Observability Analysis of Graph SLAM-Based Joint Calibration of Multiple Microphone Arrays and Sound Source Localization

Yuanzheng He, Jiang Wang, Daobilige Su, Kazuhiro Nakadai, Junfeng Wu, Shoudong Huang, Youfu Li, and He Kong

Abstract—Multiple microphone arrays have many applications in robot audition, including sound source localization, audio scene perception and analysis. However, accurate calibration of multiple microphone arrays remains a challenge because there are many unknown parameters to be identified, including the Euler angles, geometry, asynchronous factors between the microphone arrays. This paper is concerned with joint calibration of multiple microphone arrays and sound source localization using graph simultaneous localization and mapping (SLAM). By using a Fisher information matrix (FIM) approach, we focus on the observability analysis of the graph SLAM framework for the above-mentioned calibration problem. We thoroughly investigate the identifiability of the unknown parameters, including the Euler angles, geometry, asynchronous factors between the microphone arrays, and the sound source locations. We establish necessary/sufficient conditions under which the FIM and the Jacobian matrix have full column rank, which implies the identifiability of the unknown parameters. These conditions are closely related to the variation in the motion of the sound source and the configuration of microphone arrays, and have intuitive and physical interpretations. We also discover several scenarios where the unknown parameters are not uniquely identifiable. All theoretical findings are demonstrated using simulation data.

I. INTRODUCTION

Microphone array-based robot audition systems can be used for a range of applications, such as sound source localization, active multi-mode perception, speech separation, and recognition of multiple sound sources [1]-[8]. However, accurate calibration of microphone array-based robotic auditory sensors, as for other sensing modalities such as camera and LIDAR [9]-[12], is crucial for satisfactory performance. Hence, calibration of microphone array-based robot audition systems have received much attention in the recent literature.

For example, a calibration technique was proposed in [13], which allowed estimating microphone position, source position and time offset independent of the calibration signal. Some researchers have tried to use frameworks combining SLAM and beamforming algorithms to perform online calibration of asynchronous microphones without many measurements of transfer functions [14]-[15]. For microphone arrays with asynchronous effects (i.e., clock difference and initial time offset), a systematic examination and observability analysis of SLAM-based microphone array calibration and sound source localization was presented in [16]-[18] via a FIM approach. However, the above-mentioned methods are only applicable for calibrating a single microphone array.

Methods for estimating the parameters of multiple microphone arrays have been presented in [19]-[22]. Nevertheless, these methods assumed that the hardware synchronization or orientations of the microphone arrays were known, and only considered scenarios in 2-dimensions (2D). For calibrating multiple microphone arrays, it is necessary to consider not only the geometry and asynchronous effects among the arrays, but also the orientations of microphone arrays.

Simultaneous calibration of positions, orientations, time offsets among multiple microphone arrays and sound source location was explored in [23]. In the former work, a combined cost function has been proposed that can allows for estimating the array position, orientation, and time offset concurrently, by using direction of arrival (DOA) information and the time difference of arrival (TDOA) measurements among microphone arrays. However, a thorough analysis regarding the parameter observability in the joint calibration of multiple microphone arrays and sound source localization is still lacking.

In this study, we will use graph SLAM as a general framework for the above identification question, and concentrate on the parameter identifiability of the corresponding SLAM problem. By using a FIM approach, we thoroughly investigate the identifiability of the unknown parameters, including the Euler angles, geometry, asynchronous factors between the microphone arrays, and the sound source locations. We establish necessary/sufficient conditions under which the FIM and the Jacobian matrix have full column rank, which implies the identifiability of the unknown parameters. These conditions are closely related to the variation in the motion of the sound source and the configuration of microphone arrays, and have intuitive and physical interpretations. We also discover several scenarios where the unknown parameters are not uniquely identifiable. All theoretical findings have been validated using simulation data. For readability, most proofs of the theoretical results are put in the Appendix.
Fig. 1. Geometry of the problem setup and graph-based SLAM framework

Notation: Denote \( x, \mathbf{x}, \) and \( \mathbf{X} \) as scalars, vectors, and matrices, respectively. \( X^T \) represents the transpose of matrix \( X. I_n \) stands for the identity matrix of \( n \) dimensions. \( \mathbb{R}^n \) denotes the \( n \)-dimensional Euclidean space. \( [a_1, \cdots, a_n] \) denotes \( [a_1^T, \cdots, a_n^T]^T \), where \( a_1, \cdots, a_n \) are scalars/vectors/matrices with proper dimensions. \( \text{diag}(\mathbf{A}) \) denotes a block diagonal matrix with \( \mathbf{A} \) as block diagonal entries for \( n \) times; \( \text{diag}(\mathbf{A}, \mathbf{B}) \) denotes a block diagonal matrix with \( \mathbf{A} \) and \( \mathbf{B} \) as its block diagonal entries; and \( \mathbf{0}_{a \times b} \) as a dimension \( a \times b \) with all its entries as 0. \( \mathbf{X}^T \mathbf{P} = \mathbf{x}^T \mathbf{P} \mathbf{x} \). Vectors/matrices, with dimensions not explicitly stated, are assumed to be algebraically compatible.

II. PRELIMINARIES AND PROBLEM STATEMENT

A. Graph SLAM for Multiple Microphone Arrays Calibration

In a calibration scene containing \( N \) distributed microphone arrays, the microphone arrays capture \( K \) consecutive acoustic signals emitted by a single acoustic source at several spatial positions. As shown in Fig. 1 (here we take \( N=3 \) as an example), in this paper, simultaneous sound source localization and multiple microphone arrays calibration are performed in a graph-based SLAM framework with three microphone arrays and one moving source.

In Fig. 1, \( \mathbf{x}^0_{\text{arr}, i} \) represents the location of the \( i \)-th microphone array in the global reference frame, and any two of arrays are in different positions. We assume that there is a reference frame \( \mathbf{x}^0_{\text{arr}, 1} \) attached to every microphone array; we choose \( \mathbf{x}^0_{\text{arr}, 1} \) as the global reference frame; \( \mathbf{R}_i \) is the rotation matrix of reference frame \( \{\mathbf{x}_{\text{arr}, i}\} \) to the frame \( \{\mathbf{x}_{\text{arr}, 1}\} \) with the rotation angle vector \( \mathbf{s}^0_{\text{arr}, j} \); \( \mathbf{s}^k \) is the sound source position at time \( \mathbf{t}^k, k=1, \ldots, K \), with respect to \( \{\mathbf{x}_{\text{arr}, 1}\} \), where \( K \) is the total number of time steps; \( d^k \) is the distance between the \( i \)-th microphone array and the sound source at time instance \( \mathbf{t}^k \). Note that in the calibration process, the multiple microphone arrays remain static while the sound source moves around in the environment.

Here we consider the most general scenario where there are starting time offset and clock drift among different microphone arrays (we assume that the configuration of each microphone array, including its geometry, is known). When the sound source sends the \( k \)-th acoustic signal, the DOA information, i.e., the direction vector of sound source relative to the \( i \)-th microphone array frame \( \{\mathbf{x}_{\text{arr}, i}\} \) is obtained as follows:

\[
d^k_{\text{arr}, i} = \mathbf{R}^T_{\text{arr}, i} \mathbf{s}^k_{\text{arr}, i} - \mathbf{x}^0_{\text{arr}, i}.
\]

(1)

Denote \( d^k_i \), for \( i = 1, \ldots, N \), as the distance between the \( i \)-th microphone array and the sound source at the \( k \)-th sampling instant. The TDOA information between the \( i \)-th and the first microphone arrays can be expressed as follows:

\[
T^k_{\text{arr}, i} = \frac{d^k_i - d^k_0}{c} + \mathbf{x}^0_{\text{arr}, i} - \mathbf{x}^0_{\text{arr}, 0},\quad i = 2, \ldots, N,
\]

(2)

for \( i = 2, \ldots, N \), where \( c \) represents the sound speed in the air; the scalar (unknown) constant variables \( d^k_0 \) and \( \mathbf{x}^0_{\text{arr}, 0} \) represent the starting time offset and the clock difference per second of each microphone array, respectively; \( \Delta t \) is the time interval between two consecutive sound signals. As mentioned above, the first microphone array is used as the reference, hence

\[
\mathbf{x}^0_{\text{arr}, 1} = 0, \quad \mathbf{x}^0_{\text{arr}, 1} = 0, \quad \mathbf{x}^0_{\text{arr}, 1} = 0, \quad \mathbf{x}^0_{\text{arr}, 1} = 0.
\]

The Euler angles, starting time offsets, and clock differences of the microphone arrays will be determined along with the source positions in the calibration process.

The location and the rotation angle vector of \( i \)-th microphone array (where \( i = 2, \ldots, N \), i.e., \( \mathbf{x}^0_{\text{arr}, i} \) and \( \mathbf{x}^0_{\text{arr}, i} \)), can be expressed as:

\[
\mathbf{x}^0_{\text{arr}, i} = \begin{bmatrix} x^0_{\text{arr}, i}^1; x^0_{\text{arr}, i}^2; x^0_{\text{arr}, i}^3 \end{bmatrix}, \quad \mathbf{x}^0_{\text{arr}, i} = \begin{bmatrix} \theta^0_{\text{arr}, i}^x; \theta^0_{\text{arr}, i}^y; \theta^0_{\text{arr}, i}^z \end{bmatrix},
\]

respectively, where \( \theta^0_{\text{arr}, i}^x, \theta^0_{\text{arr}, i}^y, \theta^0_{\text{arr}, i}^z \) take values in the range of \([0, 2\pi],[0, \pi]\), and \([0, 2\pi]\), respectively. Denote the unknown parameters w.r.t. the \( i \)-th microphone array as:

\[
\mathbf{x}_{\text{arr}, i} = \begin{bmatrix} x^0_{\text{arr}, i}^1; x^0_{\text{arr}, i}^2; x^0_{\text{arr}, i}^3; \theta^0_{\text{arr}, i}^x; \theta^0_{\text{arr}, i}^y; \theta^0_{\text{arr}, i}^z \end{bmatrix}^T.
\]

Hence, all the unknown parameters w.r.t. the microphone arrays are:

\[
\mathbf{x}_{\text{arr}} = \left[ x_{\text{arr}, 2}; \cdots; x_{\text{arr}, N} \right]^T.
\]

Denote the sound source position at time \( \mathbf{t}^k, k=1, \ldots, K \) as:

\[
\mathbf{s}^k = \begin{bmatrix} s^k_1; s^k_2; s^k_3 \end{bmatrix}.
\]

Thus, all unknown parameters to be identified are:

\[
\mathbf{x} = \left[ \mathbf{x}_{\text{arr}}; \mathbf{s}^1; \cdots; \mathbf{s}^K \right]^T.
\]

We denote the ideal TDOA and DOA measurement information at the \( k \)-th time instance as:

\[
\mathbf{z}^k = \begin{bmatrix} T^k_{\text{arr}, 1}; \mathbf{d}^k_{\text{arr}, 1}; T^k_{\text{arr}, 2}; \mathbf{d}^k_{\text{arr}, 2}; \cdots; T^k_{\text{arr}, N}; \mathbf{d}^k_{\text{arr}, N} \end{bmatrix}^T \in \mathbb{R}^{4(N-1)}.
\]

(3)

The real values of DOA and TDOA measurements at time \( k \) are also subject to the influence of Gaussian noise as follows:

\[
\mathbf{y}^k = \mathbf{z}^k + \mathbf{v}^k
\]

(4)

where \( \mathbf{z}^k \) is defined in (3), \( \mathbf{v}^k \sim \mathcal{N}(0,\mathbf{P}) \), with \( \mathbf{P} > 0 \in \mathbb{R}^{4(N-1) \times 4(N-1)} \). We assume that the sound source relative
The question of interests is formally stated as follows.

*Problem:* Given the problem setup described as above, find conditions under which the FIM \( I_{FIM} \) defined in (9) is non-singular, or equivalently, the Jacobian matrix \( J \) is of full column rank.

### III. MAIN RESULTS

By leveraging the structure of the Jacobian matrix \( J \) associated with the SLAM formulation, we next establish necessary/sufficient conditions for the non-singularity of the \( I_{FIM} \) and the observability of the SLAM problem. In addition, we will reveal some special cases when the Jacobian matrix or FIM cannot have full column rank.

#### A. Main Results

From the definition of the Jacobian matrix [28, pp. 569], we know that \( J \in \mathbb{R}^{q \times g^2} \). \( g_1 = 4(N-1)K + 3(K-1) \), \( g_2 = 8(N-1) + 3K \). From (9)-(10), a necessary and sufficient condition for \( I_{FIM} \) to be nonsingular is that \( J \) has full column rank. For \( J \) to be of full column rank, it is necessary that

\[
4(N-1)K + 3(K-1) \geq 8(N-1) + 3K
\]

\[
\implies K \geq \left[ \frac{2 + \frac{3}{4(N-1)}}{N} \right],
\]

where \( \lceil \cdot \rceil \) stands for the ceiling operation generating the least integer not less than the number within the operator. We then have the following results.

**Proposition:** The Jacobian \( J \) can be written as

\[
\begin{bmatrix}
\mathbf{L}^1 & \mathbf{T}^1 & \mathbf{0}_{4N \times 3} & \cdots & \mathbf{0} & \mathbf{0} \\
0_{3 \times 8N} & -\mathbf{I}_3 & \mathbf{I}_3 & \cdots & \mathbf{0} & \mathbf{0} \\
\mathbf{L}^2 & \mathbf{T}^2 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\
0 & 0 & -\mathbf{I}_3 & \cdots & \mathbf{0} & \mathbf{0} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\mathbf{L}^{K-1} & 0 & 0 & \cdots & \mathbf{T}^{K-1} & 0 \\
0 & 0 & \cdots & \mathbf{I}_3 & \mathbf{I}_3 & \mathbf{0} \\
\mathbf{L}^{K} & 0 & \cdots & \mathbf{0} & \mathbf{T}^K \\
\end{bmatrix}
\]

(11)

where \( \mathbf{N} = N-1 \), expressions of \( \mathbf{L}^k, \mathbf{T}^k \), for \( k = 1, \ldots, K \), can be found in (15) and (19).

**Theorem 1:** The Jacobian matrix \( J \) is of full column rank if and only if the following matrix

\[
\mathbf{F} = \begin{bmatrix}
\mathbf{L}^1 & \mathbf{T}^1 \\
\mathbf{L}^2 & \mathbf{T}^2 \\
\vdots & \vdots \\
\mathbf{L}^{K} & \mathbf{T}^K \\
\end{bmatrix}
\]

is of full column rank.

**Proof:** The proof follows similarly from [18] and is skipped here.

**Theorem 2:** The Jacobian matrix \( J \) is of full column rank only if the matrix \( \mathbf{T} \) and \( \mathbf{L}^i \), for \( i = 2, \ldots, N \), are of full column rank, respectively, where
where \( h, U, \) and \( V \), can be found in (16).

**Theorem 3:** The Jacobian matrix \( J \) is of full column rank if one or more of the following conditions hold:

(i) Any matrix consisting of the \((j-1)\)-th column block and the last column block in \( F \) is of full column rank, \( 2 \leq j \leq N \).

(ii) All matrices \( \bar{L}_4 \) in (20), for the multiple microphone arrays system, \( i = 2, \ldots, N \) and \( i \neq j \) are of full column rank.

**B. Special Cases When Observability is Impossible**

Next, we state some exceptional cases when observability is impossible.

**Theorem 4:** The matrix \( T \) is not of full column rank if one or more of the following conditions hold:

(i) For all microphone arrays, there exists at least five time steps information (for this to hold, we must have \( K \geq 5 \) in (12)), i.e., when \( K < 5 \), the Jacobian matrix \( J \) is not of full column rank.

(ii) The coordinates of the sound source at all moments are collinear (together with the origin) in \( \{x_{arr,1}\} \), i.e., \( s^k = \lambda s^{k-1} \) does always hold, where \( \lambda \) is an arbitrary real number.

(iii) The sound source keeps moving at any plane of \( x = \alpha y, x = \beta z \), and \( y = \gamma z \) w.r.t. \( \{x_{arr,1}\} \) in all moments, where \( \alpha, \beta, \gamma \) are arbitrary real numbers.

**Theorem 5:** The matrix \( \bar{L}_4, i = 2, 3, \cdots, N \), are not of full column rank if one or more of the following conditions hold:

(i) The coordinates of the sound source at all moments are proportional w.r.t. \( \{x_{arr,1}\} \), i.e., \( (s^k - x_{arr,1})/\lambda = (s^{k-1} - x_{arr,1}) \) does always hold, where \( \lambda \) is an arbitrary real number.

(ii) For the \( i \)-th microphone array, one of the Euler angles satisfies \( \theta_{arr,j}^i = \frac{\pi}{2} \).

**IV. Numerical Simulations and Results**

We next use numerical simulations to illustrate the theoretical findings obtained above. The whole experimental scheme is shown in Fig. 1, where the multiple microphone arrays remain static while the sound source moves around in the environment. To generate the data, we assume that the characteristic parameters of each microphone array and sound source positions are known. All TDOA and DOA measurements are corrupted by Gaussian noise. Here we consider the case with eight microphone arrays.

In the simulation process, we set \( \{x_{arr,1}\} \) as the global reference coordinate system. The sound source always moves at a speed of 0.1 m/s and emits acoustic signal once per second. The starting time offset of each microphone array is randomly generated in \( 0 \sim 0.1 \) s, and the clock drift constant is randomly generated in \( 0 \sim 0.1 \) ms to restore the real scene as much as possible.

**A. Observable Cases**

We firstly give two observable scenarios for which the motion trajectories of the sound source in 3D space are shown in Fig. 2(a). The variation of \( F \) matrix rank with time steps is shown in Fig. 2(b). It can be seen that as time steps increase and the sound source moves along the two trajectories, the \( F \) matrix gradually becomes full column rank which indicates that the Jacobian matrix also gradually

**Fig. 2.** Two observable cases and the variation of the \( F \) matrix’s rank. (a) Geometric relationship between the source and the microphone arrays during the movement. (b) Variation of the \( F \) matrix rank with the movement of the source.
B. Unobservable Cases

Several unobservable scenarios are presented in the following to verify the conclusions in Theorems 4-5.

(i) For the Jacobian matrix to have full column rank, it is necessary that the time steps are greater than or equal to 3 so that the number of rows of the Jacobian matrix is greater than the number of columns. As can be seen from Fig. 2(b), when the number of time steps is greater than or equal to 3 but less than 5, the Jacobian matrix is not of full column rank. This reflects that the system is unobservable when the number of time steps is less than 5.

(ii) For the trajectories of the sound source shown in Fig. 3(a), the first case is that the sound source stays co-linear with \( \{x_{\text{arr1}}\} \) during the moving process, and the second case is that the sound source remains co-planar with \( \{x_{\text{arr1}}\} \). From Fig. 3(b), it can be seen that both are permanently unobservable due to the lack of information.

(iii) For the sound source trajectories shown in Fig. 4(a), the first case is that the sound source keeps co-linear with the origin of \( \{x_{\text{arr2}}\} \) during the movement. In the second case, the Euler angles \( \theta^y_{\text{arr4}} \) and \( \theta^y_{\text{arr7}} \) of \( \{x_{\text{arr4}}\} \) and \( \{x_{\text{arr7}}\} \) are \( \frac{\pi}{2} \), and the sound source travels along the route of the observable scenario mentioned in case 1 of Fig. 2(a). The rotation angle is at the singular point of observation, rendering the system unobservable. Hence, the simulations presented above validate the conclusions in Theorems 4-5.

V. CONCLUSION

This paper is concerned with the observability analysis of graph SLAM-based joint calibration of multiple microphone arrays and sound source localization. Via a FIM approach, we thoroughly investigate the identifiability of the unknown parameters, including the Euler angles, geometry, asynchronous effects between the microphone arrays, and the sound source locations. We establish necessary/sufficient conditions under which the FIM and the Jacobian matrix have full column rank, which implies the identifiability of the unknown parameters. These conditions are closely related to the variation in the motion of the sound source and the configuration of microphone arrays, and have intuitive and physical interpretations. Based on these conditions, we also find some special cases when the Jacobian matrix does

---

Fig. 3. The sound source remains co-linear or co-planar with \( \{x_{\text{arr1}}\} \) during the movement. 
(a) Geometric relationship between the source and the microphone arrays during the movement. 
(b) Variation of the F matrix rank with the movement of the source.

Fig. 4. The sound source remains co-linear with \( \{x_{\text{arr2}}\} \) or \( \theta^y_{\text{arr4,7}} = \frac{\pi}{2} \) during the movement. 
(a) Geometric relationship between the source and the microphone arrays during the movement. 
(b) Variation of the F matrix rank with the movement of the source.
not have full column rank, and provide some geometric and physical interpretations. Extensive simulations have been conducted to demonstrate the theoretical findings. The focus of our current and further work is to develop and validate calibration algorithms for multiple microphone arrays.

VI. ACKNOWLEDGMENT

This work was supported by the Science, Technology, and Innovation Commission of Shenzhen Municipality [Grant No. ZDSYS20200811143601004].

REFERENCES

[1] F. Grondin and F. Michaud, Lightweight and optimized sound source localization and tracking methods for open and closed microphone array configurations, *Robotics and Autonomous Systems*, Vol. 113, pp. 87–99, 2019.
[2] L. Wang, R. Sanchez-Matilla, and A. Cavallaro, Tracking a moving sound source from a multi-robot drone, *Proc. of the IEEE/RSJ IROS*, pp. 2511–2516, 2018.
[3] C. Evers and P. A. Naylor, Acoustic SLAM, *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, Vol. 26, No. 9, pp. 1484–1489, 2018.
[4] M. Strauss, P. Mordel, V. Miguez, and A. Delerfge, DREGON: Dataset and methods for UAV-embedded sound source localization, *Proc. of the IEEE/RSJ IROS*, pp. 5735–5742, 2018.
[5] Z. Shi, X. Chang, C. Yang, Z. Wu, and J. Wu, An acoustic-based surveillance system for amateur drones detection and localization, *IEEE Trans. on Vehicular Technology*, Vol. 69, No. 3, pp. 2731–2739, 2020.
[6] S. Effert, N. Wallace, H. Kong, N. Pirmarzadshahi, and S. Sukkarieh, Resource and response aware path planning for long term autonomy of ground robots in agriculture, *Field Robotics*, Vol. 2, pp. 1–33, 2022.
[7] J. Wakulicz, H. Kong, and S. Sukkarieh, Active information acquisition under arbitrary unknown disturbances, *Proc. of the IEEE ICRA*, Vol. 96, pp. 8429–8435, 2021.
[8] C. Rascon and I. Meza, Localization of sound sources in robotics: A review, *Robotics and Autonomous Systems*, Vol. 96, pp. 184–210, 2017.
[9] J. Lv, X. Zuo, K. Hu, J. Xu, G. Huang, and Y. Liu, Observability-aware intrinsic and extrinsic calibration of LiDAR-IMU systems, *IEEE Trans. on Robotics*, Early access, pp. 1–20, 2022.
[10] J. Brown, D. Su, H. Kong, S. Sukkarieh, and E. Kerrigan, Improved noise covariance estimation in visual servoing using an autocovariance least-squares approach, *Mechatronics*, Vol. 68, pp. 1–10, 2020.
[11] M. Liu, Y. Li, and H. Liu, Robust 3-D gaze estimation via data optimization and saliency aggregation for mobile eye-tracking systems, *IEEE Trans. on Instrumentation and Measurement*, Vol. 70, Article ID. 5008010, pp. 1–10, 2021.
[12] J. Jiao, Y. Yu, Q. Liao, H. Ye, R. Fan, and M. Liu, Automatic calibration of multiple 3D lidars in urban environments, *Proc. of the IEEE/RSJ IROS*, pp. 15–20, 2019.
[13] P. Perrodin, J. Nikolic, J. Busset, and R. Siegwart, Design and calibration of large microphone arrays for robotic applications, *Proc. of the IEEE/RSJ IROS*, pp. 4596–4601, 2012.
[14] H. Miura, T. Yoshida, K. Nakamura, and K. Nakadai, SLAM-based online calibration of asynchronous microphone array for robot audition, *Proc. of the IEEE/RSJ IROS*, pp. 524–529, 2011.
[15] H. Miura, T. Yoshida, K. Nakamura, and K. Nakadai, SLAM-based online calibration for asynchronous microphone array, *Advanced Robotics*, Vol. 26, No. 17, pp. 1941–1965, 2012.
[16] D. Su, T. Vidal-Calleja, and J. V. Miro, Simultaneous asynchronous microphone array calibration and sound source localisation, *Proc. of the IEEE/RSJ IROS*, pp. 5561–5567, 2015.
[17] D. Su, T. Vidal-Calleja, and J. V. Miro, Asynchronous microphone arrays calibration and source sound tracking, *Autonomous Robots*, Vol. 44, No. 2, pp. 183–204, 2020.
[18] D. Su, H. Kong, S. Sukkarieh, and S. Huang, Necessary and sufficient conditions for observability of SLAM-based TDOA sensor array calibration and source localisation, *IEEE Trans. on Robotics*, Vol. 37, No. 5, pp. 1451–1468, 2021.

APPENDIX

Proof of Proposition. Firstly, we note that the relative position of the sound source satisfies
$$s_{\alpha}^{-k} = s^k - s^{-1} + w^{-1}$$
whose corresponding Jacobian matrices are
$$\frac{\partial s_{\alpha}^{-k}}{\partial s^{-1}} = I_3, \quad \frac{\partial s_{\alpha}^{-1}}{\partial s^{-1}} = I_3.$$ Secondly, for \(i = 2, \ldots, N\), the distance between the \(i\)-th microphone array and the sound source at time instance \(k\) can be computed as
$$d_i^k = \sqrt{(\Delta x_i^k)^2 + (\Delta y_i^k)^2 + (\Delta z_i^k)^2}$$
where \(\Delta x_i^k = s_{\alpha,i}^k - s_{\alpha,i}^{-1}\), \(\Delta y_i^k = s_{\alpha,i}^k - s_{\alpha,i}^{-1}\), \(\Delta z_i^k = s_{\alpha,i}^k - s_{\alpha,i}^{-1}\).

When \(i = 1\), i.e., for the first microphone array, we have
$$d_1^k = \sqrt{(\Delta x_1^k)^2 + (\Delta y_1^k)^2 + (\Delta z_1^k)^2}.$$ (14)

Denote \(L_k = \frac{\partial x^k}{\partial s_{\alpha,i}^{-1}}\), i.e., \(L_k\) is the derivative of \(z^k\) (the measurements at the time step \(k\)) w.r.t. \(s_{\alpha,i}^{-1}\). Based on the DOA and TDOA information in (1)–(2), we then have:
$$L_k = \frac{\partial x^k}{\partial s_{\alpha,i}^{-1}} = \begin{bmatrix} j_{\alpha,i}^1, \ldots, j_{\alpha,i}^N \end{bmatrix} \in \mathbb{R}^{4N \times 8N}$$ (15)
where for \(i = 2, \ldots, N\) and \(k = 1, \ldots, K\) and only entries of \(j_{\alpha,i}^k\) on its \((4i - 7 : 4i - 4)\) rows are nonzero. Denote \(h_{\alpha,i}^k U_{\alpha,i}^k\) as the partial derivative of TDOA and DOA w.r.t. microphone array position, respectively; denote \(V_{\alpha,i}^k\) as the
partial derivative of DOA w.r.t. \(X, Y, Z\) Euler angles. We then have:

\[
\mathbf{H}_{\text{arr},i} = \mathbf{J}_{\text{arr},i}(4i - 7 : 4i - 4,:)
\]

\[
\begin{bmatrix}
\mathbf{h}_{i}^T \\
\mathbf{U}_{i}^T \\
\mathbf{V}_{i}^T
\end{bmatrix}
= \begin{bmatrix}
\mathbf{0}_{1 \times 3} & 1 & k a_i \\
\mathbf{0}_{3 \times 1} & 0_3 & 0_3
\end{bmatrix}
\in \mathbb{R}^{4 \times 8}
\]  \hspace{1cm} (16)

where

\[
\mathbf{h}_{i}^T = \begin{bmatrix}
-\Delta \Phi_i \\
-\Delta \Psi_i \\
-\Delta \Theta_i
\end{bmatrix}
\]

\[
\mathbf{U}_{i}^T = -\mathbf{R}_i^T \mathbf{A}
\]

\[
= -\mathbf{R}_i^T
\begin{bmatrix}
(\Delta \Phi_i)^2 + (\Delta \Psi_i)^2 & -\Delta \Phi_i \Delta \Psi_i & -\Delta \Phi_i \Delta \Theta_i \\
-\Delta \Phi_i \Delta \Psi_i & (\Delta \Psi_i)^2 + (\Delta \Theta_i)^2 & -\Delta \Psi_i \Delta \Theta_i \\
-\Delta \Phi_i \Delta \Theta_i & -\Delta \Psi_i \Delta \Theta_i & (\Delta \Theta_i)^2
\end{bmatrix}
\]  \hspace{1cm} (17)

and

\[
\mathbf{V}_{i}^T = \frac{1}{\partial_z^T} = \begin{bmatrix}
(\mathbf{R}_x^T \mathbf{R}_z^T) & (\mathbf{R}_y^T \mathbf{R}_z^T) & (\mathbf{R}_y^T \mathbf{R}_z^T)
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\begin{bmatrix}
\frac{\partial \mathbf{R}_x^T}{\partial \theta_x} & \frac{\partial \mathbf{R}_y^T}{\partial \theta_y} & \frac{\partial \mathbf{R}_z^T}{\partial \theta_z}
\end{bmatrix}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\mathbf{R}_x^T \mathbf{R}_z^T & \mathbf{R}_y^T \mathbf{R}_z^T & \mathbf{R}_y^T \mathbf{R}_z^T
\end{bmatrix}
\]  \hspace{1cm} (18)

where \(\mathbf{R}_{x,y,z}\) and \(\mathbf{R}_{z,x}\) are the rotation matrices about coordinate frame axes \(x, y,\) and \(z,\) respectively. The expression of \(\mathbf{R}_i^T\) is as follows:

\[
\mathbf{R}_i^T = \mathbf{R}_{x,i} \mathbf{R}_{y,i} \mathbf{R}_{z,i};
\]

with

\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & \cos \theta_x & -\sin \theta_x \\
0 & \sin \theta_x & \cos \theta_x
\end{bmatrix}

\begin{bmatrix}
0 & 1 & 0 \\
-\sin \theta_y & 0 & \cos \theta_y \\
\cos \theta_y & 0 & -\sin \theta_y
\end{bmatrix}

\begin{bmatrix}
0 & 0 & 1 \\
\sin \theta_z & \cos \theta_z & 0 \\
-\cos \theta_z & \sin \theta_z & 0
\end{bmatrix}
\]

Denote \(\mathbf{T}_i^k\) as the partial derivative of TDOA and DOA measurements w.r.t. sound source position at time instance \(t^k\), for \(k = 1, \ldots, K\). We then have the expression of \(\mathbf{T}_i^k\) as follows:

\[
\mathbf{T}_i^k = \frac{\partial \mathbf{x}^k}{\partial \mathbf{y}^k} = \begin{bmatrix}
\mathbf{J}_i^k & \mathbf{J}_y^k & \mathbf{J}_z^k
\end{bmatrix}
\begin{bmatrix}
\mathbf{h}_i^k \\
\mathbf{U}_i^k \\
\mathbf{V}_i^k
\end{bmatrix}
\]

\[
= \begin{bmatrix}
-\mathbf{h}_i^k \\
-\mathbf{U}_i^k \\
-\mathbf{V}_i^k
\end{bmatrix}
\]

\[
\mathbf{F} = \begin{bmatrix}
\mathbf{H}_{\text{arr},1}^1 & \mathbf{T}_{\text{arr},1}^1 \\
\vdots & \vdots \\
\mathbf{H}_{\text{arr},N}^N & \mathbf{T}_{\text{arr},N}^N
\end{bmatrix}
\]

\[
\mathbf{F}_{\text{arr},i}^j = \begin{bmatrix}
\mathbf{H}_{\text{arr},i}^j; \mathbf{U}_{\text{arr},i}^j; \mathbf{h}_{\text{arr},i}^j; \mathbf{U}_{\text{arr},i}^j
\end{bmatrix}
\]

\[
\mathbf{F}_{\text{arr},i} = \begin{bmatrix}
\mathbf{M}_{\text{arr},i} \mathbf{0}_{K \times 1} \mathbf{k}_{\text{arr},i} \mathbf{0}_{K \times 1} \mathbf{0}_{K \times 1}
\end{bmatrix}
\]  \hspace{1cm} (19)

The results then follow the definition of the Jacobian matrix [28, pp. 569]. This completes the proof. \(\blacksquare\)

**Proof of Theorem 2.** First, \(\mathbf{L}_i^k\) can be expressed as:

\[
\mathbf{L}_i^k = \text{diag}(\mathbf{H}_{\text{arr},i}^k, \mathbf{h}_{\text{arr},i}^k, \mathbf{U}_{\text{arr},i}^k, \mathbf{U}_{\text{arr},i}^k).
\]

By performing elementary row transformation of \(\mathbf{F}\), we can obtain:

\[
\begin{bmatrix}
\mathbf{H}_{\text{arr},1}^1 & \mathbf{T}_{\text{arr},1}^1 \\
\vdots & \vdots \\
\mathbf{H}_{\text{arr},N}^N & \mathbf{T}_{\text{arr},N}^N
\end{bmatrix}
\]

\[
\mathbf{F}_{\text{arr},i} = \begin{bmatrix}
\mathbf{M}_{\text{arr},i} \mathbf{0}_{K \times 1} \mathbf{k}_{\text{arr},i} \mathbf{0}_{K \times 1} \mathbf{0}_{K \times 1}
\end{bmatrix}
\]

for \(i = 2, \ldots, N\). Apparently, it holds that \(\text{rank}(\mathbf{F}) = \text{rank}(\mathbf{F})\).

Also, due to the structure of \(\mathbf{H}_{\text{arr},i}\), their columns are independent of each other. For each microphone array, denote \(\mathbf{F}_{\text{arr},i} = \begin{bmatrix}
\mathbf{H}_{\text{arr},i}^1; \mathbf{T}_{\text{arr},i}^1
\end{bmatrix} \cdots \begin{bmatrix}
\mathbf{H}_{\text{arr},i}^K; \mathbf{T}_{\text{arr},i}^k
\end{bmatrix}\).

We then perform the following elementary transformation on the matrix \(\mathbf{F}_{\text{arr},i}\):

(i) adding the first column block \(\begin{bmatrix}\mathbf{h}_{\text{arr},i}^1; \mathbf{U}_{\text{arr},i}^1; \cdots; \mathbf{h}_{\text{arr},i}^K; \mathbf{U}_{\text{arr},i}^K\end{bmatrix}\) of \(\mathbf{H}_{\text{arr},i}\) to \(\mathbf{T}_{\text{arr},i}\);

(ii) exchanging row blocks to collect all \(\mathbf{h}_{\text{arr},i}^k\) and \(\mathbf{U}_{\text{arr},i}^k\) together, respectively, thereby obtaining

\[
\mathbf{F}_{\text{arr},i} = \begin{bmatrix}
\mathbf{M}_{\text{arr},i} \mathbf{0}_{K \times 1} \mathbf{k}_{\text{arr},i} \mathbf{0}_{K \times 1} \mathbf{0}_{K \times 1}
\end{bmatrix}
\]
Here we take \( k = 1;2,\ldots;K \), \( M_{\alpha j} = [h_1^\alpha;h_2^\alpha;\ldots;h_K^\alpha] \), \( M_{U,j} = [U_1^j;U_2^j;\ldots;U_K^j] \), \( M_{V,j} = [V_1^j;V_2^j;\ldots;V_K^j] \), and \( t_K = \begin{bmatrix} x_1^\alpha \cr y_1^\alpha \cr z_1^\alpha \end{bmatrix} \). 

We further perform the following elementary operations on \( F_{arr,j} \), \( i = 2,3,\ldots,N \): 
1. dividing the fourth column block by \( \Delta_i \); 
2. for \( k = 2,3,\ldots,K \), deducing the \( k \)-th row by the first row; 
3. transforming the elements in the first row (except the third one) to zero by the third column block (the first element therein equals 1 while the other elements equal zero after the elementary operations listed above); 
4. for \( k = 3,4,\ldots,K \), deducing the \( k \)-th row by the second row multiplied by \( (k - 1) \); 
5. transforming the elements in the second row (except the fourth one) to zero by the fourth column block (the second element therein equals 1 while the other elements equal zero after the elementary operations listed above); 
6. moving column blocks 3 and 4 to columns blocks 1 and 2, respectively.

After the above operations, we obtain
\[
F_{arr,j} = \begin{bmatrix} \hat{L}_4 & \hat{T} \end{bmatrix}
\]
where \( \hat{L}_4 \) and \( \hat{T} \) are shown in (13) and (12), respectively. 

With the above elementary transformations, we have
\[
F \sim F' = \begin{bmatrix} L_2 & \hat{T} \\ \hat{L}_3 & \hat{T} \\ \vdots & \vdots \\ \hat{L}_N & \hat{T} \end{bmatrix}.
\] 

It holds that \( \text{rank}(F) = \text{rank}(F') = \text{rank}(F_{arr,j}) \). From the structure of \( F' \), we can see that the blocks containing \( \hat{L}_4 \), \( i = 2,\ldots,N \), are independent of each other. A necessary condition for \( F \) to be of full column rank is that \( \hat{L}_4 \) and \( \hat{T} \) are of full column rank, respectively, \( i = 2,\ldots,N \). This completes the proof.

**Proof of Theorem 3.** Here we take \( j = 2 \) as an example. For \( F' \), we could perform elementary row block changes: for \( i = 3,\ldots,N \), deduce \( \hat{L}_4 \) row block by the first-row block and obtain:
\[
\begin{bmatrix} L_2 & \hat{T} \\ \hat{L}_3 & 0 \\ \vdots & \vdots \\ \hat{L}_N & 0 \end{bmatrix}
\]

Denote the submatrix of this matrix as:
\[
M_{2,T} = \begin{bmatrix} L_2 & \hat{T} \\ \vdots & \vdots \\ -L_2 & 0 \end{bmatrix}.
\]

From the structure in (21), we can see clearly that if:
1. \( M_{2,T} \) is of full column rank, and (ii) \( \text{diag}(L_3,\ldots,L_N) \) is of full column rank, then \( F' \) will be of full column rank. Due to the fact that \( \text{rank}(F) = \text{rank}(F') = \text{rank}(F_{arr,j}) \), the Jacobian matrix \( J \) is of full column rank. Similarly, the same conditions hold when \( j \) equals to 3,\ldots,\( N \). So the Jacobian matrix \( J \) is of full column rank if any matrix consisting of the \((j-1)\)-th column block and the last column block in \( F' \) is of full column rank, \( 2 \leq j \leq N \), and \( \hat{L}_4 \) are of full column rank, \( i = 2,\ldots,N \) and \( i \neq j \). This completes the proof.

**Proof of Theorem 4.** (i) If \( \hat{T} \) in (12) is of full column rank only if a \( 3 \times 3 \) matrix formed by at least one of the three-permutation of its rows is full rank. For \( (s^k)^T \in \mathbb{R}^{1 \times 3} \), \( 1 \leq k \leq K \), the necessary condition for \( \hat{T} \) to be of full column rank is \( K \geq 5 \). If \( K < 5 \), \( T \) can not be of full column rank.

(ii) Based on (14), when \( s^k = \lambda s^{k-1} \), we could derive \( s^{k-1} = \frac{s^k}{\lambda} \). From the expression of \( T \), we can see that \( T \) cannot be of full rank if \( s^k \) is proportional to each other, \( k = 1,\ldots,K \). In this situation, the sound source positions at all time steps are collinear (together with the origin) w.r.t. the reference microphone array frame.

(iii) If the sound source keeps moving in any planes of \( x = \alpha y, x = \beta z, y = \gamma z \) w.r.t. \( \{x_{arr,j}\} \) at all moments, where \( \alpha, \beta, \gamma \) are arbitrary real numbers, the sound source position \( s^k \), \( 1 \leq k \leq K \), could be expressed as \( [\alpha x_k^1,\alpha x_k^2,\alpha x_k^3], [\beta x_k^1,\beta x_k^2,\beta x_k^3], \) and \( [\gamma x_k^1,\gamma x_k^2,\gamma x_k^3] \), respectively. \( T \) will not be of full column rank.

Specifically, if \( \alpha = 0 \) or \( \beta = 0 \) or \( \gamma = 0 \), the sound source position of \( s^k \) will have \( s_k^3 = 0 \), \( s_k^2 = 0 \), and \( s_k^2 = 0 \), respectively, i.e., YOZ, XOZ, and XOY planes. If the sound source keeps moving in the line of \( x = \alpha y = \beta z \), the situation will change to (ii). This completes the proof.

**Proof of Theorem 5.** (i) If the sound source positions w.r.t. \( \{x_{arr,j}\} \) at all of \( K (K \geq 5) \) time steps are collinear, i.e., \( (s^k-x^0_{arr,j}) = \lambda(s^{k-1}-x^0_{arr,j}) \) is always true. For \( i \geq 2, k = 2,\ldots,K \), we can get the following expression:
\[
\begin{bmatrix} \Delta x^1_k; \Delta x^2_k; \Delta x^3_k \end{bmatrix} = \begin{bmatrix} \Delta x^1_k; \Delta x^2_k; \Delta x^3_k \end{bmatrix},
\]

\[
h_k = h_k^0, U_k = U_k^0, V_k = V_k^0.
\]

where \( h, U, \) and \( V \) are defined in (16).

For an arbitrary single time step, we have \( \text{rank}(U_k^j) = \text{rank}(U_k^j A) \) as shown in (17). It can also be seen that \( \text{det}(A) = 0 \) and the second-order sub-determinant of \( A \) is not equal to zero, we know that \( \text{rank}(A) = 2 \). \( R_k^j \) is a rotation matrix, \( \text{rank}(R_k^j) = 3 \), thus \( \text{rank}(U_k^j) = 2 \). Therefore, \( \hat{L}_4 \) will not be of full column rank.

(ii) When \( \theta_{arr,j} = \frac{\pi}{2} \), for the corresponding microphone array at any different time steps, \( V_k^j \) defined in (18) has the same structure, i.e.,
\[
V_k^j = \begin{bmatrix} 0 & \Delta x_k^1;\cos\theta_{arr,j};\Delta x_k^2;\sin\theta_{arr,j} & 0 \\
\Delta x_k^1;\sin\theta_{arr,j} & -\Delta x_k^2;\cos\theta_{arr,j} & -\Delta x_k^3;\cos\theta_{arr,j}+\Delta x_k^3;\sin\theta_{arr,j} \\
\Delta x_k^1;\cos\theta_{arr,j} & -\Delta x_k^2;\sin\theta_{arr,j} & -\Delta x_k^3;\sin\theta_{arr,j}+\Delta x_k^3;\cos\theta_{arr,j} \end{bmatrix},
\]

where \( s,c \) represent \( \sin,\cos \), respectively and \( \text{rank}(V_k^j) \equiv 2 \). Therefore, the matrix of \( \hat{L}_4 \) in (20) will not be of full column rank. This completes the proof.