STRUCTURE FROM SOUND WITH INCOMPLETE DATA

Miranda Kreković [†], *Gilles Baechler* [†], *Ivan Dokmanić* [‡], *and Martin Vetterli* [†]

[†] School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland

[‡] Coordinated Science Lab
University of Illinois at Urbana-Champaign
Urbana Illinois 61801, USA

ABSTRACT

In this paper, we consider the problem of jointly localizing a microphone array and identifying the direction of arrival of acoustic events. Under the assumption that the sources are in the far field, this problem can be formulated as a constrained low-rank matrix factorization with an unknown column offset. Our focus is on handling missing entries, particularly when the measurement matrix does not contain a single complete column. This case has not received attention in the literature and is not handled by existing algorithms, however it is prevalent in practice. We propose an iterative algorithm that works with pairwise differences between the measurements eliminating the dependence on the unknown offset. We demonstrate state-of-the-art performance both in terms of accuracy and versatility.

Index Terms— Low-rank matrix factorization, node localization, sensor array self-calibration, far field, missing and uncertain data.

1. INTRODUCTION

Many problems in acoustics, computer vision, localization and signal processing rely on low-rank matrix factorization. A famous example in computer vision is *structure from motion* (SFM)—recovery of scene geometries and camera motions from images—proposed by Tomasi and Kanade [1]. A large number of improvements of the original method have been published: some of them propose to handle missing entries [2, 3], and some focus on different camera matrices or dynamic scenes consisting of multiple motions [4–7]. These methods rely on non-linear optimization or EM-like alternation to reach a solution that is not necessarily optimal.

Thrun adapted the original SFM algorithm to the joint localization of microphones and acoustic events with unknown emission times [8], which he calls *structure from sound* (SFS). This problem has received considerable attention and many methods have been proposed, including both near-field [9–12] and far-field assumptions [13, 14]. The former leads to variants of multidimensional scaling (MDS). The idea of the latter, which is also studied in this paper, is to factorize a low-rank measurement matrix into a product of a projection and a coordinate matrix with problem-dependent constraints on the projection matrix.

A number of authors generalized Thrun's initial work. A similar setup appears in [13], with the goal of localizing passive nodes in a communication network using ambient radio or sound signals. The proposed ellipsoid time difference of arrival (TDOA) method gives a closed form solution, but it requires a set of measurements with no

missing values and exactly three receivers in the plane. The same authors in [14] also address the problem with only two receivers. In both approaches, the solution is sub-optimal if more receivers are available. A thorough study of the far-field approximation for the calibration problem is given in [15]. The authors implement the original SFS algorithm [8] in 3D and analyze its failure modes. In addition, they propose two minimization strategies—an alternating optimization and a solution relying on the Levenberg–Marquardt algorithm—and evaluate them with regard to accuracy and convergence rate.

The main shortcoming of [8] is the inability to handle missing data. However, *some* SFS algorithms can be generalized to work with a certain amount of missing entries [15], although they do not explicitly explore it. In practice, the measurements are also subject to arbitrary delays that appear as unknown column shifts in the SFS formulation and that need to be estimated jointly with the projection and coordinate matrices. Provided that the measurement matrix contains a complete column, we can use it to recenter the data points and anchor the shifts; doing so brings us back to the factorization problem with no shift, which can be solved with [15]. All the aforementioned techniques fail without at least one complete column.

Our contributions are twofold. First, we propose an iterative algorithm that can factorize a noisy and incomplete data matrix in both 2D and 3D. Second, we demonstrate that our formulation not only allows us to work in a regime where existing algorithms fail, but it also outperforms existing algorithms in terms of accuracy. In addition, we recast SFS as the problem of sampling sinusoids at unknown locations, a topic of interest in sampling theory [16].

2. PROBLEM STATEMENT

We formally define the SFS problem under the far-field assumption. For simplicity, we first analyze the 2D case; the generalization to 3D is straightforward and detailed in Section 3.1.

The idea of SFS is to localize a set of N microphones, together with a set of K acoustic events whose emission times are unknown. The far-field assumption implies that the sound propagates as a plane wave and the incident angle θ_k of a fixed acoustic signal $k = 1, \ldots, K$ is the same for all microphones $n = 1, \ldots, N$. Microphones register the absolute times of the events, denoted D_{kn} for the kth event and nth microphone. These values are arranged in a matrix $\boldsymbol{D} \in \mathbb{R}^{K \times N}$. By defining

$$\boldsymbol{A} \overset{\text{def}}{=} \begin{bmatrix} \cos\theta_1 & \ldots & \cos\theta_K \\ \sin\theta_1 & \ldots & \sin\theta_K \end{bmatrix}, \boldsymbol{X} \overset{\text{def}}{=} \begin{bmatrix} x_1 & \ldots & x_N \\ y_1 & \ldots & y_N \end{bmatrix}, \boldsymbol{c} \overset{\text{def}}{=} \begin{bmatrix} c_1 \\ \vdots \\ c_K \end{bmatrix},$$

it is easy to see that

$$\boldsymbol{D} = \boldsymbol{A}^\top \boldsymbol{X} + \boldsymbol{c}\boldsymbol{1}^\top, \tag{1}$$
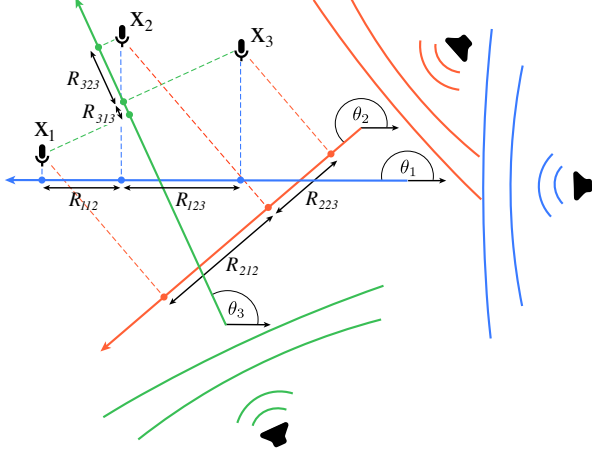
Fig. 1: Three microphones are placed at unknown locations $\boldsymbol{X}_n$. Three loudspeakers emit sound at unknown times; the incident angles θ_k of these sound waves are also unknown. We measure the time of occurrence of the same sound signal at different microphones.

where $\boldsymbol{A} \in \mathbb{R}^{2 \times K}$ is the *projection* matrix, whose columns contain unit normals of the incident waves, $\boldsymbol{X} \in \mathbb{R}^{2 \times N}$ is the *coordinate* matrix, whose columns are the microphones' positions in space, and $\boldsymbol{c}$ is an unknown *offset* vector that represents the propagation time between the emitter and an arbitrary reference point. The matrix $\boldsymbol{A}$ is uniquely determined by the angles θ_k, and vice versa. We denote by $\mathcal{U}_c$ the set of matrices with unit columns.

In mathematical terms, the SFS problem can be abstracted as the recovery of the original 2D point coordinates from a subset of their orthogonal projections on unknown lines. A particular instance of this abstraction is when the lines pass through the origin; this is equivalent to setting $\boldsymbol{c} = \boldsymbol{0}$ in (1):

$$\boldsymbol{D} = \boldsymbol{A}^\top \boldsymbol{X}. \tag{2}$$

It has been shown in [8, 15] that (1) can be reduced to (2) at the expense of losing one measurement. Indeed, one can easily eliminate $\boldsymbol{c}\boldsymbol{1}^\top$ by choosing an arbitrary microphone n to be the origin, and subtracting the corresponding column $\boldsymbol{D}_n$ from all columns in $\boldsymbol{D}$. There are a number of techniques that aim at solving (2). The traditional way is to compute the singular value decomposition (SVD) of $\boldsymbol{D}$ [8]. However, this can be done only if at least one column of $\boldsymbol{D}$ does not contain missing entries. Otherwise, the global offset $\boldsymbol{c}$ has to be estimated jointly with $\boldsymbol{A}$ and $\boldsymbol{X}$. This motivates the classification of SFS instances into three categories: 1) $\boldsymbol{D}$ has no missing entries, 2) $\boldsymbol{D}$ has missing entries, but at least one complete column, 3) $\boldsymbol{D}$ has missing entries and no complete columns.

2.1. Missing entries

The main limitation of SVD-based methods is that they require the complete measurement matrix $\boldsymbol{D}$. In practice, however, missing entries are common. For instance, it can happen that the microphones are simply too far from the sound events and the recorded peaks fall below the noise level. For this reason, we define a binary mask matrix $\boldsymbol{W}$ that indicates the known entries. In addition to missing entries, we assume that $\boldsymbol{D}$ is corrupted by noise, $\widetilde{\boldsymbol{D}} = \boldsymbol{W} \circ (\boldsymbol{D} + \boldsymbol{Z})$, where the entries of $\boldsymbol{Z}$ are independent noise realizations and $\circ$ is the Hadamard product. We can now state our problem as a low-rank

matrix factorization in the presence of noise and missing data as

$$\hat{\boldsymbol{X}}, \hat{\boldsymbol{A}} = \underset{\boldsymbol{X}, \boldsymbol{A} \in \mathcal{U}_c}{\arg \min} \|\widetilde{\boldsymbol{D}} - \boldsymbol{W} \circ (\boldsymbol{A}^\top \boldsymbol{X})\|^2. \tag{3}$$

We propose to use polar coordinates to represent the vectors in $\boldsymbol{A}$ and rewrite the problem as

$$\hat{\boldsymbol{X}}, \hat{\boldsymbol{\theta}} = \underset{\boldsymbol{X}, \theta}{\arg \min} \sum_{k=1}^{K} \sum_{n=1}^{N} \left(\widetilde{D}_{kn} - W_{kn}(x_n \cos \theta_k + y_n \sin \theta_k) \right)^2$$

$$= \underset{\boldsymbol{X}, \theta}{\arg \min} \sum_{k=1}^{K} \sum_{n=1}^{N} \left(\widetilde{D}_{kn} - W_{kn} r_n \sin(\theta_k + \alpha_n) \right)^2, \tag{4}$$

with $r_n = \sqrt{x_n^2 + y_n^2}$ and $\alpha_n = \arctan \frac{x_n}{y_n}$. It is interesting to note that this formulation relates to the problem of sampling sinusoids of unknown amplitudes and phases at unknown locations. It is a specific case of the more general problem of irregular sampling with unknown locations for bandlimited signals, studied in [16, 17].

2.2. Missing entries with unknown shift

The challenge of missing data becomes apparent when one has to estimate the unknown translation vector $\boldsymbol{c}$ jointly with the unknown points in $\boldsymbol{X}$ and unit vectors in $\boldsymbol{A}$. As mentioned in Section 2, casting (1) into (2) is possible both for complete and incomplete matrices $\boldsymbol{D}$, as long as there is at least one microphone that has registered all the events (i.e. at least one complete column of $\boldsymbol{D}$). This assumption cannot always be satisfied in practice.

To relax it, we propose to consider the pairwise differences between the columns of $\boldsymbol{D}$, since they do not depend on the translation vector $\boldsymbol{c}$. Specifically, we construct the measurement tensor $\boldsymbol{R} \in \mathbb{R}^{K \times N \times N}$ with relative distances $R_{knm} = D_{kn} - D_{km}$. When the measurements $\widetilde{\boldsymbol{D}}$ are incomplete and corrupted with noise, we obtain an incomplete and noisy tensor as well, denoted $\widetilde{\boldsymbol{R}} = \boldsymbol{V} \circ \boldsymbol{R}$. The element V_{knm} of the mask $\boldsymbol{V} \in \mathbb{R}^{K \times N \times N}$ is equal to 1 when both projections D_{kn} and D_{km} are registered, 0 otherwise. It can be expressed as the product of the two corresponding elements of the mask $\boldsymbol{W}$, $V_{knm} = W_{kn} W_{km}$. For simplicity, we adopt the convention that $V_{knn} = 0$. With this notation at hand, we formally define the problem to solve as:

Problem 1. *Given a subset of noisy observations of distances $\boldsymbol{D}$, jointly recover the points $\boldsymbol{X}$, the column-unitary matrix $\boldsymbol{A}$, and the translation vector $\boldsymbol{c}$, such that*

$$\hat{\boldsymbol{X}}, \hat{\boldsymbol{A}}, \hat{\boldsymbol{c}} = \underset{\boldsymbol{X}, \boldsymbol{A} \in \mathcal{U}_c, \boldsymbol{c}}{\arg \min} \left\| \widetilde{\boldsymbol{D}} - \boldsymbol{W} \circ \left(\boldsymbol{A}^\top \boldsymbol{X} + \boldsymbol{c}\boldsymbol{1}^\top \right) \right\|^2. \tag{5}$$

By observing the pairwise differences between the columns of $\widetilde{\boldsymbol{D}}$ given in the tensor $\widetilde{\boldsymbol{R}}$, we eliminate $\boldsymbol{c}$ and reformulate Problem 1 as

$$\hat{\boldsymbol{X}}, \hat{\boldsymbol{\theta}} = \underset{\boldsymbol{X}, \theta}{\arg \min} \sum_{k=1}^{K} \sum_{n,m=1}^{N} \left(\widetilde{R}_{knm} - V_{knm}(\Delta_{x_{nm}} \cos \theta_k \right.$$

$$\left. + \Delta_{y_{nm}} \sin \theta_k) \right)^2, \tag{6}$$

where $\Delta_{x_{nm}} = x_n - x_m$ and $\Delta_{y_{nm}} = y_n - y_m$.

3. PROPOSED ALGORITHM

The objective function defined in (6) is non-convex, and has many local minima. Moreover, given $\boldsymbol{R}$, we cannot distinguish between

---
Algorithm 1 Proposed algorithm.

Input: An incomplete noisy measurement matrix $\widetilde{D}$, the mask matrix W, convergence criterion ϵ.
Output: $\hat{X}$ and $\hat{\theta}$ that minimize (6).
  Let $\widetilde{V}_{knm} = W_{kn}W_{km}$.
  Let $\widetilde{R}_{knm} = (\widetilde{D}_{kn} - \widetilde{D}_{km})V_{knm}$.
  Randomly initialize X^0 and θ^0, where θ denotes a vector of length K with values θ_k.
  Repeat for each iteration i until $\|\hat{R}^i - \hat{R}^{i-1}\|_F^2 < \epsilon$:
    Step 1. For fixed θ^{i-1}, find X^i by solving the linear system (7).
    Step 2. For fixed X^i, find θ^i by solving (8).
    Step 3. $\hat{R}_{knm}^i = V_{knm}\left(\Delta_{x_{nm}}^i \cos\theta_k^i + \Delta_{y_{nm}}^i \sin\theta_k^i\right)$.
  return $\hat{X}, \hat{\theta}$
---

translated, rotated and reflected setups in Fig. 1; we consider these solutions to be equivalent. However, we observe that we can find the global minimizer of (6) over θ for fixed X, and analogously, the optimal solution of X for fixed θ. Therefore, we propose to solve the factorization of R by alternating between estimations of θ and X, where at each step i the matrix X^i and vector θ^i are optimized separately keeping the other one fixed.

To derive the optimal X for fixed θ, we take the first derivative of the cost function in (6) with respect to x_n and y_n, and set the resulting equations to 0. The system of equations is then written as

$$\begin{bmatrix} M^{xx} & M^{xy} \\ M^{xy} & M^{yy} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} p^x \\ p^y \end{bmatrix} \text{ or } M\begin{bmatrix} x \\ y \end{bmatrix} = p, \qquad (7)$$

where $M^{**} \in \mathbb{R}^{N \times N}$ and $p^* \in \mathbb{R}^N$; their entries are derived in Appendix A. The vector $\begin{bmatrix} x & y \end{bmatrix}^\top$ represents the flattened matrix X: the first N elements x are the values of the first row of X and the last N elements y are the values of the second row of X. Therefore, by solving the linear system in (7), we can recover X.

The second step of the algorithm estimates θ for fixed X by following the same idea as in the first step. We set the first derivatives of the cost function in (6) with respect to θ_k to 0:

$$\sin\theta_k \sum_{n,m=1}^N \widetilde{R}_{knm}\Delta_{x_{nm}} - \frac{1}{2}\sin 2\theta_k \sum_{n,m=1}^N V_{knm}(\Delta_{x_{nm}}^2 - \Delta_{y_{nm}}^2)$$
$$-\cos\theta_k \sum_{n,m=1}^N \widetilde{R}_{knm}\Delta_{y_{nm}} + \cos 2\theta_k \sum_{n,m=1}^N V_{knm}\Delta_{x_{nm}}\Delta_{y_{nm}} = 0.$$
$$(8)$$

By replacing trigonometric functions with complex exponentials and substituting $t_k = e^{j\theta_k}$, we obtain quartic polynomials in t_k. To find the global solution, we evaluate the objective function for all four roots and choose t_k that yields the smallest cost.

The algorithm stops when $\|\hat{R}^i - \hat{R}^{i-1}\|_F^2$ falls below some threshold ϵ. Once we have $\hat{X}$ and $\hat{\theta}$, it is straightforward to estimate c via least-squares. This procedure is summarized in Algorithm 1.

The proposed algorithm is an instance of more general alternating optimization problems, where the variables are partitioned in two subsets and the optimal solution is estimated for each subset separately at every iteration. Such algorithms are ubiquitous and their properties are well-studied [18, 19]. In [20], it was shown that this class of algorithms is locally, Q-linearly (i.e. quotient-linearly) convergent to some value x^*, which means that there exists $r \in (0, 1)$ such that $\frac{|x_{k+1}-x^*|}{|x_k-x^*|} \le r$ for all k sufficiently large.

3.1. Generalization to 3D space

The generalization to 3D is straightforward; the most significant modification takes place in step 2 of Algorithm 1. To ensure that $A \in \mathcal{U}_c$, we represent its columns in spherical coordinates:

$$A = \begin{bmatrix} \cos\theta_1 \sin\phi_1 & \dots & \cos\theta_K \sin\phi_K \\ \sin\theta_1 \sin\phi_1 & \dots & \sin\theta_K \sin\phi_K \\ \cos\phi_1 & \dots & \cos\phi_K \end{bmatrix}, \; X = \begin{bmatrix} x_1 & \dots & x_N \\ y_1 & \dots & y_N \\ z_1 & \dots & z_N \end{bmatrix}.$$

The minimization problem in (6) is generalized to include the angles θ and ϕ and the additional coordinate in X as

$$\underset{X,\theta,\phi}{\text{minimize}} \sum_{k=1}^K \sum_{n,m=1}^N \widetilde{R}_{knm} - V_{knm}\big(\Delta_{x_{nm}}\cos\theta_k \sin\phi_k$$
$$(9)$$
$$-\Delta_{y_{nm}}\sin\theta_k \sin\phi_k - \Delta_{z_{nm}}\cos\phi_k\big)^2.$$

Step 1 of Algorithm 1 is adjusted by fixing both vectors θ and ϕ. Then, the minimization over X is formalized as in (7), with M having an additional column and row. Their values can be computed analogously to the 2D case, by replacing the polar coordinates of A with a spherical representation. Step 2 however becomes slightly more complicated, as we need to take derivatives over two angles, θ_k and ϕ_k. We perform a similar substitution as in 2D case, but instead of obtaining one quatric equation, it results in two bivariate polynomials of degree 8.

4. EXPERIMENTAL RESULTS

In this section, we evaluate performance and versatility of our algorithm. To quantify the performance, we use the root mean squared error (RMSE) with respect to the noiseless measurement matrix D and to the original configuration X. The former, RMSE(D), quantifies the consistency of the algorithms, and the latter, RMSE(X), determines the capacity to accurately estimate the points' locations, which is typically the goal. Our experiments are conducted for the three categories defined in Section 2: 1) no missing entries, 2) missing entries with at least one complete column, 3) missing entries with no complete columns. We examine the performance of our algorithm against the most relevant existing methods for the first two categories. In practice, there is a wide range of situations falling into the third category that existing solutions cannot deal with.

4.1. Performance evaluation: the complete case

We first turn to the complete case and compare our algorithm with the SVD-based estimator proposed in [8], and the alternating optimization (AO) from [15]. The SVD-based algorithm only handles this case, while the AO method includes the second category as well. Unfortunately, none of these three methods guarantee a global minimum. Besides, the estimates obtained by AO and the proposed algorithm depend on the initial values of the point locations X and angles θ. As the convergence of both methods is within a few steps (in the order of milliseconds for about 30 measurements), we can afford to repeat the estimation several times with a different initialization to avoid local minima. In the experiment below, we run the iterative algorithms ten times and keep the solution with the lowest cost. We empirically observed that AO has a stronger tendency than our algorithm to get trapped in local minima.

We assume to have $N = 6$ microphones and $K = 5$ acoustic events. The measurement matrix D is corrupted with Gaussian noise such that the input signal-to-noise ratio (SNR) ranges from 0 to
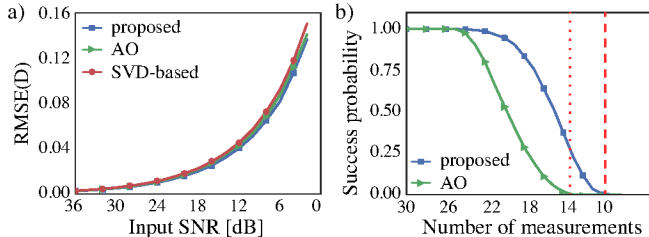
Fig. 2: a) Comparison of the proposed method with the SVD-based approach and the alternating optimization for different values of input SNR, with complete $\boldsymbol{D}$. b) Likelihood of the algorithms to work for an randomly created mask $\boldsymbol{W}$ with a given number of measurements. Each experiment is an average of 1000 realizations.

36 dB. The dependency of the RMSE($\boldsymbol{D}$) on the input SNR is shown in Fig. 2a. Our algorithm outperforms the algorithms in [8] and [15], even though [8] was tailored exclusively for complete matrices.

4.2. Performance evaluation: missing entries

Next, we analyze the algorithms in the presence of missing entries. For a successful reconstruction, the arrangement of the measurements in $\widetilde{\boldsymbol{D}}$ has to satisfy several conditions. As mentioned, AO needs a complete column, while our difference-based algorithm obviously does not work with less than two entries per row. Moreover, both algorithms require a *connected* $\widetilde{\boldsymbol{D}}$. This means that there must exist a path between any two entries $\widetilde{D}_{kn}$ and $\widetilde{D}_{lm}$, where the matrix entries are said to be connected if they share a common column or a row. For a given number of measurements M, we repeatedly create random masks $\boldsymbol{W}$ with M non-zero entries and verify if the matrices $\widetilde{\boldsymbol{D}}$ satisfy the above conditions. This gives us an estimate of the likelihood of the algorithms to work with M entries, as illustrated in Fig. 2b. In addition, we observe that the smallest possible M for our algorithm is $2K$, or equivalently two entries per row. Similarly, the threshold for AO is $3K - 1$, since with less measurements it is impossible to construct a connected $\widetilde{\boldsymbol{D}}$ with a complete column. These thresholds separate success from failure modes for our algorithm and AO, and are indicated in Fig. 2b with dashed and dotted lines, respectively.

In Fig. 3, we compare the performance of the two algorithms for varying levels of noise and number of measurements. Again, our method achieves lower RMSE($\boldsymbol{D}$) and RMSE($\boldsymbol{X}$) for all combinations of input SNR and missing entries. Fig. 3a also suggests that the reconstruction is stable and the RMSE($\boldsymbol{D}$) increases with the amount of noise in the same manner as in Fig. 2. Interestingly, we notice that the RMSE($\boldsymbol{D}$) values are not affected by the number of missing entries; indeed, they remain fairly constant along the x-axis. This is not the case for the RMSE($\boldsymbol{X}$), which grows significantly as the number of measurements decreases. In other words, even though it is always possible to find a low-rank matrix $\hat{\boldsymbol{D}}$ that is consistent with the observed measurements, it is not guaranteed that its factorization is close in any way to the original point configuration; this is especially true as $\widetilde{\boldsymbol{D}}$ becomes sparser. What's more, we observe that both algorithms can sometimes yield an estimate that is far from the original points, causing outliers in RMSE($\boldsymbol{X}$). In Fig. 3b we discarded the experiments resulting in an RMSE($\boldsymbol{X}$) larger than a chosen threshold value $\zeta = 1$. The success of the reconstruction is depicted by transparency, where a high failure rate translates to a more transparent value. In Fig. 3a, the reconstruction is considered successful when $\widetilde{\boldsymbol{D}}$ is connected, as explained earlier and depicted in Fig. 2b. To as-

(a) RMSE($\boldsymbol{D}$)

(b) RMSE($\boldsymbol{X}$)
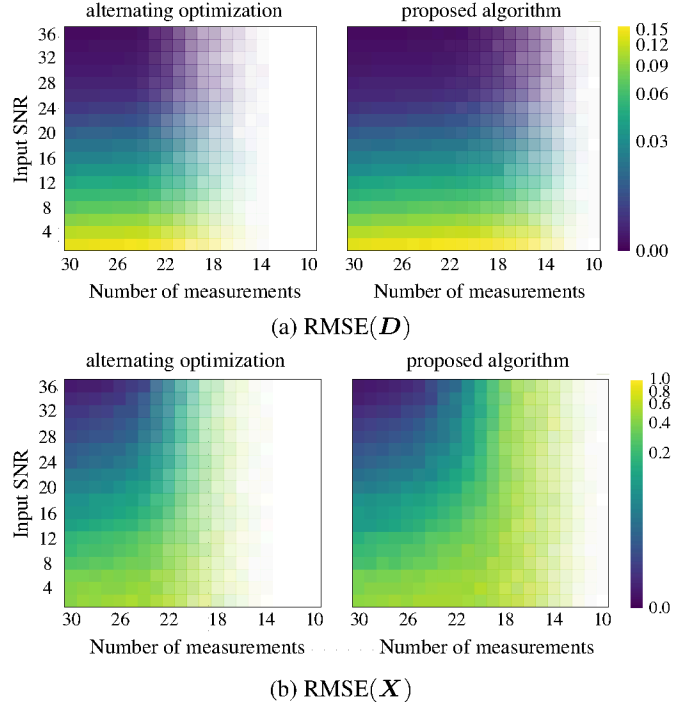
Fig. 3: Performance of the proposed method and the alternating optimization for different input SNR and number of missing entries. Reported errors are averaged over 300 realizations.

sess the success of Fig. 3b, we also take into account the percentage of outliers. We observe that our approach surpasses AO with respect to RMSE($\boldsymbol{D}$) and RMSE($\boldsymbol{X}$) both in performance and versatility.

5. CONCLUSION

We have proposed an iterative algorithm for low-rank matrix factorization with uncertain and missing data. Extensive numerical simulations show that our method outperforms existing solutions, while at the same time allowing larger number of missing entries in the measurement matrices. This creates opportunities for more robust solutions relying on self-calibrating sensor arrays, as well as indoor localization based on ambient sound signals. Ongoing research studies the uniqueness of the reconstruction, as well as the relationship between consistency and accurate recovery of the points.

A. APPENDIX

The entries of the matrices $\boldsymbol{M}$ and vectors $\boldsymbol{p}$ from (7) are given as:

$$M_{nm}^{xx} = \begin{cases} \sum_{k=1}^{K} \sum_{m=1}^{N} \cos\theta_k{}^2 V_{knm} & \text{if } n = m \\ -\sum_{k=1}^{K} \cos\theta_k{}^2 V_{knm}, & \text{otherwise} \end{cases},$$

$$M_{nm}^{yy} = \begin{cases} \sum_{k=1}^{K} \sum_{m=1}^{N} \sin\theta_k{}^2 V_{knm} & \text{if } n = m \\ -\sum_{k=1}^{K} \sin\theta_k{}^2 V_{knm} & \text{otherwise} \end{cases},$$

$$M_{nm}^{xy} = M_{mn}^{yx} = \begin{cases} \sum_{k=1}^{K} \sum_{m=1}^{N} \cos\theta_k \sin\theta_k V_{knm} & \text{if } n = m \\ -\sum_{k=1}^{K} \cos\theta_k \sin\theta_k V_{knm} & \text{otherwise} \end{cases},$$

$$p_n^x = \sum_{k=1}^{K} \sum_{m=1}^{N} \widetilde{R}_{knm} V_{knm} \cos\theta_k, \quad p_n^y = \sum_{k=1}^{K} \sum_{m=1}^{N} \widetilde{R}_{knm} V_{knm} \sin\theta_k.$$

6. REFERENCES

[1] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorization method," *International Journal of Computer Vision*, vol. 9, no. 2, pp. 137–154, 1992.

[2] D. Jacobs, "Linear fitting with missing data: Applications to structure-from-motion and to characterizing intensity images," *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 206–212, 1997.

[3] R. Hartley and F. Schaffalitzky, "Powerfactorization: 3D reconstruction with missing or uncertain data," *Australia-Japan Advanced Workshop on Computer Vision*, vol. 74, pp. 76–85, 2003.

[4] R. Vidal and R. Hartley, "Motion segmentation with missing data using powerfactorization and GPCA," *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 310–316, 2004.

[5] A. M. Buchanan and A. W. Fitzgibbon, "Damped newton algorithms for matrix factorization with missing data," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 316–322, 2005.

[6] J. Yan and M. Pollefeys, "A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate," *European Conference on Computer Vision*, pp. 94–106, 2006.

[7] L. Torresani, A. Hertzmann, and C. Bregler, "Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors," *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 5, pp. 878–892, 2008.

[8] S. Thrun, "Affine structure from sound," *Advances in Neural Information Processing Systems*, pp. 1353–1360, 2006.

[9] N. Ono, H. Kohno, N. Ito, and S. Sagayama, "Blind alignment of asynchronously recorded signals for distributed microphone array," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 161–164, 2009.

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

[11] N. D. Gaubitch, W. B. Kleijn, and R. Heusdens, "Auto-localization in ad-hoc microphone arrays," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 106–110, 2013.

[12] I. Dokmanić, L. Daudet, and M. Vetterli, "From acoustic room reconstruction to SLAM," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6345–6349, 2016.

[13] J. Wendeberg, T. Janson, and C. Schindelhauer, "Self-localization based on ambient signals," *Theoretical Computer Science*, vol. 453, pp. 98–109, 2012.

[14] T. Janson, C. Schindelhauer, and J. Wendeberg, "Self-localization application for iPhone using only ambient sound signals," *IEEE International Conference on Indoor Positioning and Indoor Navigation*, pp. 1–10, 2010.

[15] Y. Kuang, E. Ask, S. Burgess, and K. Åström, "Understanding TOA and TDOA network calibration using far field approximation as initial estimate." *International Conference on Pattern Recognition Applications and Methods*, pp. 590–596, 2012.

[16] P. Marziliano and M. Vetterli, "Reconstruction of irregularly sampled discrete-time bandlimited signals with unknown sampling locations," *IEEE Transactions on Signal Processing*, vol. 48, no. 12, pp. 3462–3471, 2000.

[17] M. Pacholska, B. B. Haro, A. Scholefield, and M. Vetterli, "Sampling at unknown locations, with an application in surface retrieval," *IEEE International Conference on Sampling Theory and Applications*, pp. 364–368, 2017.

[18] J. C. Bezdek and R. J. Hathaway, "Some notes on alternating optimization," *AFSS*, vol. 2, pp. 288–300, 2002.

[19] J. Gorski, F. Pfeuffer, and K. Klamroth, "Biconvex sets and optimization with biconvex functions: a survey and extensions," *Mathematical Methods of Operations Research*, vol. 66, no. 3, pp. 373–407, 2007.

[20] J. C. Bezdek, R. J. Hathaway, R. E. Howard, C. A. Wilson, and M. P. Windham, "Local convergence analysis of a grouped variable version of coordinate descent," *Journal of Optimization Theory and Applications*, vol. 54, no. 3, pp. 471–477, 1987.