

# FROM ACOUSTIC ROOM RECONSTRUCTION TO SLAM

Ivan Dokmanić<sup>†‡</sup>, Laurent Daudet<sup>‡</sup>, and Martin Vetterli<sup>†</sup>

<sup>†</sup>School of Computer and Communication Sciences  
Ecole Polytechnique Fédérale de Lausanne (EPFL)  
CH-1015 Lausanne, Switzerland  
{ivan.dokmanic, martin.vetterli}@epfl.ch

<sup>‡</sup>Institut Langevin  
ESPCI, Université Paris Diderot/CNRS UMR 7587  
1 rue Jussieu 75005 Paris, France  
laurent.daudet@espci.fr

## ABSTRACT

Recent works on reconstruction of room geometry from echoes assume that the geometry of the sensor array is known. In this paper, we show that such an assumption is not essential; echoes provide sufficient clues to reconstruct the room's and the array's geometries jointly, even from a single acoustic event. Rather than focusing on the combinatorial problem of matching the walls and the recorded echoes, we provide algorithms for solving the joint estimation problem in practical cases when this matching is known and the number of microphones is small. We then explore intriguing connections between this problem and simultaneous localization and mapping (SLAM), and show that SLAM can be solved by the same methods. Finally, we demonstrate how effective the proposed methods are by numerical simulations and experiments with real measured room impulse responses.

**Index Terms**—Room geometry, microphone array, localization, acoustics, simultaneous localization and mapping (SLAM).

## 1. INTRODUCTION

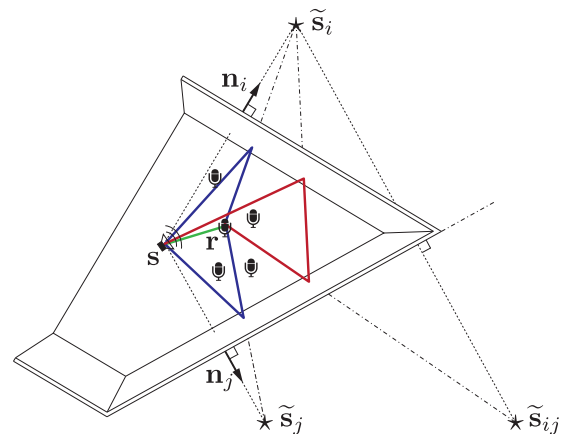
This paper is inspired by three problems that have recently received considerable attention. First one is acoustic room geometry reconstruction: in a series of works [1–3] it has been shown that we can reconstruct polyhedral rooms from RIRs recorded by a few microphones, and we can do it even with a single source [1].

The second problem is self-localization of ad hoc microphone arrays [4–6]: we now know that it is not necessary to use a tape meter or devise special calibration rigs [7] to localize the microphones; one can use spurious acoustic events.

The third problem is simultaneous localization and mapping (SLAM) [8]. In SLAM, a robot moves along an unknown trajectory in an unknown environment. Often, we have a noisy idea about the trajectory. The goal is then to reconstruct both the trajectory and the map of the environment. The map is often specified in terms of *landmarks*, and reconstructing the map is equivalent to localizing the landmarks. There is a variety of sophisticated measurements that can be used to do SLAM. The most rudimentary one is a pure range measurement, leading to what is known as range-only SLAM [9–12].

We want to show that echoes contain a surprising amount of information and that very little need be known a priori in order to reason about the geometry of the sensing setup or the room. Along the way, we show that room reconstruction (with or without knowing the array geometry), microphone self-localization, and range-only SLAM are all instances of the multidimensional unfolding (MDU) problem [13].

This work was supported by the Swiss National Science Foundation grant number 200021\_138081, “Non-linear Sampling Methods”.



**Fig. 1.** Illustration of the image source model for first- and second-order echoes. The vector  $\mathbf{n}_i$  is the outward-pointing unit normal associated with the  $i$ th wall. Stars denote the image sources, and  $\tilde{\mathbf{s}}_{ij}$  is the image source corresponding to the second-order echo. Sound rays corresponding to first reflections are shown in purple, and the ray corresponding to the second-order reflection is shown in green.

Finally, an important contribution of this work is to show that these problems can be solved with very few measurements using Euclidean distance matrices (EDMs) [14, 15]. This approach also allows us to take advantage of prior geometric knowledge, much like the inertial information in SLAM.

## 2. ROOM RECONSTRUCTION

We start by discussing room reconstruction (“Can one hear the shape of a room?”), and formulating it as an optimization problem:

**Problem 1.** Let  $\mathcal{R} \subset \mathbb{R}^3$  be a convex polyhedral<sup>1</sup> room described by  $K$  walls with normals  $\{\mathbf{n}_k\}$  and centroids  $\{\mathbf{p}_k\}$ . Given a set of  $M$  microphones at known locations  $\{\mathbf{r}_m \in \mathcal{R}\}$ , and a single impulsive source at an unknown location  $\mathbf{s} \in \mathcal{R}$ , determine the shape of the room.

We extract parts of the solution from [1] that will provide the context for the remainder of the article. First, we model the walls using the image source model [16, 17], thus replacing every reflection by

<sup>1</sup>Convexity is only one possible model that enables us to reconstruct the room. The algorithm will correctly localize the image sources regardless of whether the room is convex or not. If we are only after the image sources and the microphone locations (and in most audio applications that is what we need), then we do not need convexity. If we do want to reconstruct the room shape, then a model is necessary, especially with few microphones.

an equivalent point source. With reference to Fig. 1, a first order reflection—corresponding to a single bounce off a wall—can be modeled by a virtual source that is the mirror image of the real source across the  $k$ th wall. An image source  $\tilde{\mathbf{s}}_k$  of  $\mathbf{s}$  corresponding to the  $k$ th wall is given as  $\tilde{\mathbf{s}}_k = \mathbf{s} + 2\langle \mathbf{p}_k - \mathbf{s}, \mathbf{n}_k \rangle \mathbf{n}_k$ . Higher order reflections are simply images of lower order images. An idealized room impulse response (RIR) measured by a microphone at position  $\mathbf{r} \in \mathbb{R}^3$  is a train of pulses,

$$h(t; \mathbf{r}) = \sum_{k \geq 0} \alpha_k \delta(t - \tau_k(\mathbf{r})) = \sum_{k \geq 0} \alpha_k \delta(t - \|\mathbf{r} - \mathbf{s}_k\|/c), \quad (1)$$

where  $c$  is the speed of sound,  $(\mathbf{s}_k, k \geq 0)$  lists the real source and all its images up to an arbitrarily high order, and  $\alpha_k$  are the received magnitudes, which depend on the wall absorption coefficients and the distance of the image source from the microphone. From a set of recorded RIRs  $\{h(t, \mathbf{r}_m)\}_{m=1}^M$  as in (1), we estimate the set of distances  $\tilde{d}_{mk}$  between the  $m$ th microphone and the  $k$ th image source.

Knowing the locations of the image sources is equivalent to knowing the room geometry, thus Problem 1 can be recast as a problem of simultaneously localizing multiple point sources. The optimal locations are found as the unique minimizer of the *s-stress* cost function:

$$\{\hat{\mathbf{s}}_k\}_{k=1}^K = \arg \min_{\{\mathbf{s}_k\}} \sum_{m=1}^M \sum_{k=0}^K (\tilde{d}_{m\pi_m(k)}^2 - \|\mathbf{r}_m - \mathbf{s}_k\|^2)^2, \quad (2)$$

where permutations  $\pi_m(k)$  model the ambiguity in grouping the pulses in  $M$  measured impulse responses corresponding to the same echo. A discussion of *echo sorting*, which solves this problem (see [1, 18]), is outside of the scope of this article.<sup>2</sup> We assume that this assignment is known, and in the following set  $\pi_m(k) = k$ .

Thanks to the assumption that the array geometry is known, (2) decouples over image sources—we can solve it one  $\mathbf{s}_k$  at a time:

$$\hat{\mathbf{s}}_k = \arg \min_{\mathbf{s}_k} \sum_{m=1}^M (\tilde{d}_{mk}^2 - \|\mathbf{r}_m - \mathbf{s}_k\|^2)^2. \quad (3)$$

The problem (3) is a variant of *multilateration*, and it is not convex. Nevertheless, it can be solved exactly using efficient algorithms [19].

### 3. ZERO-KNOWLEDGE CALIBRATION

So far, we assumed that we know the geometry of the microphone array. The locations of the real source and of the image sources were then produced by the algorithm. Imagine now that we simply place the microphones in the room and measure the impulse responses from a single source, but we forget to measure the inter-microphone distances. Can we recover the geometry of the microphone array and save the experiment? If yes, then we can clearly recover the room by the results of Section 2. We consider the following problem:

**Problem 2.** *Let a polyhedral room be described by  $K$  walls with normals  $\{\mathbf{n}_k\}$  and centroids  $\{\mathbf{p}_k\}$ . Given a set of  $M$  microphones at unknown locations  $\{\mathbf{r}_m\}$ , and a single impulsive source at an unknown location  $\mathbf{s}$ , determine the microphone locations and the shape of the room.*

We show that, remarkably, the array geometry is not necessary—echoes can provide it. A preliminary discussion of this phenomenon was given in [18].

<sup>2</sup>For small microphone arrays, or in the case of SLAM, this is not an issue. See the brief discussion in Section 4. For details on how echo sorting can be solved in the case of room reconstruction, see [1]

Similarly to the previous section, we want to solve the problem

$$\{\hat{\mathbf{s}}_k\}, \{\hat{\mathbf{r}}_m\} = \arg \min_{\{\mathbf{s}_k\}, \{\mathbf{r}_m\}} \sum_{m=1}^M \sum_{k=0}^K (\tilde{d}_{mk}^2 - \|\mathbf{r}_m - \mathbf{s}_k\|^2)^2 \quad (4)$$

where  $\mathbf{s}_0 = \mathbf{s}$ ,  $(\mathbf{s}_k, k \geq 1)$  are the image sources. However, (4) is considerably more challenging, as we optimize over both  $\{\mathbf{r}_m\}$  and  $\{\mathbf{s}_k\}$ .

This problem is equivalent to the ad hoc microphone array localization problem. There, we aim to recover the geometry of an ad hoc microphone array without tape meters or specialized infrastructure, using distance measurements between a set of *acoustic events* and the microphones. Here, the acoustic events are virtual—they are the image sources corresponding to walls. Both problems are instances of MDU [13], where the aim is to localize a set of points partitioned in two subsets from distances measured between the points belonging to different subsets.

A simple degree-of-freedom-counting exercise tells us that there are  $MK$  distance measurements,  $3K + 3M$  unknown coordinates, and 6 scalars required to specify the translation and the rotation of the coordinate system (which cannot be reconstructed). Requiring that  $MK \geq 3K + 3M - 6$  gives that the minimal cases are  $(M = 4, K = 6)$ ,  $(M = 6, K = 4)$ , and  $(M = 5, K = 5)$ . In these cases the solution set is a zero-dimensional variety [20] comprising many points (38 when  $M = 4, K = 6$ ), some of which are complex, and some real. Thus, to guarantee uniqueness without additional constraints, we need more than the minimal number of points. In what follows, we discuss the strategy to solve (4) using Euclidean distance matrices.

#### 3.1. Euclidean Distance Matrices

An EDM corresponding to the point set  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]$  is the matrix  $\mathbf{D} = [d_{ij}^2]$ , where

$$d_{ij}^2 = \|\mathbf{x}_i - \mathbf{x}_j\|^2, \quad (5)$$

and  $\|\cdot\|$  is the Euclidean norm. By expanding the norm, we see that  $\mathbf{D}$  can be computed at once as

$$\mathbf{D} = \mathcal{K}(\mathbf{G}) \stackrel{\text{def}}{=} \text{diag}(\mathbf{G})\mathbf{1}^\top + \mathbf{1} \text{diag}(\mathbf{G})^\top - 2\mathbf{G}, \quad (6)$$

where  $\mathbf{G} \stackrel{\text{def}}{=} \mathbf{X}^\top \mathbf{X}$  is the Gram matrix. Let  $\mathbf{J} \stackrel{\text{def}}{=} \mathbf{I}_N - \frac{1}{N}\mathbf{1}\mathbf{1}^\top$  be the geometric centering matrix of size  $N$ . Then assuming that the point set is centered,  $\mathbf{X}\mathbf{1} = \mathbf{0}$ , straightforward computation shows that

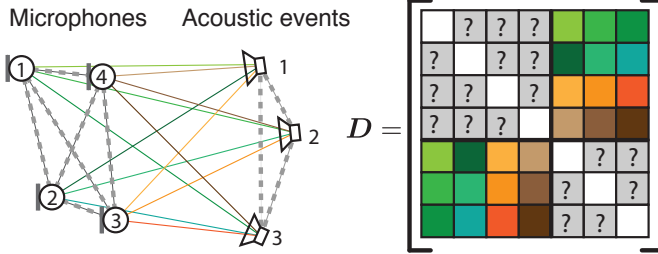
$$\mathbf{G} = -\frac{1}{2}\mathbf{J}\mathbf{D}\mathbf{J} \quad (7)$$

because the terms  $\text{diag}(\mathbf{G})\mathbf{1}^\top$  and  $\mathbf{1} \text{diag}(\mathbf{G})^\top$  are annihilated by  $\mathbf{J}$ . The point set  $\mathbf{X}$  can then be reconstructed up to a rotation and reflection by factoring  $\mathbf{G} = \mathbf{X}^\top \mathbf{X}$ . For more intuitions about (6) and (7) see [14, 15].

#### 3.2. EDM Completion for Multidimensional Unfolding

A typical EDM problem is to localize the point set from noisy and incomplete distances. A characterization of EDMs by Gower [21] states that  $\mathbf{D}$  is an EDM if and only if the corresponding geometrically centered matrix  $-\frac{1}{2}\mathbf{J}\mathbf{D}\mathbf{J}$  is positive-semidefinite (that is, if it is a Gram matrix).

We can use this correspondence to cast EDM completion and approximation as a semidefinite program. While the above characterization describes an EDM of an  $N$ -point configuration in any dimension, we are interested in a situations where  $d \ll N$  ( $d = 2$  or  $d = 3$ ). It is



**Fig. 2.** Microphone calibration as an example of MDU. We can measure only the propagation times from acoustic sources at unknown locations, to microphones at unknown locations. The corresponding revealed part of the EDM has a particular off-diagonal structure.

easy to adjust for this case by requiring that the rank of the centered Gram matrix  $-\frac{1}{2}\mathbf{J}\mathbf{D}\mathbf{J}$  be bounded by  $d$ .

Let  $\tilde{\mathbf{D}} \in \mathbb{R}^{N \times N}$  be a noisy measurement of an EDM, with some entries missing (we can arbitrarily set them to zero). Let further  $\mathbf{W} \in \mathbb{R}^{N \times N}$  have ones at positions where  $\tilde{\mathbf{D}}$  is observed, and zeros for the missing entries. Then the EDM corresponding to points in  $\mathbb{R}^d$  which best fits the measurements  $\tilde{\mathbf{D}}$  in the sense of the  $s$ -stress metric is given as an optimizer of the following program:

$$\begin{aligned} & \underset{\mathbf{G}}{\text{minimize}} && \left\| \mathbf{W} \circ (\tilde{\mathbf{D}} - \mathcal{K}(\mathbf{G})) \right\|_F^2 \\ & \text{subject to} && \mathbf{G} \geq 0, \mathbf{G}\mathbf{1} = \mathbf{0}, \text{rank}(\mathbf{G}) \leq d. \end{aligned} \quad (8)$$

Unfortunately, the rank constraint makes (8) hard to solve directly. A common strategy is to relax the rank by the trace norm; however, perhaps surprisingly, trace maximization is often a better choice when working with EDMs [15]. In the next section, we discuss an alternative technique that is particularly well-suited to our problems—the so-called convex iteration [14].

With these notations in hand, it is now simple to frame (4) as an EDM completion problem. Let  $\mathbf{D}_{\#} \stackrel{\text{def}}{=} [d_{mk}^2] \in \mathbb{R}^{M \times K}$ . Then we can solve (4) by defining

$$\mathbf{W}_{\text{MDU}} \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{0}_{M \times M} & \mathbf{1}_{M \times K} \\ \mathbf{1}_{K \times M} & \mathbf{0}_{K \times K} \end{bmatrix}, \tilde{\mathbf{D}} \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{0}_{M \times M} & \mathbf{D}_{\#} \\ \mathbf{D}_{\#}^{\top} & \mathbf{0}_{K \times K} \end{bmatrix}. \quad (9)$$

and plugging them into (8).

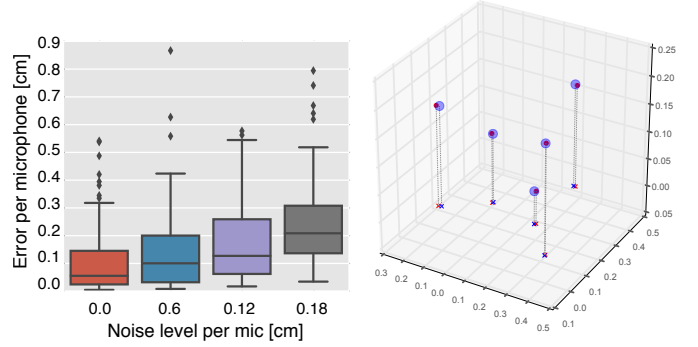
### 3.3. The Challenge of Few Microphones

The relaxation of (8) works well for multidimensional unfolding (that is, it returns a point set with the correct embedding dimension) when the number of points is large enough (empirically at least 10 or 15 microphones and equally many acoustic events [22]). This number grows larger as the quality of measurements deteriorates. But in a typical room, the number of echoes that we may clearly detect (while keeping the sorting easy) will often be smaller. Furthermore, we would like to localize smaller arrays, perhaps with only four or five microphones.

Thus we need a method that promotes low rank (low embedding dimension) more aggressively, while still minimizing  $s$ -stress subject to measurement constraints. We use the iteration of two convex programs proposed by Dattorro [14]:

$$\begin{aligned} \text{(i)} \quad \mathbf{G}^* &= \arg \min \left\{ \langle \mathbf{G}, \mathbf{Z}^* \rangle + \lambda \left\| \mathbf{W} \circ (\tilde{\mathbf{D}} - \mathcal{K}(\mathbf{G})) \right\|_F^2 : \right. \\ & \quad \left. \mathbf{G} \geq 0, \mathbf{G}\mathbf{1} = \mathbf{0} \right\} \\ \text{(ii)} \quad \mathbf{Z}^* &= \arg \min \left\{ \langle \mathbf{G}^*, \mathbf{Z} \rangle : \mathbf{0} \leq \mathbf{Z} \leq \mathbf{I}, \text{trace}(\mathbf{Z}) = N - d \right\}, \end{aligned} \quad (10)$$

with  $\mathbf{Z}^*$  initialized as  $\mathbf{I}$  (that is, starting with an ordinary trace minimization).



**Fig. 3.** Microphone localization in zero-knowledge calibration. *Left:* For each of the 4 different noise levels, we generated 100 random configurations of 6 microphones in a  $1 \text{ m}^3$  cube at the center of a  $5.1 \text{ m} \times 3.7 \text{ m} \times 2.6 \text{ m}$  shoebox room. For every microphone configuration, a source location was randomly generated inside the room. We used the distances to this source and 5 nearest image sources to estimate the geometry of the microphone array. *Right:* Experiment with real recorded RIRs in a non-shoebox room from [1]. Unlike in [1], we assume no knowledge of the microphone positions; rather, we estimate them from echoes. Blue disks show the locations estimated from tape meter measurements; red circles are *blindly* estimated locations from echoes only; ticks are in meters.

The crude physics of this method are as follows: the value of the second program at optimum is  $\sum_{i=d+1}^N \lambda_i(\mathbf{G}^*)$  where  $\lambda_1(\mathbf{G}^*) \geq \lambda_2(\mathbf{G}^*) \geq \dots \geq \lambda_N(\mathbf{G}^*)$  are the sorted eigenvalues of  $\mathbf{G}^*$ . Thus with a so-chosen  $\mathbf{Z}^*$ , minimizing  $\langle \mathbf{G}, \mathbf{Z}^* \rangle$  will have the effect of reducing the (positive) sum of the  $N - d$  smallest eigenvalues of  $\mathbf{G}$ . A deeper intuition of why this method succeeds with high likelihood is available in [14]. As we will see, the convex iteration performs very well in practice. This formulation will be particularly convenient when we come to SLAM.

### 3.4. Results

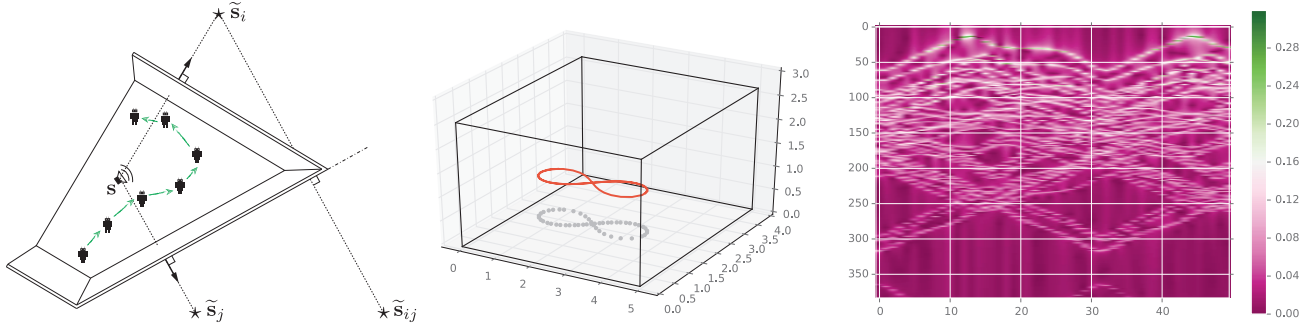
Fig. 3 shows the average error in microphone localization for a random 6-microphone array and a random source position inside a shoebox room. Even with noise and a near-minimal configuration, we obtain accurate reconstructions.

On the right, we show the results with the real data from a non-shoebox room used in experiments in [1]. We stress that the reconstruction was obtained with only 5 microphones and 6 sources (the real one and 5 virtual). There is no prior knowledge about the microphone locations—they are reconstructed from echoes. The accuracy of localization is surprisingly high, better than 1 cm per microphone. Applying the convex iteration (10) is essential to achieve such accuracy with few measurements.

## 4. ACOUSTIC SLAM

We started with the inverse problem of room geometry where we assumed that we know the geometry of the microphone array. Next, we have shown that if can detect only a couple of reflections, then the information on array geometry can be recovered from the measurements. We now assume that we only have a single microphone, albeit one that moves, and we reinterpret the derived results in the context of SLAM. Effectively, we show that range-only SLAM is yet another instance of MDU.

Imagine a robot moving about in a room, as illustrated in Fig. 4, and denote its position by  $\mathbf{r}(t)$ . Let the robot record the room response (*i.e.*, the distances to the image sources) excited by the loudspeaker at



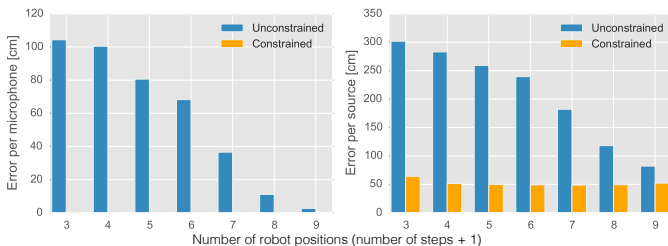
**Fig. 4.** Acoustic range-only SLAM. *Left:* Illustration of a robot moving in a room with the notation. *Middle:* A lemniscate trajectory (12) and the projection of the actual steps. *Right:* The induced RIR matrix—the horizontal axis corresponds to robot’s steps, the vertical axis to the time in the RIRs, and the pixel color encodes the magnitude of the RIR at the given robot’s position and time.

$\mathbf{s}$  at times  $t_1, \dots, t_M$ . Setting  $\mathbf{r}_m \stackrel{\text{def}}{=} \mathbf{r}(t_m)$ , we see that range-only SLAM precisely corresponds to zero-knowledge calibration (Problem 2), thus we can address it as an instance of EDM completion.

This perspective is different from what is typically done in SLAM. The usual approach is through some variant of Bayesian inference, using the Kalman (or extended Kalman) filter, or more general approaches such as FastSLAM [23]. In particular, leveraging ultra-wideband (UWB) multipath for SLAM in a probabilistic framework has been demonstrated by Leitinger and coauthors [24]. Our aim here is to show that, at its core, this is a Euclidean geometry problem.

We note that in the case of SLAM, echo sorting is much less of an issue. To understand why, take a look at Fig. 4B and C. If the robot follows a trajectory such as a lemniscate in Fig. 4B and measures the impulse response at every step, the matrix of impulse responses will be the one in Fig. 4C. Such a visual representation facilitates following the contributions of the same echo across many steps.<sup>3</sup>

A major appeal of Bayesian methods is that it is relatively straightforward to incorporate information about the robot’s kinematics or landmark distribution into the estimation. We now show by an example that similar feats are possible with EDMs.



**Fig. 5.** Position estimation errors in SLAM. The robot was following a lemniscate (12), with the step size  $t_n = 0.3n$  ( $\approx 30$  cm) in the room of dimensions  $5 \text{ m} \times 4 \text{ m} \times 2.6 \text{ m}$ . We varied the number of microphones, and used the real source and the 6 first-order image source for the estimation. Errors are shown for the microphone localization (left), and the image source localization (right), with and without the step size constraints. The constraints are given as bounds on the step size between  $-20\%$  and  $+20\%$  of the true value.

Assume, for instance, that the source emits pulses at regular intervals, and that we know approximately the robot’s speed. Then it is realistic to have certain bounds on the length of the robot’s “step” between two measurements. We can easily add this constraint to the

iteration. Namely, we modify the step (i) of (10) as

$$(i') \mathbf{G}^* = \arg \min \left\{ \langle \mathbf{G}, \mathbf{Z}^* \rangle + \lambda \left\| \mathbf{W} \circ (\tilde{\mathbf{D}} - \mathcal{K}(\mathbf{G})) \right\|_F^2 : \right. \\ \left. \mathbf{G} \geq 0, \mathbf{G}\mathbf{1} = \mathbf{0}, \right. \\ \left. \mathcal{K}(\mathbf{G})_{i,i+1} \in [\Delta_1^2, \Delta_2^2], 1 \leq i < M \right\}. \quad (11)$$

The new constraint defines the set of Gram matrices that satisfy the step bounds. We note that instead of adding constraints, we could include the priors in the Lagrangian form, which would allow for more sophisticated error models. Numerical experiments reveal that such constraints can significantly improve the localization accuracy.

#### 4.1. Numerical Experiments

Fig. 5 shows the error in the reconstruction of the trajectory and of the virtual sources for a robot that moves along the lemniscate

$$x(t) = 2 + \frac{1.5 \cos(t)}{1 + \sin(t)^2}, \quad y(t) = 2 + \frac{\sin(t) \cos(t)}{1 + \sin(t)^2}, \quad z(t) = 0.9, \quad (12)$$

Clearly, introducing the bounds on the robot’s step size dramatically improves the quality of the localization, even when the number of steps is very small. Introducing such simple constraints enables us to localize fewer microphones than what is theoretically possible in the minimal *unconstrained* case. Note that no knowledge about the direction of the robot was assumed, only bounds on the distance traveled.

## 5. CONCLUSION

We have shown that it is possible to efficiently reconstruct the geometry of the microphone array from echoes only, and consequently solve the room reconstruction problem without first measuring the microphone array geometry. The key is to pose the problem as a constrained optimization problem and use adequate methods for promoting low embedding dimension.

A particularly interesting finding is that the range-only SLAM can be posed and solved as a pure geometric problem. This gives an interesting alternative to the more common probabilistic solutions often involving the application of the extended Kalman filter, or of particle filters. In fact, the original convex formulation is flexible enough to allow to add constraints resembling *priors* in traditional approaches to SLAM. We have shown that with such knowledge one can obtain very accurate localization in only a few steps.

Future work involves dealing with measurement uncertainties, unknown emission times, efficient implementations, and a theoretical analysis of the algorithm performance.

<sup>3</sup>We leave a detailed analysis of Fig. 4C to a forthcoming publication.

## 6. REFERENCES

- [1] I. Dokmanić, R. Parhizkar, A. Walther, Y. M. Lu, and M. Vetterli, "Acoustic Echoes Reveal Room Shape," *Proc. Natl. Acad. Sci.*, vol. 110, no. 30, June 2013.
- [2] F. Ribeiro, D. A. Florencio, D. E. Ba, and C. Zhang, "Geometrically Constrained Room Modeling With Compact Microphone Arrays," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 20, no. 5, pp. 1449–1460, 2012.
- [3] F. Antonacci, J. Filos, M. R. P. Thomas, E. A. P. Habets, A. Sarti, P. A. Naylor, and S. Tubaro, "Inference of Room Geometry From Acoustic Impulse Responses," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 20, no. 10, pp. 2683–2695, 2012.
- [4] N. D. Gaubitch, W. B. Kleijn, and R. Heusdens, "Auto-Localization in Ad-Hoc Microphone Arrays," in Proc. *IEEE ICASSP*, pp. 106–110. Vancouver: IEEE, 2013.
- [5] R. Heusdens and N. Gaubitch, "Time-Delay Estimation for TOA-Based Localization of Multiple Sensors," in Proc. *IEEE ICASSP*, pp. 609–613. IEEE, 2014.
- [6] M. Pollefeys and D. Nister, "Direct Computation of Sound and Microphone Locations From Time-Difference-of-Arrival Data," in Proc. *IEEE ICASSP*, pp. 2445–2448. Las Vegas: IEEE, 2008.
- [7] J. M. Sachar, H. F. Silverman, and W. R. Patterson, "Microphone Position and Gain Calibration for a Large-Aperture Microphone Array," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 42–52, Jan. 2005.
- [8] H. Durrant-Whyte and T. Bailey, "Simultaneous Localization and Mapping: Part I," *IEEE Robotics & Automation Magazine*, vol. 13, no. 2, pp. 99–110, 2006.
- [9] P. Newman and J. Leonard, "Pure Range-Only Sub-Sea SLAM," *ROBOT-03*, vol. 2, pp. 1921–1926 vol.2, 2003.
- [10] J. Djugash, S. Singh, G. Kantor, and W. Zhang, "Range-only SLAM for Robots Operating Cooperatively with Sensor Networks," in Proc. *2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, pp. 2078–2084. IEEE, 2006.
- [11] G. Kantor and S. Singh, *Preliminary Results in Range-Only Localization and Mapping*. IEEE, 2002, vol. 2.
- [12] J.-L. Blanco, J.-A. Fernández-Madriral, and J. González, "Efficient Probabilistic Range-Only SLAM," in Proc. *IEEE IROS*, Nice, France, Sep 2008.
- [13] P. H. Schönemann, "On Metric Multidimensional Unfolding," *Psychometrika*, vol. 35, no. 3, pp. 349–366, 1970.
- [14] J. Dattorro, *Convex Optimization & Euclidean Distance Geometry*. Meboo, 2011.
- [15] 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, 2015.
- [16] J. B. Allen and D. A. Berkley, "Image Method For Efficiently Simulating Small-room Acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.
- [17] J. Borish, "Extension of the Image Model To Arbitrary Polyhedra," *J. Acoust. Soc. Am.*, vol. 75, no. 6, pp. 1827–1836, 1984.
- [18] I. Dokmanić, L. Daudet, and M. Vetterli, "How to Localize Ten Microphones in One Fingersnap," in Proc. *EUSIPCO*, 2014.
- [19] A. Beck, P. Stoica, and J. Li, "Exact and Approximate Solutions of Source Localization Problems," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1770–1778, 2008.
- [20] Y. Kuang, S. Burgess, A. Torstensson, and K. Astrom, "A Complete Characterization and Solution to the Microphone Position Self-Calibration Problem," in Proc. *ICASSP*, pp. 3875–3879. IEEE, 2013.
- [21] J. C. Gower, "Euclidean Distance Geometry," *Math. Sci.*, vol. 7, no. 1, pp. 1–14, 1982.
- [22] I. Dokmanić, J. Ranieri, and M. Vetterli, "Relax and Unfold: Microphone Localization with Euclidean Distance Matrices," in Proc. *EUSIPCO*, Nice, 2015.
- [23] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "Fast-SLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem," *AAAI/IAAI*, 2002.
- [24] E. Leitinger, P. Meissner, M. Lafer, and K. Witrisal, "Simultaneous Localization and Mapping using Multipath Channel Information," in Proc. *IEEE ICCW*, pp. 754–760, 2015.