m, j = k, \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

Note that this approach can be used in the presence of outliers by first estimating them using RANSAC for instance. Then those outliers can be treated as missing entries.

V. NUMERICAL SIMULATIONS

In this section, we evaluate our approach on synthetic data. We implemented our algorithms in Matlab and used the CVX package for specifying convex programs [51], [52] with the SeDuMi solver. For reverberation simulations, we use the Pyroomacoustics [53] package in Python. In the following, we describe the setup and present a series of results for different scenarios showcasing the accuracy and flexibility of our method.

A. Setup

We consider an equal number of sensors and sources $M = K$ in the range of 7 to 12. For each pair, we generated 200 configurations of points randomly distributed in a volume of dimensions 10 m $\times$ 10 m $\times$ 3 m. The time offsets are uniformly generated in the range $[-1, 1]$ s. The TOAs were finally corrupted by zero-mean random Gaussian noise with standard deviation $\sigma \in \{10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 0\}$ s or equivalently $\sigma \in \{34.3, 3.4, 0.3, 0.03, 0\}$ cm. The parameters are summarized in Table II.
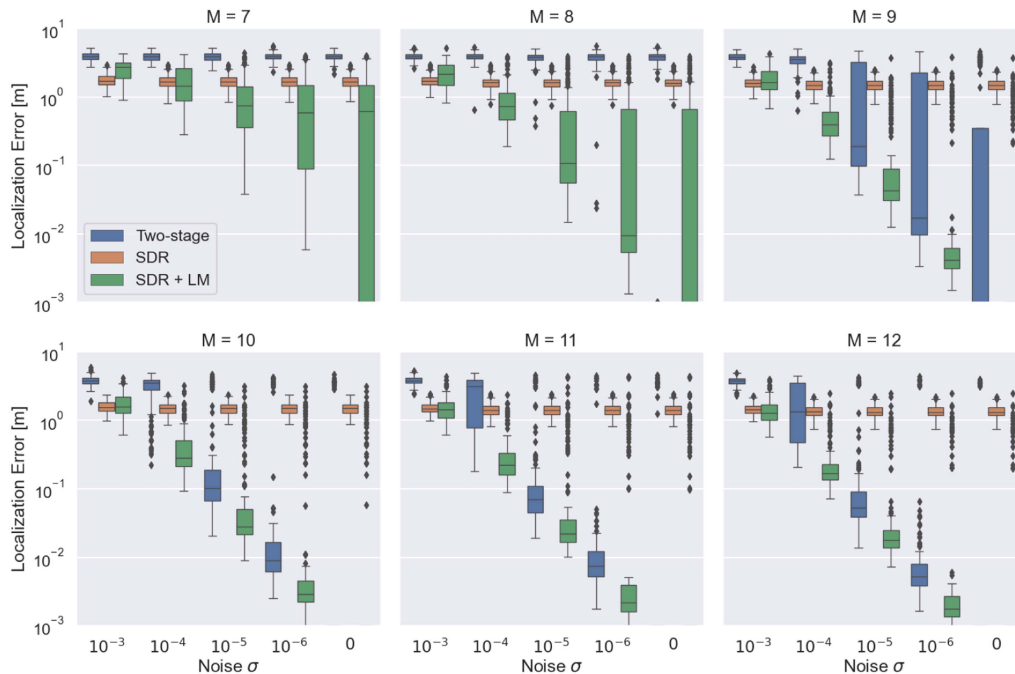
Fig. 2. Localization results for an unsynchronized setup comparing our approach (SDR+LM) to the two-stage baseline (Wang et al. [16]). We also show the results before the LM refinement (SDR). Results are shown at different noise levels and for $M = K$ in the range of 7 to 12.

B. Evaluation

The reconstructed points are first optimally aligned using Procrustes analysis [54]. The localization error between the true $[\boldsymbol{R}, \boldsymbol{S}]$ and reconstructed $[\hat{\boldsymbol{R}}, \hat{\boldsymbol{S}}]$ points is then

$$E_{\mathrm{rs}} = \frac{\sum_{m=1}^{M} \|\boldsymbol{r}_m - \hat{\boldsymbol{r}}_m\| + \sum_{k=1}^{K} \|\boldsymbol{s}_k - \hat{\boldsymbol{s}}_k\|}{M + K}. \quad (23)$$

We present the results using box plots depicting the median, first quartile, and third quartile. The whiskers correspond to 1.5 times the inter-quartile range. Values over or below this are shown as outliers.

To facilitate comparison, we will also present the median and the 95% confidence interval. In all figures, we clip the errors to a minimum of 10^{-3} which we assume is sufficient for most applications.

C. Results

As a baseline algorithm, we choose to compare to the two-stage approach of Wang et al. [16]. We consider two aspects in our comparison: localization error and runtime. Note that we only report the results of Wang et al.'s approach without the further third-stage refinement using Ono et al.'s [30] algorithm. We had observed that the refinement takes a long time due to the required number of iterations and did not improve the results in the case $M = K = 7$.

The localization results comparing our approach (SDR+LM) to the two-stage baseline are shown in Fig. 2. We also show our SDR results before the LM refinement. Note that while the SDR-only results are poor, they indeed provide a good initialization for the subsequent LM refinement as evidenced by the SDR+LM
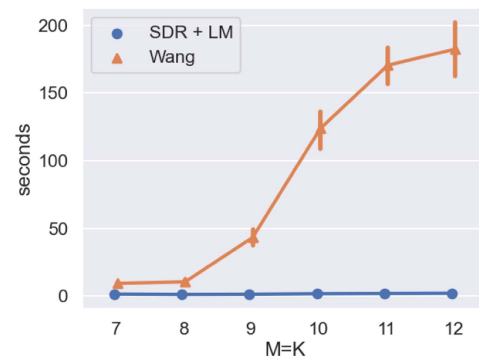

Fig. 3. Average runtime of our approach (SDR+LM) compared to the two-stage baseline (Wang et al. [16]). Results are for $M = K$ in the range of 7 to 12. The noise standard deviation is fixed $\sigma = 10^{-5}$. The runtime of our approach is consistently less than 5 seconds.

results. We further note that the outlier errors correspond to geometrically unfavorable source/receiver configurations that have attractive local minima. As can be seen, our timing-invariant approach outperforms the two-stage method, especially in the near-minimal configurations. As for the runtime, we show in Fig. 3 a comparison of the average runtime over 20 runs as we increase the number of sensors and receivers for a fixed noise standard deviation $\sigma = 10^{-5}$. The runtime of our approach is consistently less than 5 seconds, whereas the runtime for the two-stage approach increases significantly with the number of sensors. Thus, our timing-invariant approach avoids both having timing-estimation errors that propagate to the position estimation as well as the need for multiple random initializations that increase the runtime.
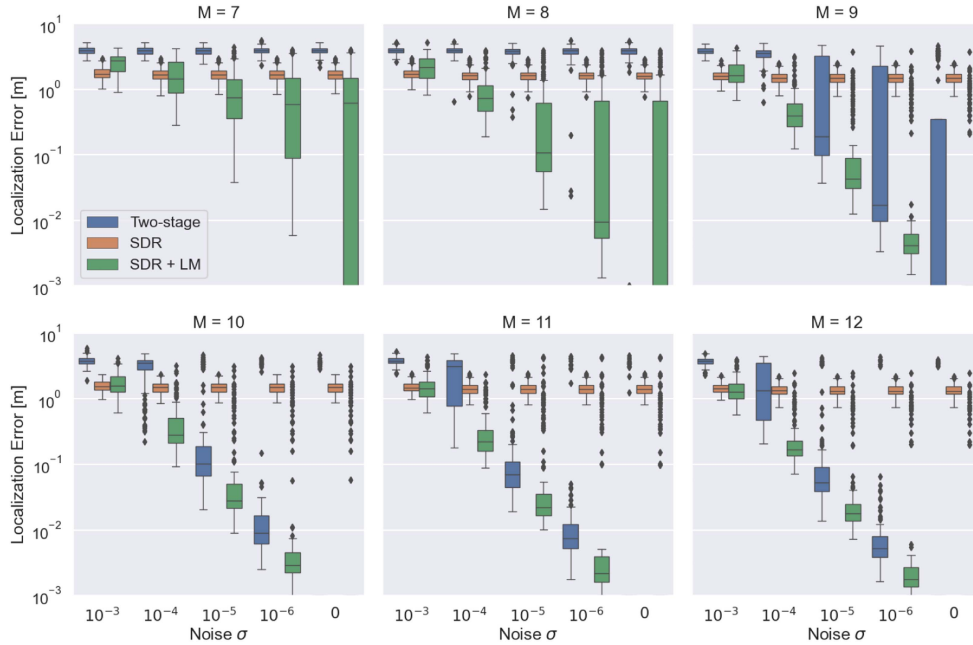
Fig. 4.   Localization errors for an unsynchronized setup in different scenarios. Results are shown at different noise levels and for $M = K$ in the range of 7 to 12. In the case without any synchronization, our approach (SDR+LM) outperforms the two-stage baseline (Wang et al. [16]). Side information such as partial synchronization or knowledge of some distances improves the results.

We now evaluate the case when the microphones are synchronized but the source emission times are still unknown. The results are shown in Fig. 4 along with the results of the fully unsynchronized case. As can be seen, the localization performance with partial synchronization is better, especially for $(M = 7, K = 7)$ and $(M = 8, K = 8)$.

Next, we test the case when the inter-microphone distances are known, but the whole setup is still fully unsynchronized. The localization errors are shown in Fig. 4. As would be expected, having side information significantly improves the localization performance.

For the last experiment, we attempt localization when some TOA data is missing. We test a range of missing entries from 4% to 20% of the total. Fig. 5 shows the results for $M = K = 12$ where we also compare the performance to the complete data case. We can see that we are still able to localize when few entries are missing. However, as we have found some intrinsically bad configurations that cannot be accurately localized with complete data, it is expected that the problem would be exacerbated when some TOA entries are further missing.

In summary, our approach (SDR+LM) outperforms the two-stage baseline (Wang et al. [16]) in terms of localization error and runtime. Additional side information such as partial synchronization or knowledge of some distances can be easily accommodated in our formulation, and as expected improve the localization results compared to the fully unsynchronized case.

D. Reverberation Results

We now evaluate the localization performance of our approach in the presence of reverberation.

TABLE III
PARAMETERS FOR THE REVERBERATION SIMULATION

Room dimensions	10 m × 10 m × 3 m
M=K sensors and sources	12
Time offset	[0,2] s
Source duration	0.5 s
Speed of sound	343 m/s
Sampling rate	48000 Hz for noise and 16000 Hz for speech
Reverberation time	0 to 0.5 s
SNR	15 dB

We consider a room of dimensions $10 \times 10 \times 3$. The $M = K = 12$ sensors and sources are randomly placed in the room. The sources are random Gaussian noise and do not overlap such that the segmentation at each microphone is known. We also test with speech sources from the Speech Commands dataset [55]. The SNR is set to 15 dB.

We vary the reverberation time by adjusting the wall absorption coefficients and maximum order for the image-source method. For each reverberation time, we simulate 20 realizations. While the desired reverberation times are 0 s, 0.1 s, 0.2 s, 0.3 s, 0.4 s, and 0.5 s, the actual reverberation times obtained in simulation are slightly different but close. The parameters are summarized in Table III.

The TDOA are estimated using the method of Yamaoka et al. [56]. The error between the true t_{mk} and estimated $\hat{t}_{mk}$ TDOA is calculated as

$$E_t = \frac{\sum_{m=1}^{M} \sum_{k=1}^{K} |t_{mk} - \hat{t}_{mk}|}{MK}. \tag{24}$$

We then use the estimated TDOA as input to our localization algorithm as well as the two-stage baseline of Wang et al. [16].
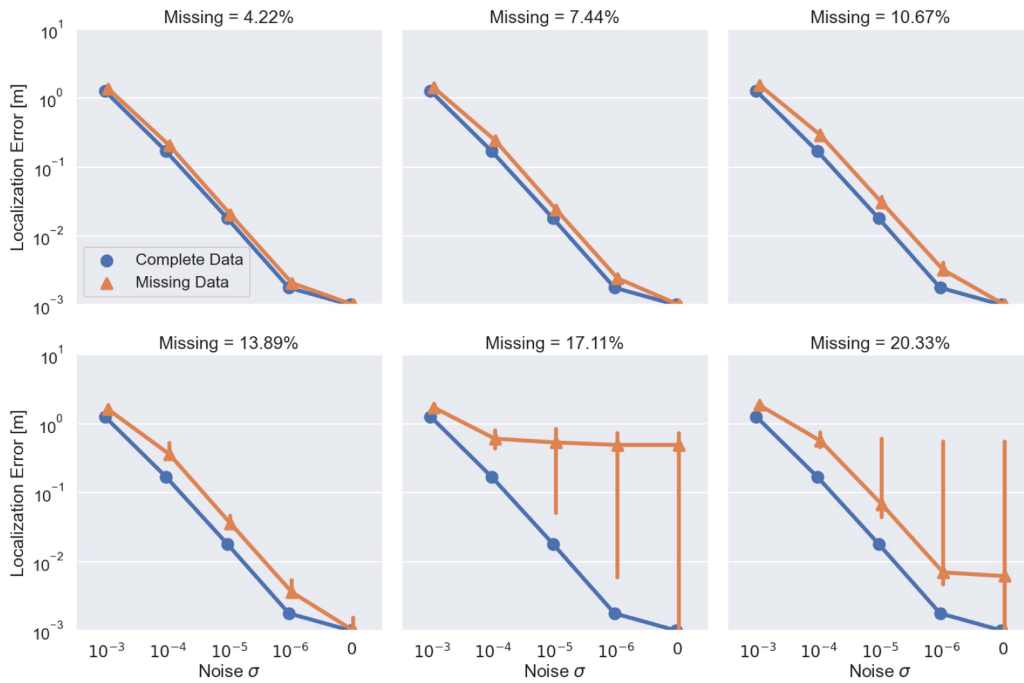
Fig. 5. (Missing Data.) Median localization errors for $M = K = 12$ in the case of missing TOA entries at different levels of noise. Between 4% and 20% of entries are missing. The whole setup is unsynchronized. The results are plotted against those without any missing data.
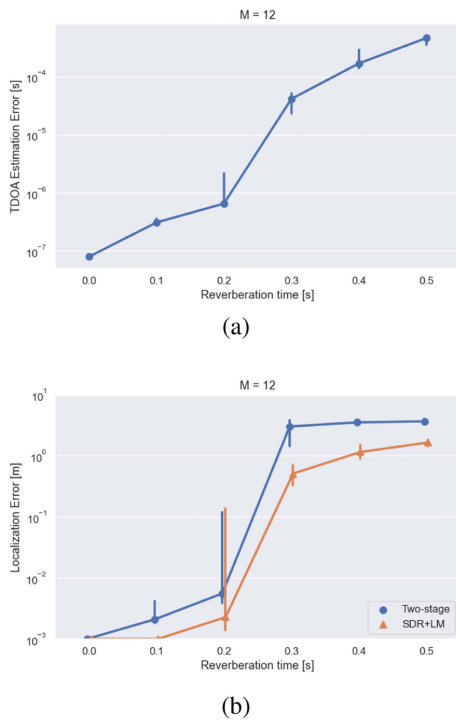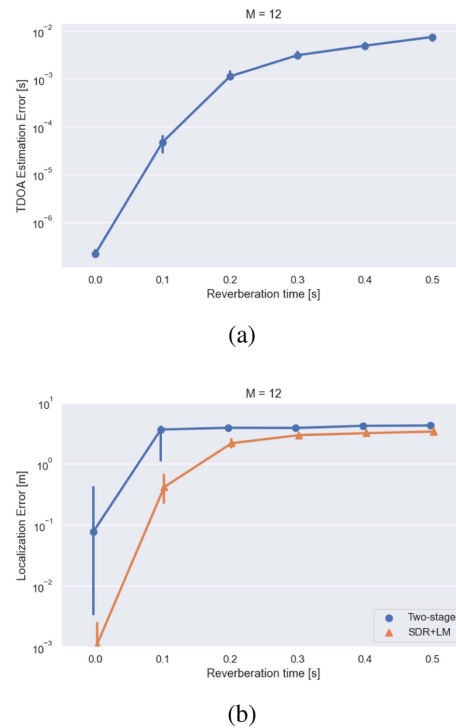

Fig. 6. (Reverberation.) Results with noise sources as the reverberation time increases. (a) Median TDOA estimation error. (b) Median localization error. Our timing-invariant approach outperforms the two-stage baseline.

Fig. 7. (Reverberation.) Results with speech sources as the reverberation time increases. (a) Median TDOA estimation error. (b) Median localization error. The TDOA estimation errors are larger with speech compared to noise sources. The subsequent localization is also worse. Our timing-invariant approach outperforms the two-stage baseline.

The results are shown in Figs. 6 and 7. In Fig. 6(a), we show the median TDOA estimation error. As expected, the errors are larger as the reverberation time increases. Subsequently, this affects the localization errors as shown in Fig. 6(b). A similar trend is observed with the speech sources in Fig. 7(a). For both source types, our timing-invariant approach still outperforms the two-stage baseline.
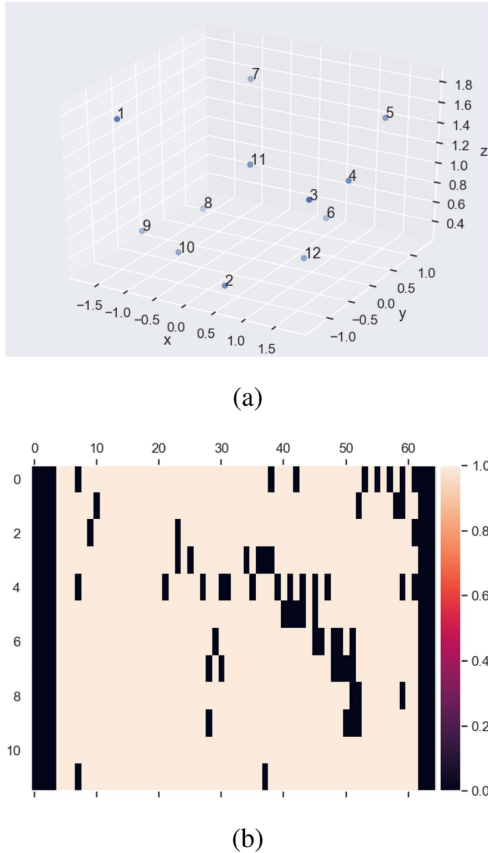
(a)

(b)

Fig. 8.   (Real Data.) Data from a real experiment carried out in an office environment. (a) True positions of 12 microphones. (b) Mask showing outliers in the TOA data.

VI.   EXPERIMENTS

In this section, we evaluate our approach on real data [46] recorded in an office with most of the furniture removed. A loudspeaker played a chirp from 65 positions. We have access to the ground truth positions for 12 microphones measured using a laser, the 12×65 TOA matrix, and a mask indicating the positions of outlier TOA entries. While the TOA was extracted from the recordings knowing the chirp, our algorithms also work for TDOA matrices which could have been extracted without any assumption on the source as shown in Section V-D.

A.   Setup

Fig. 8 shows the TOA matrix, outlier positions, and the true microphone positions. A total of 23 loudspeakers provide clean data without outliers. Thus, for the first experiments, we will use the 12×23 subset of TOAs. Then, for the full 12×65 TOAs, we will use our missing data approach to handle the outliers. We report results over 200 runs where we randomly choose a subset of the loudspeakers to localize the 12 microphones.

Since we only have the ground truth for the microphone positions, we calculate the localization error between the true $\boldsymbol{R}$ and reconstructed $\hat{\boldsymbol{R}}$ microphone positions as

$$E_{\mathrm{r}} = \frac{\sum_{m=1}^{M} \|\boldsymbol{r}_m - \hat{\boldsymbol{r}}_m\|}{M}. \tag{25}$$
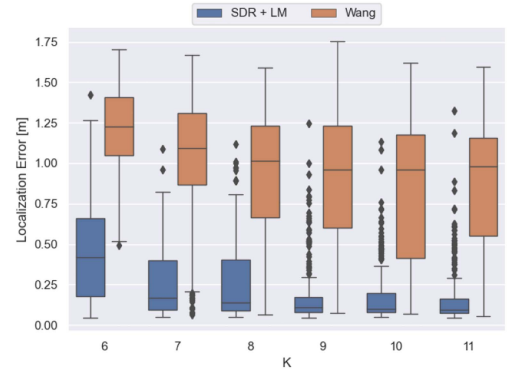

Fig. 9.   (Real Data.) Localization errors for localizing 12 microphones using 6 to 11 sources. We compare our approach (SDR+LM) to a two-stage approach (Wang). The minimum error with our approach is consistently less than 5 cm. The input TOA matrix is a subset of the 12×23 outlier-free TOA matrix.
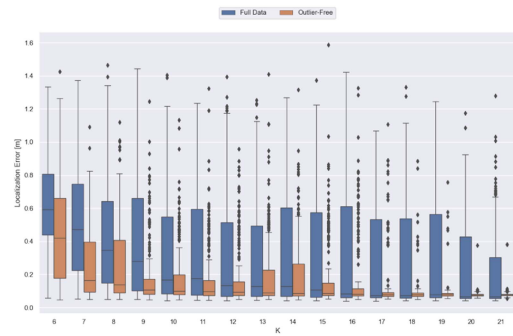

Fig. 10.   (Real Data.) Localization errors for localizing 12 microphones using 6 to 21 sources. We compare using subsets of the available 12×65 TOA matrix (Full Data) to using subsets of the 12×23 outlier-free TOA matrix (Outlier-Free). For the former, we use the missing data approach where we treat outliers as missing entries.

B.   Results

We first test our approach on subsets of the 12×23 TOA matrix that is outlier-free. We also compare to the Wang et al. [16] two-stage baseline. In Fig. 9(a), we show the errors for localizing all 12 microphones using a varying number of loudspeakers from 6 (the minimal case for $M = 12$) to 11. Once again, our approach significantly outperforms the two-stage baseline. Also similar to what we observed in the numerical simulations, while using more loudspeakers improves the average error, for each K, there are a number of large errors corresponding to intrinsically difficult configurations that have attractive non-global minima. However, the minimum error is consistent for all K and is less than 5 cm.

Next, we use subsets of the entire 12×65 TOA that contains outliers but assume knowing where the outliers are. We can thus use the missing data approach described in Section IV-D. Fig. 10 shows the errors for localizing the 12 microphones using 6 to 21 sources. On one hand, the average error is larger compared to the outlier-free case. On the other hand, the best case scenarios that result in minimum errors are comparable to the outlier-free case. The smallest localization error is 4 cm and corresponds to $K = 16$ loudspeakers.

VII. Conclusion

We formulated a timing-invariant objective that allows us to localize sensors and sources in an unsynchronized setting. Based on this objective, we proposed an SDP relaxation using the full Gram matrix of the point set and then a subsequent refinement step using the LM algorithm. We have thus eliminated the need for having to first estimate the unknown timings as well as avoided the multiple initializations required when solving non-convex problems. We also proposed an approach to handle missing data and showed how to seamlessly incorporate additional information such as knowledge of some distances or partial synchronization.

Using numerical simulations, we demonstrated the feasibility of the approach in different scenarios and in near-minimal configurations. We compared our approach to a two-stage state-of-the-art method [16]. Not only did our approach outperform the two-stage algorithm in terms of localization error, but also in terms of runtime. We also tested our algorithms on real times of arrival measured in an office environment. We were able to localize the 12 microphones to within 0.04 m average error.

Future work could focus on dealing with outliers which arise in the presence of multipath, for example. The method could be similar to our missing data approach except that the positions of the outliers are not known and need to be determined.

Appendix A
Localization of Subarrays

The LM refinement proceeds by iteratively minimizing the augmented Lagrangian over θ

$$\|f(\theta)\|^2 + g(\theta)^\top z + \mu \|g(\theta)\|^2, \qquad (26)$$

where z is the Lagrange multiplier and $\mu > 0$. Minimizing (26) is equivalent to minimizing [50]

$$\|f(\theta)\|^2 + \mu \|g(\theta) + z/(2\mu)\|^2. \qquad (27)$$

Since (27) is nonlinear in θ, the LM algorithm is once again used in which the update is

$$
\theta^{(i+1)} = \theta^{(i)} - \Big(\mathsf{D}f(\theta^{(i)})^\top \mathsf{D}f(\theta^{(i)})
$$
$$
+ \mu \mathsf{D}g(\theta^{(i)})^\top \mathsf{D}g(\theta^{(i)}) + \lambda^{(i)} \boldsymbol{I} \Big)^{-1}
$$
$$
\times \Big(\mathsf{D}f(\theta^i)^\top f(\theta^{(i)}) + \mu \mathsf{D}g(\theta^{(i)})^\top (g + z/(2\mu)) \Big)
$$

where $\mathsf{D}g \in \mathbb{R}^{|\mathcal{M}| \times d(M+K)}$ is the Jacobian defined as

$$
\mathsf{D}g_{ij,p} = \begin{cases} \boldsymbol{x}_i - 2\boldsymbol{x}_j & p = i, \\ -2\boldsymbol{x}_i + \boldsymbol{x}_j & p = j, \\ 0 & \text{otherwise.} \end{cases} \qquad (28)
$$

References

[1] G. San Antonio, D. R. Fuhrmann, and F. C. Robey, "MIMO radar ambiguity functions," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 1, pp. 167–177, Jun. 2007.

[2] R. Parhizkar, A. Karbasi, S. Oh, and M. Vetterli, "Calibration using matrix completion with application to ultrasound tomography," *IEEE Trans. Signal Process.*, vol. 61, no. 20, pp. 4923–4933, Oct. 2013.

[3] B. G. Ferguson, "Improved time-delay estimates of underwater acoustic signals using beamforming and prefiltering techniques," *IEEE J. Ocean. Eng.*, vol. 14, no. 3, pp. 238–244, Jul. 1989.

[4] A. J. Devaney, "Time reversal imaging of obscured targets from multistatic data," *IEEE Trans. Antennas Propag.*, vol. 53, no. 5, pp. 1600–1610, May 2005.

[5] D. Ciuonzo, G. Romano, and R. Solimene, "Performance analysis of time-reversal MUSIC," *IEEE Trans. Signal Process.*, vol. 63, no. 10, pp. 2650–2662, May 2015.

[6] R. Scheibler, D. Di Carlos, A. Deleforge, and I. Dokmanić, "Separake: Source separation with a little help from echoes," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018, pp. 6897–6901.

[7] H. Pan, R. Scheibler, E. Bezzam, I. Dokmanić, and M. Vetterli, "FRIDA: FRI-based DOA estimation for arbitrary array layouts," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2017, pp. 3186–3190.

[8] M. Martinez-Camara, I. Dokmanić, J. Ranieri, R. Scheibler, M. Vetterli, and A. Stohl, "The Fukushima inverse problem," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2013, pp. 4330–4334.

[9] S. Simoni et al., "Hydrologic response of an alpine watershed: Application of a meteorological wireless sensor network to understand streamflow generation," *Water Resour. Res.*, vol. 47, pp. 1–16, no. 10, 2011.

[10] M. Z. Win, F. Meyer, Z. Liu, W. Dai, S. Bartoletti, and A. Conti, "Efficient multisensor localization for the Internet of Things: Exploring a new class of scalable localization algorithms," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 153–167, Sep. 2018.

[11] H. Simkovits, A. J. Weiss, and A. Amar, "Navigation by inertial device and signals of opportunity," *Signal Process.*, vol. 131, pp. 280–287, 2017.

[12] Y. Rockah and P. Schultheiss, "Array shape calibration using sources in unknown locations–Part I: Far-field sources," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 3, pp. 286–299, Mar. 1987.

[13] Y. Rockah and P. Schultheiss, "Array shape calibration using sources in unknown locations–Part II: Near-field sources and estimator implementation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 6, pp. 724–735, Jun. 1987.

[14] A. J. Weiss and B. Friedlander, "Array shape calibration using sources in unknown locations-a maximum likelihood approach," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 1958–1966, Dec. 1989.

[15] A. Plinge, F. Jacob, R. Haeb-Umbach, and G. A. Fink, "Acoustic microphone geometry calibration: An overview and experimental evaluation of state-of-the-art algorithms," *IEEE Signal Process. Mag.*, vol. 33, no. 4, pp. 14–29, Jul. 2016.

[16] L. Wang, T.-K. Hon, J. D. Reiss, and A. Cavallaro, "Self-localization of ad-hoc arrays using time difference of arrivals," *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 1018–1033, Feb. 2016.

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

[18] J. B. Kruskal and M. Wish, *Multidimensional scaling*, no. 11 Thousand Oaks, CA, USA: SAGE Pub., 1978.

[19] I. Dokmanić, R. Parhizkar, J. Ranieri, and M. Vetterli, "Euclidean distance matrices: Essential theory, algorithms, and applications," *IEEE Signal Process. Mag.*, vol. 32, no. 6, pp. 12–30, Nov. 2015.

[20] C. Vanwynsberghe, P. Challande, J. Marchal, R. Marchiano, and F. Ollivier, "A robust and passive method for geometric calibration of large arrays," *J. Acoustical Soc. Amer.*, vol. 139, no. 3, pp. 1252–1263, 2016, doi: 10.1121/1.4944566.

[21] P. H. Schönemann, "On metric multidimensional unfolding," *Psychometrika*, vol. 35, no. 3, pp. 349–366, 1970.

[22] M. Crocco, A. Del Bue, and V. Murino, "A bilinear approach to the position self-calibration of multiple sensors," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 660–673, Feb. 2012.

[23] P. Biswas, T.-C. Lian, T.-C. Wang, and Y. Ye, "Semidefinite programming based algorithms for sensor network localization," *ACM Trans. Sensor Netw.*, vol. 2, no. 2, pp. 188–220, 2006.

[24] P. Biswas, T.-C. Liang, K.-C. Toh, Y. Ye, and T.-C. Wang, "Semidefinite programming approaches for sensor network localization with noisy distance measurements," *IEEE Trans. Automat. Sci. Eng.*, vol. 3, no. 4, pp. 360–371, Oct. 2006.

[25] K. Yang, G. Wang, and Z.-Q. Luo, "Efficient convex relaxation methods for robust target localization by a sensor network using time differences of arrivals," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2775–2784, Jul. 2009.

[26] R. M. Vaghefi and R. M. Buehrer, "Asynchronous time-of-arrival-based source localization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2013, pp. 4086–4090.

[27] G. Wang, Y. Li, and N. Ansari, "A semidefinite relaxation method for source localization using TDOA and FDOA measurements," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 853–862, Feb. 2013.

[28] A. Y. Alfakih, A. Khandani, and H. Wolkowicz, "Solving euclidean distance matrix completion problems via semidefinite programming," *Comput. Optim. Appl.*, vol. 12, no. 1, pp. 13–30, 1999.

[29] Y. Ding, N. Krislock, J. Qian, and H. Wolkowicz, "Sensor network localization, euclidean distance matrix completions, and graph realization," *Optim. Eng.*, vol. 11, no. 1, pp. 45–66, 2010.

[30] N. Ono, H. Kohno, N. Ito, and S. Sagayama, "Blind alignment of asynchronously recorded signals for distributed microphone array," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2009, pp. 161–164.

[31] J. Zhang, R. C. Hendriks, and R. Heusdens, "Structured total least squares based internal delay estimation for distributed microphone auto-localization," in *Proc. IEEE Int. Workshop Acoust. Signal Enhancement*, 2016, pp. 1–5.

[32] N. D. Gaubitch, W. B. Kleijn, and R. Heusdens, "Auto-Localization in Ad-Hoc Microphone Arrays," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2013, pp. 106–110.

[33] S. Thrun, "Affine structure from sound," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 1353–1360.

[34] M. Kreković, G. Baechler, I. Dokmanić, and M. Vetterli, "Structure from sound with incomplete data," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018, pp. 3539–3543.

[35] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan, "BeepBeep: A high accuracy acoustic ranging system using cots mobile devices," in *Proc. ACM 5th Int. Conf. Embedded Networked Sensor Syst.*, 2007, pp. 1–14.

[36] C. Vanwynsberghe, "Réseaux à grand nombre de microphones: Applicabilité et mise en œuvre," Ph.D. dissertation, Paris 6, Paris, France, 2016.

[37] I. Dokmanić, J. Ranieri, and M. Vetterli, "Relax and unfold: Microphone localization with euclidean distance matrices," in *Proc. IEEE 23rd Eur. Signal Process. Conf.*, 2015, pp. 265–269.

[38] M. Pollefeys and D. Nister, "Direct computation of sound and microphone locations from time-difference-of-arrival data," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2008, pp. 2445–2448.

[39] R. Heusdens and N. Gaubitch, "Time-delay estimation for TOA-based localization of multiple sensors," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2014, pp. 609–613.

[40] F. Jiang and Y. Kuang et al., "Time delay estimation for TDOA self-calibration using truncated nuclear norm regularization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2013, pp. 3885–3889.

[41] Y. Kuang, S. Burgess, A. Torstensson, and K. Åström, "A complete characterization and solution to the microphone position self-calibration problem," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2013, pp. 3875–3879.

[42] Y. Kuang and K. Åström, "Stratified sensor network self-calibration from TDOA measurements," in *Proc. IEEE 21st Eur. Signal Process. Conf.*, 2013, pp. 1–5.

[43] S. Zhayida, S. Burgess, Y. Kuang, and K. Åström, "TOA-based self-calibration of dual-microphone array," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 5, pp. 791–801, Aug. 2015.

[44] M. Farmani, R. Heusdens, M. S. Pedersen, and J. Jensen, "TDOA-based self-calibration of dual-microphone arrays," in *Proc. IEEE 24th Eur. Signal Process. Conf.*, 2016, pp. 617–621.

[45] S. Burgess, Y. Kuang, and K. Åström, "TOA sensor network self-calibration for receiver and transmitter spaces with difference in dimension," *Signal Process.*, vol. 107, pp. 33–42, 2015.

[46] K. Batstone et al., "Robust self-calibration of constant offset time-difference-of-arrival," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019, pp. 4410–4414.

[47] I. Dokmanić, L. Daudet, and M. Vetterli, "How to localize ten microphones in one finger snap," in *Proc. IEEE 22nd Eur. Signal Process. Conf.*, 2014, pp. 2275–2279.

[48] I. Dokmanić, L. Daudet, and M. Vetterli, "From acoustic room reconstruction to SLAM," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2016, pp. 6345–6349.

[49] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Methods Softw.*, vol. 11, no. 1–4, pp. 625–653, 1999, doi: 10.1080/10556789908805766.

[50] S. Boyd and L. Vandenberghe, *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares*. Cambridge, U.K.: Cambridge Univ. Press, 2018.

[51] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming," Mar. 2014. [Online]. Availabel: http://cvxr.com/cvx

[52] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Adv. in Learn. and Control* (Lecture Notes in Control and Information Sciences Series), V. Blondel, S. Boyd, and H. Kimura, Eds. Berlin, Germany: Springer, 2008, pp. 95–110. [Online]. Available: http://stanford.edu/boyd/graph_dcp.html

[53] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A Python package for audio room simulations and array processing algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018, pp. 351–355.

[54] P. H. Schönemann, "A generalized solution of the orthogonal procrustes problem," *Psychometrika*, vol. 31, no. 1, pp. 1–10, 1966, doi: 10.1007/BF02289451.

[55] P. Warden, "Speech commands: A public dataset for single-word speech recognition," 2017. [Online]. Available: http://download.tensorflow.org/data/speech_commands_v0.01.tar.gz

[56] K. Yamaoka, R. Scheibler, N. Ono, and Y. Wakabayashi, "Sub-sample time delay estimation via auxiliary-function-based iterative updates," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2019, pp. 130–134.