Offset estimation for microphone localization using alternating projections

Simayijiang Zhayida, Fredrik Andersson and Kalle Åström
Centre for Mathematical Sciences, Lund University, Sweden
E-mail: {zhayida, fa, kalle}@maths.lth.se

***Abstract*—In this paper, we focus on solving the time delay as a separate problem to the reconstruction of the microphone and sound locations. The time delay estimation appears as one of the main steps in sensor calibration problem, once the time delays are known or estimated, we can solve the time-difference-of-arrival problems by converting them to time-of-arrival problems. In this paper we make use of an alternating projection approach for estimating time delays. We show both for simulated and real data that alternative projecting algorithm yields good estimates of the time delays, even when provided with bad initial estimations. The method is also easy to implement and comparatively fast.**

I. Introduction

The problem of sensor network self-calibration is essential for localization and navigation, these are an important area that has attracted significant research interests. Among various existing location estimation approaches, the range-based schemes, TOA and TDOA are proved to have a very good accuracy due to the high time resolution of the signals. Although such problems have been studied extensively in the literature in the form of localization of e.g. a sound source using a calibrated detector array, see, e.g. [1]–[4], the problem of self-calibration of a sensor array is still an open problem.

Time delay occurs in the capturing of sound before it reaches the recorder, it causes the center of the recording waveform to not be at 0, but at a higher value. Time delay estimation has been studied extensively, and several previous contributions of estimating the time delay problem rely on a set of rank constraints to determine the unknown offsets. In [5] a different constraint is used, this makes it possible to solve for the time delays for at least 10 receivers and at least 5 transmitters. In [6] present two techniques for solving the unknown offset. The first scheme is an improved version of the linear factorization in [5]. Another one is to make full use of the rank constraints to the distance matrix that gives a novel formulation. In [7], time delays are recovered by solving a truncated nuclear norm minimization problem using the alternating direction method of multipliers (ADMM). Once time delays are calculated, then it can be used as an initial solution for converting TDOA measurements to TOA measurements.

In [8], have developed an automatic system for microphone self-localization using ambient sound. The system is based on a first finding several time-difference matching vectors for the recording. These are then used as input to robust geometric algorithms based on minimal solvers and RANSAC to provide initial estimates of the time delay. This paper appear as a improvement of time delay estimation step of [8]. Here we study the time delay estimation of the TDOA network calibration problem for general dimensions. We implement alternating projection algorithm for estimation of the time delays. The method of alternating projections finds a point in the intersection of two manifolds by iteratively projecting a point onto one set and then the other. Popular because of its simplicity and intuitive appeal, the method has been rediscovered many times in the literature. The proposed algorithm tested on synthetic and real data, and experiments shows that alternating projection algorithm is more accurate, fast and stable method for time delay estimation.

II. Time delay estimation for time-difference-of-arrival network

For a sensor network with M receivers and N transmitters, we denote the spatial coordinated of the receivers and transmitters by $\mathbf{m}_i$ for $i = 1, 2, ..., M$ and $\mathbf{s}_j$ for $j = 1, 2, ..., N$. For the TDOA problem, the receivers are synchronized, but transmitters are not. Arrival time instances t_{ij} of signals are measured time difference between departing from sound $\mathbf{s}_j$ and arriving at the microphone $\mathbf{m}_i$ and each transmitter $\mathbf{s}_j$ has clock offset q_j (time of recording), then we have the following model between the positions and measurements.

$$||\mathbf{m_i} - \mathbf{s_j}|| = c(t_{ij} - q_j) = ct_{ij} - o_j, \tag{1}$$

where $|| \cdot ||$ is Euclidean norm, c is the speed of sound, assumed to be known and constant.

In [8], we have introduced matching vector u_{ij} which is time matchings of signals in different channels at some time instant. In this section, we introduce the idea of using an alternating projection method to determine the time delays $\mathbf{o}$ from the matching vectors (or relative distance measurements) such that $u_{ij} = ||\mathbf{m_i} - \mathbf{s_j}|| + o_j$.

A. Time matching vector and offset

Let $(\mathbf{x_1}, \dots, \mathbf{x_M})$ be sound recordings with M channels and microphones are at unknown positions $(\mathbf{m_1}, \dots, \mathbf{m_M})$. We assume that among the sounds there are one or several possibly moving sound sources and sounds occurs at unknown positions $(\mathbf{s_1}, \dots, \mathbf{s_N})$. This means that at several time instances along the sound channels there are one or several matchings. Each such match corresponds to a set of time instants of arrival times to the microphones. At one sound instant we have one offset value, let j be used as an index for different sound instants and each time vector $(t_{1j}, \dots, t_{Mj})$ correspond to a sound made at instant t_{0j} at 3D position $\mathbf{s_j}$ that fulfilling

$$c(t_{ij} - t_{0j}) = ||\mathbf{m_i} - \mathbf{s_j}||.$$

Without loss of generality, we will in the sequel assume that all time differences are measured against channel 1. We introduce $u_{ij} = c(t_{ij} - t_{1j})$, which can be interpreted as

$$u_{ij} = c(t_{ij} - t_{0j}) - c(t_{1j} - t_{0j}) = ||\mathbf{m_i} - \mathbf{s_j}|| - ||\mathbf{m_1} - \mathbf{s_j}||. \tag{2}$$

Also introduce $o_j = c(t_{1j} - t_{0j}) = ||\mathbf{m_1} - \mathbf{s_j}||$ as the offset. This can be interpreted as the distance from the sound to the microphone 1. Using this notation the measurement equation (2) becomes

$$u_{ij} = ||\mathbf{m_i} - \mathbf{s_j}|| - o_j. \tag{3}$$

If we have p column of matching vectors, then offset $\mathbf{o}$ can be written as

$$\mathbf{o}=\begin{pmatrix}o_1\\o_2\\\vdots\\o_p\end{pmatrix}=\begin{pmatrix}c(t_{11}-t_{01})\\c(t_{12}-t_{02})\\\vdots\\c(t_{1p}-t_{0p})\end{pmatrix}=\begin{pmatrix}||\mathbf{m_1}-\mathbf{s_1}||\\||\mathbf{m_1}-\mathbf{s_2}||\\\vdots\\||\mathbf{m_1}-\mathbf{s_p}||\end{pmatrix}.$$

B. Alternating projection algorithm

Let $\mathcal{K}$ be a finite dimensional Hilbert space over $\mathbf{R}$ and let $\mathcal{M}$ be a manifold. Suppose $\mathcal{M}_1$ and $\mathcal{M}_2$ are two manifolds in $\mathbf{R}^n$, our goal is to find the intersection point $x\in\mathcal{M}_1\cap\mathcal{M}_2$, i.e. given initial point x_0 find $x_{j+1}=P_{\mathcal{M}_2}(P_{\mathcal{M}_1}(x_j))$, here $P_{\mathcal{M}_1}$ and $P_{\mathcal{M}_2}$ denote projection on $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively. It is not certain that there are multiple intersection points, nor certain that there are any at all. The purpose of the method is to find an intersection point as close as possible to the starting point. If there is no intersection point, one can hope it jumps back and forth between the manifolds, i.e., that the sequence containing every point converges, and that this convergence point is close to the original. The algorithm starts with any $x_0\in\mathcal{M}_2$, and then alternately projects onto $\mathcal{M}_1$ and $\mathcal{M}_2$:

$$y_k=P_{\mathcal{M}_1}(x_k),\quad x_{k+1}=P_{\mathcal{M}_2}(y_k),\quad k=0,1,2,... \tag{4}$$

Under suitable conditions, this generates a sequence of points $x_k\in\mathcal{M}_1$ and $y_k\in\mathcal{M}_2$. If $\mathcal{M}_1\cap\mathcal{M}_2\neq\emptyset$, then the sequence x_k and y_k both converge to a point $x^*\in\mathcal{M}_1\cap\mathcal{M}_2$.

Alternating projection schemes for non-linear subsets have been used in a number of applications, for instance, $\mathcal{K}$ can be the set of $m\times n$-matrices and the set $\mathcal{M}_j$ be subsets with a certain structure, e.g. matrices with a certain rank, self-adjoint matrices etc. For a detailed overview of optimization methods on matrix manifolds, see [9]. Much emphasis has been put towards the use of alternating projections for the case of convex sets $\mathcal{M}_1$ and $\mathcal{M}_2$, however, for non-convex sets the field remained rather undeveloped until the 90's. One of the first attempt at dealing with the method of alternating projections for non-convex sets were made in [10], recent results for manifolds are given in [11], [12]. The manifolds studied in this paper falls under the framework of the manifolds discussed in [12]. It described theory for larger class of manifolds, under appropriate conditions, it proved not only the sequence of alternating projection converges, but then the limit point is fairly close to optimal point, in a manner to the distance between starting point and optimal point is not too far.

If we set $D_{ij}=||\mathbf{m_i}-\mathbf{s_j}||^2=d_{ij}^2$, then we have $D_{ij}=||\mathbf{m_i}||^\mathbf{2}-2\langle\mathbf{m_i},\mathbf{s_j}\rangle+||\mathbf{s_j}||^\mathbf{2}$ and it also can be written as $D_{ij}=(u_{ij}+o_j)^2$. By constructing the vectors $\mathbf{g_i}=(\|\mathbf{m_i}\|^\mathbf{2}\quad\mathbf{m_i}\quad\mathbf{1})^\mathbf{T}$, $\mathbf{h_j}=(1\quad-2\mathbf{s_j}\quad\|\mathbf{s_j}\|^\mathbf{2})^T$ and collecting them into matrices $\mathbf{G}$ and $\mathbf{H}$, we have $\mathbf{D}=\mathbf{G^TH}$. It is a matrix with elements $D_{ij}=\sum_{k=1}^{5}\mathbf{g_i}(k)\mathbf{h_j}(k)$, and matrix $\mathbf{D}$ has at most rank 5. By simple manipulations, it is possible to eliminate the terms $\|\mathbf{s_j}\|$, and in that way construct a rank-4 formulation, i.e.,

$$D_{ij}-D_{1j}=||\mathbf{m_i}||^2-||\mathbf{m_1}||^2+2\langle\mathbf{m_1}-\mathbf{m_i},\mathbf{s_j}\rangle. \tag{5}$$

$$D_{ij}-D_{1j}=(u_{ij}^2-u_{1j}^2)+2o_j(u_{ij}-u_{1j}). \tag{6}$$

From equation (5), we see that the matrix with elements $D_{ij}-D_{1j}$ has rank 4. Hence, if $A_{ij}=u_{ij}^2-u_{1j}^2$ and $B_{ij}=u_{ij}-u_{1j}$, then

$$\hat{\mathbf{D}}=\mathbf{A}+\mathbf{2diag(o)B}. \tag{7}$$

is a rank 4 matrix. Here $\hat{\mathbf{D}}$ is a matrix with size $N\times M$ and the first row is all zeros, N and M are for a number of sound and microphone, respectively. More precisely, it is

$$\hat{\mathbf{D}}=\begin{pmatrix}D_{11}-D_{11} & D_{12}-D_{12} & \cdots & D_{1M}-D_{1M}\\ D_{21}-D_{11} & D_{22}-D_{12} & \cdots & D_{2M}-D_{1M}\\ \vdots & \vdots & \ddots & \vdots\\ D_{N1}-D_{11} & D_{N2}-D_{12} & \cdots & D_{NM}-D_{1M}\end{pmatrix}.$$

The condition that the matrix $\hat{D}$ has rank four gives constraints on which offsets $\mathbf{o}$ that are consistent with our model, and hence it provides a way to estimate the offsets $\mathbf{o}$. Our aim is to find $\mathbf{o}$ such that $rank(\mathbf{A}+2diag(\mathbf{o})\mathbf{B})=4$, i.e. $o_j\in\mathcal{M}_1\cap\mathcal{M}_2$, where $\mathcal{M}_1$ is rank four matrices and $\mathcal{M}_2$ is set of matrices that can be written as $\mathbf{A}+diag(\mathbf{o})\mathbf{B}$ for a vector $\mathbf{o}$. More precisely, if $\mathbf{A}=\sum_{i=j}\mathbf{u}_j\sigma_j\mathbf{v}_j^T$, then by the Eckart-Young theorem, it follows that the best rank four approximation of $\mathbf{A}$ is obtained by including only the singular vector corresponding to the four largest singular values, i.e.,

$$P_{\mathcal{M}_1}(A)=\sum_{i=1}^{4}\mathbf{u}_i\sigma_i\mathbf{v}_i^T.$$

The algorithm starts with an arbitrary value for $\mathbf{o}_0$ and then generate the sequence $\mathbf{o_{j+1}}=P_{\mathcal{M}_2}(P_{\mathcal{M}_1}(\mathbf{o_j})), j=1,2,...,n$, n is the number of iterations. If we say $\mathbf{C}\approx\mathbf{A}+2diag(\mathbf{o})\mathbf{B}$, this can be written as

$$\begin{pmatrix}o_1 & & & \\ & o_2 & & \\ & & \ddots & \\ & & & o_N\end{pmatrix}\mathbf{B}_{N\times M}=\mathbf{C}_{N\times M}-\mathbf{A}_{N\times M}$$

where N and M is number of sound and microphones, respectively. Each offset can be written $2o_k\mathbf{B}(k,:)=\mathbf{C}(k,:)-\mathbf{A}(k,:)$ and we get $o_k=\frac{\sum B(C-A)}{\sum B^2/2}$.

Due to the noise, the manifolds in this problem do not intersect, hence the alternating projection method will not always converge, as it may oscillate between the manifolds. The sequences obtained by taking every other element seems, however, to converge in the tests conducted in the paper, and the method works pretty well for the offset estimation problem.

III. Experimental Validation

In this section, we present experimental results of our method on synthetic data, and we use real data for comparing it with a RANSAC based algorithm which appear as an offset calculation step of the system in [8].

A. Synthetic data

It is of interest to see the performance of the method on synthetic data regarding speed of convergence, accuracy and its sensitivity to noise. We simulate the position of the transmitters and receivers by drawing independently from the standard normal distributions, i.e. $m_i\sim\mathcal{N}(0,1)$ and $s_j\sim\mathcal{N}(0,1)$ for $i=1,2,...,M$ and $j=1,2,...,N$.

For the purpose of illustration, we use pretty bad initial guesses for $\mathbf{o}$, e.g., $\mathbf{o}_0=\mathbf{1}_{N\times1}$. Despite this, the method performs well, which indicates that it is fairly robust to initial guesses even though the underlying problem is non-convex. From Figure 1, we can see that estimated offset value did converge well to ground truth, which is calculated by $\mathbf{o}=||\mathbf{m_1}-\mathbf{s}||$.

B. Comparison with other method using real data

We have made several experiments with 8 microphones (Shure SV100). These are connected to an audio interface (M-Audio Fast Track Ultra 8R) connected to a laptop. The 8 sound channels

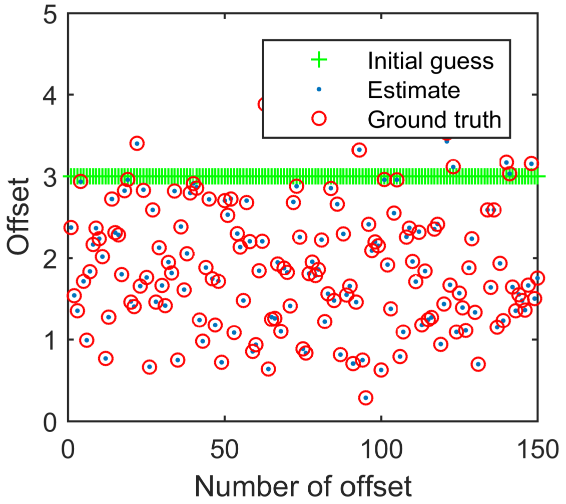

Fig. 1: Offset estimation with synthetic data. The x- and y-axis represents the number of offsets and the offset values, respectively.

were sampled at 96000 Hz. Our real data collected by two sets of experiments in which the eight microphones are placed so that they span 3D space and the sound source is moving slowly through the room. Experiment 1, part of a choir song played by a mobile phone through a small speaker. Experiment 2, part of a punk song played by a mobile phone through a small speaker and the sound source path also goes through the microphone cluster. We assume the speed of sound $c = 343m/s$, in room temperature.

As a starting point, we need to know matching vectors, we use matching algorithm in [8] for finding matching vectors of these two sets of data, the matching algorithm produces 110 and 266 matching vectors, separately. It also includes missing values due to the fact that there are no matches found at some time instants between channels. Matching vectors has size of $U_{8\times110}$ (or $U_{8\times266}$) and each row represent matching of 110 (or 266) time instances between channel 1 and channel $i = 1, 2, ..., 8$. Each entry u_{ij} is shifted values from channel 1 to channel i at time instant t_j. More precisely, we have

$$\mathbf{U_{8\times p}} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ u_{21} & u_{22} & \cdots & u_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ u_{81} & u_{82} & \cdots & u_{8p} \end{pmatrix}, \text{ here } p = 110 \text{ (or 266).}$$

For the alternating projection algorithm, we first remove missing data from the matching vectors, and remove columns if there are missing values included in matching vectors. After this, we end up with 60 (or 138) matches. Then set the number of iterations and initial values for estimate offset, therefor we get 60 (or 138) offsets. We use the microphone and sound positions calculated in [8] for calculating our ground truth offset. In the end, errors are calculated by computing the mean of $|\mathbf{o_t} - \mathbf{o}|$, here $|\cdot|$ is absolute value.

The two algorithms were applied to two data sets. The performance is illustrated in Table I and Figure 2. We note that alternating projection algorithm provides a better fit to the true offset values comparable to the RANSAC based algorithm. In this example, it is three times as accurate, and moreover, it is faster than the RANSAC based algorithm. Figure 3 shows one extreme case of where the RANSAC method gives pretty bad estimation results, yielding a mean error of 0.5028 on a data set I, while a mean error of 0.1459 is obtained by the alternating projection algorithm. The top panels of

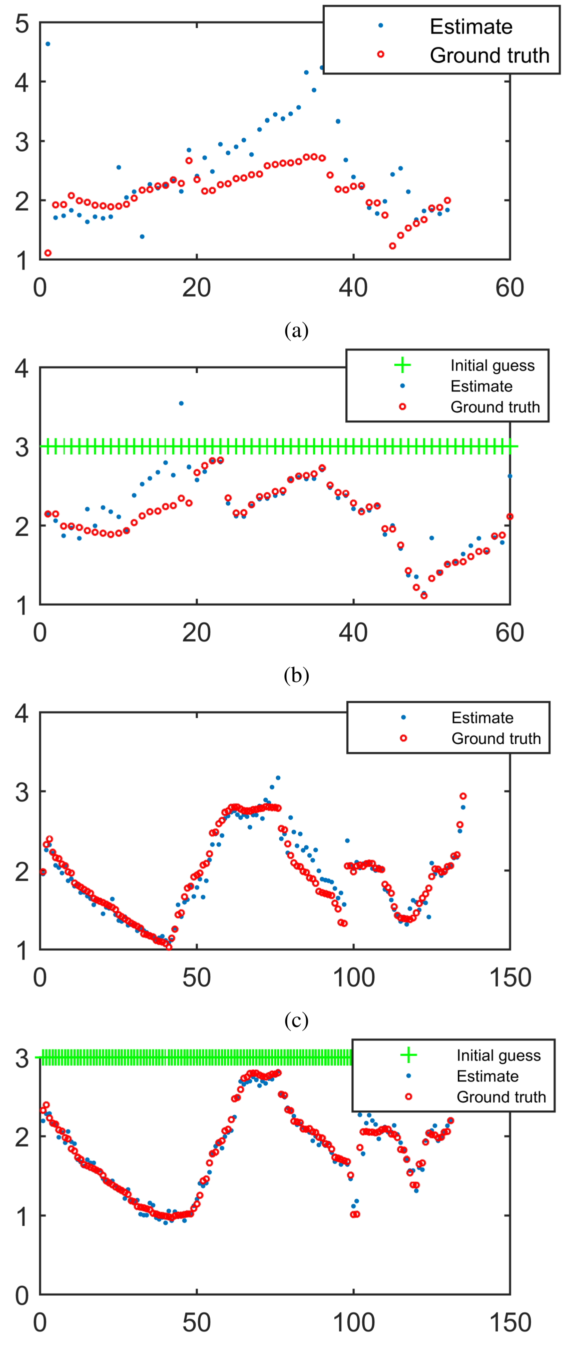

Fig. 2: Offset estimation with real data using two different methods on two different data sets. (a) and (c) are the result of using RANSAC based algorithm for offset estimation using data set I and II, respectively. (b) and (d) are the results of using alternating projection algorithm presented in this paper. Here the x- and the y-axis corresponds to the number of offsets and offset values (in meter), respectively.

	RANSAC based algorithm			
	# of matches	# of offsets	error	t (sec)
Data I	110	80	0.5358	3.1289
Data II	266	176	0.1052	4.8381
	Alternating projection algorithm			
	# of matches	# of offsets	error	t (sec)
Data I	110	60	0.1459	0.8752
Data II	266	138	0.0645	2.4880

TABLE I: Comparison of two different algorithms for single run. The alternating projection algorithm used a fixed number of iterations (10000). For the RANSAC based algorithm the number of offsets and errors are different in each single run.

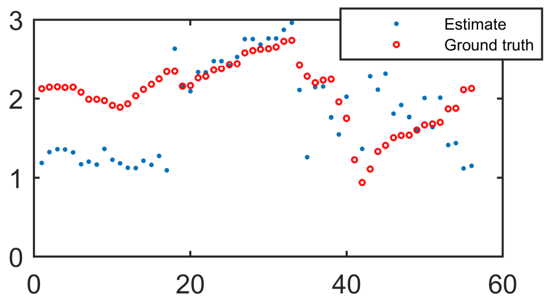

Fig. 3: One of the worst results of offset estimation using RANSAC based algorithm with data set I, in which mean error is 0.5028, elapsed time is 3.9628 sec. Here the x- and the y-axis corresponds to the number of offsets and offset values (in meter), respectively.

Figure 4 show the mean errors for both data sets, we see that the alternating projection algorithm performs substantially better in terms of the quality of the estimates. The bottom panels of Figure 4 shows timing for the two methods. We see that the alternating projection algorithm is substantially faster than the RANSAC algorithm. There is a trade-off between the quality and the speed related to the number of iterations used in the alternating projection method. However, we see that the alternating projection method will typically be superior with regards to both aspects. As an explanation to this, it seems that the alternating projection formulation seems less sensitive to local minima.

IV. Conclusion

In this paper, we study the problem of determining the unknown time delays in the TDOA self-calibration problem. We propose to use a rank constraint formulation in combination with an alternating projection method to estimate offsets. We show experimentally that this method gives a pretty good estimation for the time delays even we start with a bad initial estimation point. Our experimental comparison of the proposed method with the RANSAC based algorithm in [8] shows that it performs better both in terms of the quality of the estimates and in computational speed. Moreover, it is much easier to implement.

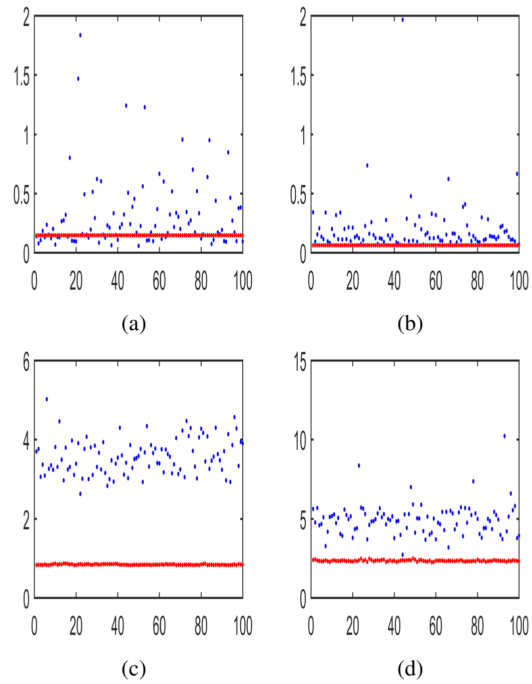

Fig. 4: Mean error and elapsed time plots for 100 executions of the two algorithms on data set I and II, blue is for RANSAC based algorithm, red for alternating projection algorithm. Mean errors for RANSAC based and the alternating projection algorithms in (a) are 0.3199 and 0.1459 for a data set I, in (b) are 0.1972 and 0.0645 for a data set II, respectively. Mean elapsed time in (c) are 3.5759 and 0.8433 for a data set I, in (d) are 4.9358 and 2.3530 for a data set II, respectively. Here the x-axis corresponds to the number of execution of the algorithm, the y-axis in (a)-(b) indicate mean error (in meter), while the y-axis in (c)-(d) indicate elapsed time (in second).

References

[1] M.S. Brandstein, J.E. Adcock, and H.F. Silverman, "A closed-form location estimator for use with room environment microphone arrays," *Speech and Audio Processing, IEEE Transactions on*, vol. 5, no. 1, pp. 45 –50, Jan. 1997.

[2] A. Cirillo, R. Parisi, and A. Uncini, "Sound mapping in reverberant rooms by a robust direct method," in *Acoustics, Speech and Signal Processing, IEEE International Conference on*, April 2008, pp. 285 – 288.

[3] M. Cobos, A. Marti, and J.J. Lopez, "A modified SRP-PHAT functional for robust real-time sound source localization with scalable spatial sampling," *Signal Processing Letters, IEEE*, vol. 18, no. 1, pp. 71 – 74, Jan. 2011.

[4] Hoang Do, H.F. Silverman, and Ying Yu, "A real-time SRP-PHAT source location implementation using stochastic region contraction(SRC) on a large-aperture microphone array," in *ICASSP 2007*, April 2007, vol. 1, pp. 121–124.

[5] M. Pollefeys and D. Nister, "Direct computation of sound and microphone locations from time-difference-of-arrival data," in *Proc. of ICASSP*, 2008.

[6] Yubin Kuang and Kalle Åström, "Stratified sensor network self-calibration from tdoa measurements," in *21st European Signal Processing Conference 2013*, 2013.

[7] F. Jiang, Y. Kuang, and K. Åström, "Time delay estimation for tdoa self-calibration using truncated nuclear norm," in *Proc. of ICASSP*, 2013.

[8] Simayijiang Zhayida, Fredrik Andersson, Yubin Kuang, and Kalle Åström, "An automatic system for microphone self-localization using ambient sound," in *22st European Signal Processing Conference*, 2014.

[9] P. A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithm on matrix manifolds*, Princeton University Press, 2008.

[10] P. L. Combettes and H. J. Trussell, "Method of successive projections for finding a common point of sets in metric spaces," *Journal of optimization theory and application*, vol. 67, no. 487-507, 1990.

[11] A. S. Lewis and J. Malick, "Alternating projections on manifolds," *Mathematics of operations research*, vol. 33, no. 1, 2008.

[12] F. Andersson and M. Carlsson, "Alternating projections on non-tangential manifolds," *Constructive approximation*, vol. 38, no. 10, 2013.