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 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 \( \mathbf{r}_i \) for \( i = 1, 2, \ldots, M \) and \( \mathbf{s}_j \) for \( j = 1, 2, \ldots, N \). 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 \( s_j \) and arriving at the microphone \( \mathbf{r}_i \) and each transmitter \( s_j \) has clock offset \( q_j \), (time of recording), then we have the following model between the positions and measurements.

\[
\| \mathbf{r}_i - \mathbf{s}_j \| = c(t_{ij} - q_j) = ct_{ij} - o_j, \tag{1}
\]

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 \( o \) from the matching vectors (or relative distance measurements) such that \( u_{ij} = \| \mathbf{r}_i - \mathbf{s}_j \| + o_j \).

A. Time matching vector and offset

Let \( (x_1, \ldots, x_M) \) be sound recordings with M channels and microphones are at unknown positions \( (x_1, \ldots, x_M) \). We assume that among the sounds there are one or several possibly moving sound sources and sources occurs at unknown positions \( (s_1, \ldots, s_N) \). 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 \( (t_{ij_1}, \ldots, t_{ij_j}) \) correspond to a sound made at instant \( t_{ij} \) at 3D position \( s_j \) that fulfilling

\[
c(t_{ij} - t_{ij}) = \| \mathbf{r}_i - \mathbf{s}_j \|.
\]

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

\[
u_{ij} = c(t_{ij} - t_{ij}) = c(t_{ij} - t_{ij}) = \| \mathbf{r}_i - \mathbf{s}_j \| - \| \mathbf{r}_1 - \mathbf{s}_j \|. \tag{2}
\]

Also introduce \( o_j = c(t_{ij} - t_{ij}) = \| \mathbf{r}_1 - \mathbf{s}_j \| \) 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

\[
u_{ij} = \| \mathbf{r}_1 - \mathbf{s}_j \| - o_j. \tag{3}
\]

If we have p column of matching vectors, then offset \( o \) can be written as
Let $K$ be a finite dimensional Hilbert space over $\mathbb{R}$ and let $M$ be a manifold. Suppose $M_1$ and $M_2$ are two manifolds in $\mathbb{R}^n$, our goal is to find the intersection point $x \in M_1 \cap M_2$, i.e. given initial point $x_0$ find $x_{k+1} = P_{M_2}(P_{M_1}(x_k))$, here $P_{M_1}$ and $P_{M_2}$ denote projection on $M_1$ and $M_2$, 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 $x_0 \in M_2$, and then alternately projects onto $M_1$ and $M_2$: 

$$y_k = P_{M_1}(x_k), \quad x_{k+1} = P_{M_2}(y_k), \quad k = 0, 1, 2, \ldots \quad (4)$$

Under suitable conditions, this generates a sequence of points $x_k \in M_1$ and $y_k \in M_2$. If $M_1 \cap M_2 \neq \emptyset$, then the sequence $x_k$ and $y_k$ both converge to a point $x^* \in M_1 \cap M_2$.

Alternating projection schemes for non-linear subsets have been used in a number of applications, for instance, $K$ can be the set of $m \times n$-matrices and the set $M_f$ 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 $M_1$ and $M_2$, 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, 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 $D_{ij} = \|x_i - s_j\|^2 = d_{ij}$, then we have $D_{ij} = \|x_i\|^2 - 2\langle x_i, s_j \rangle + \|s_j\|^2$ and it also can be written as $D_{ij} = \langle u_{ij} + \sigma_j, \sigma_j \rangle$. By constructing the vectors $g_k = (\|x_i\|^2, x_i 1)^T$, $h_k = (1 - 2\sigma_j \|s_j\|^2)^T$ and collecting them into matrices $G$ and $H$, we have $D = GHH^T$. It is a matrix with elements $D_{ij} = \sum_{k=1}^{N} g_i(k)h_j(k)$, and matrix $D$ has at most rank 5. By simple manipulations, it is possible to eliminate the terms $\|s_j\|^2$, and in that way construct a rank-4 formulation, i.e.,

$$D_{ij} - D_{ij} = \|x_i\|^2 - \|x_i\|^2 + 2\langle x_i, s_j \rangle. \quad (5)$$

$$D_{ij} - D_{ij} = \langle u_{ij} + \sigma_j, \sigma_j \rangle + 2\sigma_j(u_{ij} - u_{ij}). \quad (6)$$

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

$$\hat{D} = A + 2\text{diag}(\sigma)B. \quad (7)$$

is a rank 4 matrix. Here $\hat{D}$ 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

$$\hat{D} = \begin{pmatrix}
D_{11} - D_{11} & D_{12} - D_{12} & \cdots & D_{1M} - D_{1M} \\
D_{21} - D_{11} & D_{22} - D_{12} & \cdots & D_{2M} - D_{1M} \\
\vdots & \vdots & \ddots & \vdots \\
D_{N1} - D_{11} & D_{N2} - D_{12} & \cdots & D_{NM} - D_{1M}
\end{pmatrix}.$$
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 $c = 343\text{m/s}$, 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 $U_{8 \times 110}$ (or $U_{8 \times 266}$) 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

$$U_{8 \times p} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\u_{121} & u_{122} & \cdots & u_{12p} \\
\vdots & \vdots & \ddots & \vdots \\
u_{81} & u_{82} & \cdots & u_{8p} \end{bmatrix}, \quad \text{here } p = 110 \text{ (or 266)}.$$

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 $|o_t - o|$, here $|·|$ 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. 1: Offset estimation with synthetic data. The x- and y-axis represents the number of offsets and the offset values, respectively.](image1)

![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.](image2)
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.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th># of matches</th>
<th># of offsets</th>
<th>error (m)</th>
<th>t (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RANSAC based algorithm</td>
<td>110</td>
<td>80</td>
<td>0.5358</td>
<td>3.1289</td>
</tr>
<tr>
<td>Data I</td>
<td>266</td>
<td>176</td>
<td>0.1052</td>
<td>4.8381</td>
</tr>
<tr>
<td>Alternating projection algorithm</td>
<td>110</td>
<td>60</td>
<td>0.1345</td>
<td>0.8752</td>
</tr>
<tr>
<td>Data I</td>
<td>266</td>
<td>138</td>
<td>0.0845</td>
<td>2.4680</td>
</tr>
</tbody>
</table>

The alternating projection method to estimate offsets. We show experimentally that this 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 estimated times and in computational speed. Moreover, it is much easier to implement.

REFERENCES