Digit. Signal Process. 178 (2026) 106106

Contents lists available at ScienceDirect

Digital Signal Processing

journal homepage: www.elsevier.com/locate/dsp

Low rank properties for synchronizing microphones and sources in Ad-Hoc wireless acoustic sensor network

Faxian Cao [a], Yongqiang Cheng [b,*], Adil Mehmood Khan [a], Zhijing Yang [c], S. M. Ahsan Kazmi [d], Zhao Huang [e], Yingxiu Chang [a]

[a] School of Computer Science, University of Hull Cottingham Road, Hull, HU6 7RX, England, UK
[b] Faculty of Business and Technology, University of Sunderland St Peters's Way, Sunderland, SR6 0DD, England, UK
[c] School of Information Engineering, Guangdong University of Technology, Waihuan West Road, Pan Yu Qu, Guangzhou, 510006, Guangdong Province, China
[d] Faculty of Computer Science and Creative Technologies, University of the West of England, Coldharbour Ln, Stoke Gifford, Bristol, BS16 1QY, England, UK
[e] Department of Computer and Information Sciences, Northumbria University CIS Building, Ellison Place, Newcastle, NE1 8ST, England, UK

ARTICLE INFO

Keywords:
Wireless acoustic sensor network
Microphones start time
Sources emission time
Low-rank property
Combined low-rank approximation

ABSTRACT

Wireless Acoustic Sensor Networks (WASNs) designed for joint localization of acoustic sources and microphones are increasingly utilized in applications such as intelligent speech interfaces, environmental monitoring, and robotic audition. However, uncertainties in timing information-such as imprecise microphone recording start times and unknown acoustic source emission times-pose significant challenges, particularly in ad-hoc WASNs. Traditional optimization methods typically address these issues by explicitly estimating unknown timing information (UTIm) to synchronize sensors and sources. In contrast, the Low-Rank Property (LRP) based methods exploit an inherent low-rank structure to impose linear constraints on UTIm, facilitating globally optimal solutions given proper initialization. Nonetheless, these methods often rely on random initialization, leading to convergence toward suboptimal local minima. Targeting this challenge, this paper introduces a novel combined low-rank approximation (CLRA) method to synchronize microphones and sources, mitigating random initialization effects. We present three new LRP variants, mathematically proven to enhance the UTIm with richer low-rank structural information. Using this augmented information, we formulate four linear constraints on UTIm and employ the CLRA algorithm to derive globally optimal solutions. Experimental results show our method outperforms state-of-the-art approaches in both recovery number and reduced estimation errors of UTIm.

1. Introduction

Speech signals, as a natural and intuitive way for the development of human-machine interfaces, have garnered significant interest in the wireless communications network [1]. By incorporating microphones into ad-hoc wireless acoustic sensor network (WASN), we can utilize WASN in applications such as smart phone [2], environmental monitoring [3], smart city [4], health monitoring [5], etc. Among these applications, the design of local positioning systems has been significantly advanced by the use of location information over the past decade, driven by its potential for developing ambient intelligence applications [6]. For instance, in the applications of ambient-assisted living, the utilization of location information enables the living environment safer, easier and more comfortable [6]. Additionally, microphone tends to be lightweight, smaller and less intrusive than the video camera [6].

Time of arrival (TOA) and time difference of arrival (TDOA) [7,8] measurements are two typical information for localizing both microphones and sources in ad-hoc WASN. Specifically, the TOA measurements between individual microphones and sources are obtained through cross-correlation when the waveform of the source signal is known beforehand, such as its amplitude, frequency, and duration [9]. Additionally, the TDOA of a pair of microphones with respect to the corresponding source are determined by performing cross-correlation on the two corresponding microphone signals when the waveform of the source signal is unknown [10]. When microphones and sources are synchronized, meaning their recording start time and emission time align, the time difference between the emission of an audio signal by the source and its reception by a microphone is measured using TOA [11]. Similarly, TDOA measures the delay between a pair of microphones receiving corresponding audio signals [12]. Therefore, accurate

* Corresponding author.
E-mail addresses: faxiancao@foxmail.com (F. Cao), yongqiang.cheng@sunderland.ac.uk (Y. Cheng), a.m.khan@hull.ac.uk (A.M. Khan), yzhj@gdut.edu.cn (Z. Yang), ahsan.kazmi@uwe.ac.uk (S. M.A. Kazmi), zhao.huang@northumbria.ac.uk (Z. Huang), yingxiuchang@gmail.com (Y. Chang).

https://doi.org/10.1016/j.dsp.2026.106106

Available online 9 April 2026

estimation of TOA and TDOA [7,8] measurements is a critical prerequisite for various applications in ad-hoc WASN, such as microphones array calibration [13,14], source localization and tracking [15,16], joint microphones and sources localization [17–20], and simultaneous localization and mapping [21,22]. However, the accurate estimation of TOA and TDOA is complicated by the presence of unknown timing information (UTIm) arising from asynchronous microphones and sources within ad-hoc WASN in real-world applications. This lack of synchronization in the start times of microphones and emission times of sources is a significant challenge that undermines the reliability and accuracy of localization and tracking systems [23].

Existing methods [23–25] to synchronize microphones and sources in ad-hoc WASN by estimating UTIm primarily fall into two categories. The first category relies on direct optimization methods, including maximum likelihood estimation [24] and distributed damped Newton optimization [25]. Despite their popularity, these methods require both TDOA and angle of arrival information [26] and their applications can be limited in certain scenarios. The second category [23,27] hinges on exploiting the low-rank property (LRP) [28], where the low-rank structure information of UTIm is used to formulate linear constraints for UTIm estimation. The effectiveness of this method is often hindered by unknown and random initialization of UTIm, leading to local (suboptimal) minimal values and thus limited recovery rate (the ratio of number for successful initializations to all initializations in one configuration) and convergence rate (the ratio of number for successful recovery configurations to number of all configurations), particularly when there is noise in TOA measurements.

Drawing upon our prior work [29], which highlighted the sufficiency of microphone signals alone for localizing asynchronous microphones and sources within an ad-hoc WASN, we delve deeper into the relationship between TOA and TDOA formulas in our previous work [29] by presenting a novel mapping function. More specifically, expanding on the significant alignment between the transformations of TOA and TDOA measurements discovered in our previous research, we imply that the low-rank structure information exploited between UTIm and TOA could be similarly applied to TDOA, suggesting the distances between microphones and sources can be acquired once the UTIm in TOA or TDOA is estimated.

In this paper, we introduce the combined low-rank approximation (CLRA) method to estimate UTIm in both TOA and TDOA. This method integrates the linear constraints formed by LRP with three new variants of LRP designed to exploit more low-rank structure information of UTIm. Then, by using these four linear constraints to restrict UTIm, our CLRA method seeks a global optimization solution for UTIm estimation, improving both convergence and recovery rates while reducing estimation errors in noise environment. In addition, alongside presenting this novel method, we provide a mathematical proof for the proposed LRP variants, reinforcing the theoretical foundation of our approach. This paper, therefore, not only uncovers a novel methodology for UTIm estimation but also presents robust evidence to validate its effectiveness. This study thus serves as a significant advancement in the realm of ad-hoc WASN, paving the way for more accurate and efficient methods for estimating UTIm. The key contributions of our study are as follows:

- Three novel low-rank properties are presented to leverage the low-rank structure between TOA (or TDOA) and UTIm, with rigorous mathematical proofs supporting their validity.
- A CLRA method is proposed to achieve global solutions for synchronizing microphones and sources by accurately estimating the start times of microphones and the emission times of sources.
- The introduced rank properties offer potential applicability to other domains facing similar synchronization challenges, such as radio signals, and can contribute to a wide range of technological and engineering applications.

This paper is organized as follows. Section 2 reviews previous research on estimating UTIm using both TOA and TDOA measurements. Sections 3 and 4 present the problem formulation and preliminary concepts, respectively. Sections 5 and 6 introduce the proposed Low-Rank approach along with its underlying mathematical principles. Section 7 demonstrates experimental results and includes benchmark comparisons. Finally, Section 8 concludes the paper by summarizing the key findings.

2. State of the art

This section reviews the latest methods for estimating UTIm in both TOA and TDOA measurements, used for synchronizing microphones and sources in ad-hoc WASN. As stated early in Section 1, the methods for UTIm estimation are categorized into two groups, i.e., direct optimization and LRP-based [28]. Among the direct optimization techniques, probabilistic generative model [30] estimates sources emission time by using TOA measurements. However, this model ignores the start time of microphones in TOA measurements. To eliminate UTIm in TOA measurements, one work introduces Gram matrices [31], assuming that the center point among microphones and sources is known. Nonetheless, this assumption does not hold for microphones array applications. Maximum likelihood estimation [32] assumes co-located sources and microphones and estimates their locations and microphone start time using either TOA or TDOA information. In addition, optimal solutions are derived through close form equations in this case [32]. Auxiliary function-based approach [12] uses TDOA information to estimate the locations and microphone time offset, showing better convergence properties. Some approaches, such as maximum likelihood estimation [24] and distributed damped Newton optimization [25], combine TDOA with angle of arrival (AOA) measurements to estimate UTIm. However, these approaches require both TDOA and AOA measurements and this is not usually satisfied in many scenarios. Among the first group of methods reviewed, only auxiliary function approach [12] presents a general method while others require additional geometric constraints on the location of microphones and sources or assume the start or emission time to be known or necessitate more measurements.

Regarding the LRP-based methods for estimating UTIm [28], they are categorized into three algorithms: alternating minimization (AM) [33,34], nuclear truncation minimization (NTM) [35,36], and structure total least square (STLS) [23,37–40]. Of these algorithms, the STLS method is over 100 times faster than both AM and NTM. Besides, by denoting the time offset of the microphones in TDOA as the pseudo-start time of the microphones and the distance (divided by the speed of sound) between the first microphone and the corresponding source as the pseudo-source emission time [31], the LRP [28] is being used to estimate UTIm with TDOA information. In our previous work [29], we demonstrated that the transformations of both TOA and TDOA measurements are identical to one another, showing the UTIm in TOA and TDOA are the same as each other. Therefore, the values of variable in the low-rank matrix exploited by LRP with TDOA measurements can be equivalent to those in the low-rank matrix exploited by LRP with TOA information.

In summary, various methods are proposed to estimate UTIm using TOA or TDOA measurements. Among these methods, the auxiliary function-based algorithm [12] and LRP-based methods [33] are promising due to their consideration for a general formulation. However, LRP-based [28] methods are found to have better recovery and convergence rates with less estimation errors of UTIm than the auxiliary function-based algorithm [12]. In addition, by using LRP [28], STLS is found to be 100 times faster than AM [33,34] and NTM [35,36]. Nevertheless, with LRP only, resulting in a risk of solutions getting stuck in local minima. To address this issue, this paper proposed three variants of LRP that exploit more of the low-rank structure information pertaining to UTIm in TOA or TDOA. The proposed CLRA uses both LRP and the corresponding three new variants to constrain UTIm, allowing for a global solution to be found. Therefore, the proposed CLRA method helps improve the recovery and convergence rate while also reducing estimation errors of UTIm.

3. Problem formulation

Assuming there are M microphones and N sources located at unknown locations $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M]_{3\times M}$ and $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N]_{3\times N}$, respectively, where 3 represents the 3 dimensional space (For ease of reference, all notations are listed in *Table 1 in Supplementary Material*). Then, if there is a fundamental control centre for both microphones and sources, the microphones and sources can be synchronized, i.e., both the emission time of sources and the microphones start time are known. Thus TOA ($\mathbf{t}_{i,j} = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|}{c}$) between i^{th} microphone and j^{th} source is obtained with the received microphone signal and known waveform of source signal [41] (Note that $\|\bullet\|$ and c are the l_2 norm and speed of sound, respectively.) Similarly, TDOA ($\boldsymbol{\zeta}_{i,j} = \frac{\|\mathbf{r}_i - \mathbf{s}_j\| - \|\mathbf{r}_1 - \mathbf{s}_j\|}{c}$) between the 1st microphone and the i^{th} microphone with respect to j^{th} source is obtained using the received signals from a pair of microphones [10]. However, usually the microphones and sources are asynchronous [29,31,38], resulting in the emission time of sources and the start time of microphones being unknown.

If microphones and sources are asynchronous and the waveform of source signal is known, TOA between microphone and source is incomplete because of the UTIm. Denote $\boldsymbol{\delta} = [\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \ldots, \boldsymbol{\delta}_M]^T$ and $\boldsymbol{\eta} = [\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \ldots, \boldsymbol{\eta}_N]^T$ where $\boldsymbol{\delta}_i$ and $\boldsymbol{\eta}_j$ are the unknown start time of i^{th} microphone and unknown emission time of j^{th} source, respectively (i and j range from 1 to M and 1 to N, respectively). Then the TOA between microphones and sources is

$$\mathbf{t}_{i,j} = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|}{c} + \boldsymbol{\eta}_j - \boldsymbol{\delta}_i. \tag{1}$$

If microphones and sources are asynchronous and the waveform of source signal is unknown, then TDOA is available and the received signals from a pair of microphones are utilized for TDOA estimation with the method of cross-correlation [10]. Thus the relationship between TDOA and the location of microphones and sources is displayed as [31,37]

$$\boldsymbol{\zeta}_{i,j} = \mathbf{t}_{i,j} - \mathbf{t}_{1,j} = \frac{\|\mathbf{r}_i - \mathbf{s}_j\| - \|\mathbf{r}_1 - \mathbf{s}_j\|}{c} + \boldsymbol{\delta}_1 - \boldsymbol{\delta}_i = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|}{c} - \frac{\|\mathbf{r}_1 - \mathbf{s}_j\|}{c} + \boldsymbol{\delta}_i', \tag{2}$$

where $\boldsymbol{\delta}_i'$ is the time offset between the i^{th} microphone and 1st microphone.

Denote $\hat{\boldsymbol{\eta}}_j = -\frac{\|\mathbf{r}_1 - \mathbf{s}_j\|}{c}$ and $\hat{\boldsymbol{\delta}}_i = -\boldsymbol{\delta}_i'$, TDOA formula in Eq. (2) is rewritten as [31,37]

$$\boldsymbol{\zeta}_{i,j} = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|}{c} + \hat{\boldsymbol{\eta}}_j - \hat{\boldsymbol{\delta}}_i, \tag{3}$$

which shares the same structure with the TOA formula in Eq. (1).

By using TOA information $\mathbf{t}_{i,j}$ only in Eq. (1), the transformation of TOA formula in Eq. (1) can be rewritten as [29]

$$\dot{\mathbf{t}}_{i,j} = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|}{c} - \dot{\boldsymbol{\delta}}_i + \dot{\boldsymbol{\eta}}_j, \tag{4}$$

where $\dot{\mathbf{t}}_{i,j}$ is the transformation of TOA $\mathbf{t}_{i,j}$, and $\dot{\boldsymbol{\delta}}_i$ and $\dot{\boldsymbol{\eta}}_j$ are the transformation of microphone start time $\boldsymbol{\delta}_i$ and source emission time $\boldsymbol{\eta}_j$, respectively.

Similarity, by using TDOA information $\boldsymbol{\zeta}_{i,j}$ only in Eq. (2), the transformation of TDOA formula in Eq. (2) is rewritten as [29]

$$\ddot{\boldsymbol{\zeta}}_{i,j} = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|}{c} - \ddot{\boldsymbol{\delta}}_i + \ddot{\boldsymbol{\eta}}_j, \tag{5}$$

where $\ddot{\boldsymbol{\zeta}}_{i,j}$ is the transformation of TDOA $\boldsymbol{\zeta}_{i,j}$, and $\ddot{\boldsymbol{\delta}}_i$ and $\ddot{\boldsymbol{\eta}}_j$ are pseudo start time and pseudo emission time, respectively, and

$$\begin{cases} \ddot{\boldsymbol{\eta}}_j = \dot{\boldsymbol{\eta}}_j \\ \ddot{\boldsymbol{\delta}}_i = \dot{\boldsymbol{\delta}}_i \\ \ddot{\boldsymbol{\zeta}}_{i,j} = \dot{\mathbf{t}}_{i,j} \end{cases}, \tag{6}$$

where $i = 1, \ldots, M$, $j = 1, \ldots, N$ and $\ddot{\eta}_1 = \dot{\eta}_1 = 0$.

Upon inspection of Eqs. (4), (5) and (6), it is obvious that the transformation of TDOA formula in Eq. (2) can be same as the transformation of TOA formula in Eq. (1) [29]. Without further mention, we denote $\dot{\mathbf{t}}_{i,j}$ as both TOA and TDOA, and also denote $\dot{\boldsymbol{\delta}}_i$ and $\dot{\boldsymbol{\eta}}_j$ as microphone start time and source emission time, respectively. Our objective is to estimate the UTIm from TOA or TDOA information, used for synchronizing microphones and sources in ad-hoc WASN.

4. Preliminaries

The low-rank structure information of UTIm exploited by LRP [28] is stated in this section.

By taking the square of both sides of Eq. (4) [33,37], we can have

$$\frac{\mathbf{r}_i^T\mathbf{r}_i + \mathbf{s}_j^T\mathbf{s}_j - 2\mathbf{r}_i^T\mathbf{s}_j}{c^2} = \dot{\mathbf{t}}_{i,j}^2 + \dot{\boldsymbol{\eta}}_j^2 + \dot{\boldsymbol{\delta}}_i^2 - 2(\dot{\mathbf{t}}_{i,j}\dot{\boldsymbol{\eta}}_j - \dot{\mathbf{t}}_{i,j}\dot{\boldsymbol{\delta}}_i + \dot{\boldsymbol{\eta}}_j\dot{\boldsymbol{\delta}}_i), \tag{7}$$

where $i = 1, \ldots, M$ and $j = 1, \ldots, N$.

To formulate LRP, we can sequentially subtract the corresponding equation for $i = 1$ and the equation for $j = 1$ and then add the equation for $i = j = 1$ from the general form Eq. (7), finally based on the assumption $\boldsymbol{\eta}_1 = 0$ due to the invariance of relative time, it follows [33,37]

$$\begin{aligned} &\frac{-2(\mathbf{r_i} - \mathbf{r}_1)^T(\mathbf{s}_j - \mathbf{s}_1)}{c^2} \\ &= \dot{\mathbf{t}}_{i,j}^2 - \dot{\mathbf{t}}_{i,1}^2 - \dot{\mathbf{t}}_{1,j}^2 + \dot{\mathbf{t}}_{1,1}^2 + 2\dot{\boldsymbol{\delta}}_i(\dot{\mathbf{t}}_{i,j} - \dot{\mathbf{t}}_{i,1}) \\ &- 2\dot{\boldsymbol{\delta}}_1(\dot{\mathbf{t}}_{1,j} - \dot{\mathbf{t}}_{1,1}) - 2\dot{\boldsymbol{\eta}}_j(\dot{\mathbf{t}}_{i,j} - \dot{\mathbf{t}}_{1,j}) + 2\dot{\boldsymbol{\eta}}_j(\dot{\boldsymbol{\delta}}_1 - \dot{\boldsymbol{\delta}}_i), \end{aligned} \tag{8}$$

where $i = 2, \ldots, M$ and $j = 2, \ldots, N$. By defining four matrices $\mathbf{R} \in \mathbb{R}^{3\times(M-1)}$, $\mathbf{S} \in \mathbb{R}^{3\times(N-1)}$, $\mathbf{D} \in \mathbb{R}^{(M-1)\times(N-1)}$ and $\mathbf{U} \in \mathbb{R}^{(M-1)\times(N-1)}$ as

$$\begin{aligned} &\mathbf{R}_{:,i-1} = \mathbf{r}_i - \mathbf{r}_1, \\ &\mathbf{S}_{:,j-1} = \mathbf{s}_j - \mathbf{s}_1, \\ &\mathbf{D}_{i-1,j-1} = \dot{\mathbf{t}}_{i,j}^2 - \dot{\mathbf{t}}_{i,1}^2 - \dot{\mathbf{t}}_{1,j}^2 + \dot{\mathbf{t}}_{1,1}^2, \\ &\mathbf{U}_{i-1,j-1} = 2\dot{\boldsymbol{\delta}}_i(\dot{\mathbf{t}}_{i,j} - \dot{\mathbf{t}}_{i,1}) - 2\dot{\boldsymbol{\delta}}_1(\dot{\mathbf{t}}_{1,j} - \dot{\mathbf{t}}_{1,1}) - 2\dot{\boldsymbol{\eta}}_j(\dot{\mathbf{t}}_{i,j} - \dot{\mathbf{t}}_{1,j}) + 2\dot{\boldsymbol{\eta}}_j(\dot{\boldsymbol{\delta}}_1 - \dot{\boldsymbol{\delta}}_i), \end{aligned}$$

respectively, where $i = 2, \ldots, M$ and $j = 2, \ldots, N$, then Eq. (8) is rewritten as matrix form [33,37]

$$\frac{-2\mathbf{R}^T\mathbf{S}}{c^2} = \mathbf{D} + \mathbf{U}. \tag{9}$$

Upon inspection of Eq. (9), the left side contains the information on the unknown location of microphones and sources while the right side contains UTIm and TOA (or TDOA), it implies that we need to estimate the UTIm before localizing both microphones and sources. Then with Eq. (9), the low-rank structure information of UTIm in TOA (or TDOA) exploited by LRP [28] is presented here.

LRP: if

$$\begin{cases} M - 1 > 3 \\ N - 1 > 3 \end{cases}, \tag{10}$$

LRP can be stated as

$$rank(\mathbf{D} + \mathbf{U}) = rank(\mathbf{R}^T\mathbf{S}) \le 3. \tag{11}$$

Upon inspection of Eq. (11), matrix $\mathbf{U}$ contains both the UTIm and TOA (or TDOA), and matrix $\mathbf{D}$ contains TOA (or TDOA), it indicates that Eq. (11) shows the low-rank structure information of UTIm in the known TOA (or TDOA). However, usually the initialization of UTIm is random, so that the solution of UTIm is easy to stuck into local minima. It leads the following problems: First, if the number of microphones or sources is not sufficient, the convergence rate is limited, for example, if number of microphones (or sources) is less than seven (or six), the convergence rate achieved by LRP is almost zero percent. Second, the recovery rate is still limited whatever the number of microphones and sources is. Third, the estimation error of UTIm tends to be large when there are noises in TOA (or TDOA) measurements. Therefore, we are interested in investigating whether there are any other low-rank structure information of UTIm in

TOA (or TDOA), i.e., given the additional linear constraints formulated by these low-rank structure information, we are aiming to improve both the recovery and convergence rate of UTIm and reduce the estimation errors of UTIm in noise environments, facilitating the localization task of ad-hoc WASN.

5. Proposed new low-rank properties

In this section, we propose three variants of the LRP to estimate UTIm. LRP reveals the low-rank structure information of UTIm in TOA (or TDOA), and the proposed three variants enhance this property by incorporating additional low-rank structure information. These additional structures enable the recovery and convergence rates for UTIm estimation to be improved, while reducing estimation errors, especially in noisy environments.

The key motivation behind our method is to exploit the low-rank structure inherent in the timing discrepancies (UTIm) of TOA and TDOA measurements. This structure arises from the dependencies between the microphone and source timing in ad-hoc networks. By utilizing these low-rank properties, we can formulate linear constraints that provide more robust and efficient estimation of UTIm, even under noisy conditions and random initializations.

Our three variants of LRP differ from existing methods in that they offer more comprehensive low-rank structure information, leading to better performance in terms of convergence, recovery rate, and noise tolerance. Below, we define three new matrices, each of which captures a distinct aspect of the low-rank structure in UTIm, enabling the estimation of UTIm to benefit from richer structural information.

By defining three new matrices

$$\mathbf{T_1}^* = \begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix} \in \mathbb{R}^{(M-1)\times 2(N-1)}, \tag{12}$$

$$\mathbf{T}_2^* = \begin{bmatrix}\mathbf{D}^T & \mathbf{U}^T\end{bmatrix} \in \mathbb{R}^{(N-1)\times 2(M-1)}, \tag{13}$$

and

$$\mathbf{T}_3^* = \begin{bmatrix}\mathbf{D} & \mathbf{U} \\ \mathbf{U} & \mathbf{D}\end{bmatrix} \in \mathbb{R}^{2(M-1)\times 2(N-1)}, \tag{14}$$

then three variants of LRP are proposed if Eq. (10) holds.

LRPV1:

$$rank\left(\mathbf{T}_1^*\right) \le min\{M-1, N-1+3\}, \tag{15}$$

where matrix $\mathbf{T}_1^*$ is low-rank only if $M-1 > N-1+3$ (see proof in *Section A1 of Supplementary Material*).

LRPV2:

$$rank\left(\mathbf{T}_2^*\right) \le min\{N-1, M-1+3\}, \tag{16}$$

where matrix $\mathbf{T}_2^*$ is low-rank only if $N-1 > M-1+3$ (see proof in *Section A2 of Supplementary Material*).

LRPV3:

$$rank\left(\mathbf{T}_3^*\right) \le min\{N-1+3, M-1+3\}, \tag{17}$$

where matrix $\mathbf{T}_3^*$ is always low-rank if Eq. (10) holds (see proof in *Section A3 of Supplementary Material*).

From Eqs. (12), (13), (14), (15), (16) and (17), we can see that LRPV1, LRPV2 and LRPV3 always reveal low-rank structure information of UTIm in known TOA (or TDOA) if $M-1 > N-1+3$ holds for $\mathbf{T}_1^*$ and $N-1 > M-1+3$ holds for $\mathbf{T}_2^*$. The proposed LRP variants extend traditional LRP methods by introducing new constraints and structures that increase the robustness of UTIm estimation. In particular, these variants mitigate the issues of local minima and suboptimal solutions that often arise in traditional LRP methods, especially in the presence of noise or random initializations. The three variants each utilize a distinct structure, allowing for a more comprehensive exploration of the low-rank properties of UTIm in both TOA and TDOA measurements.

6. Proposed algorithm

This section illustrates the corresponding linear constraints first based on low-rank structure information exploited by both LRP in Section 4 and three variants of LRP in Section 5, and then applies them to UTIm estimation.

Both the low-rank property (LRP) and the three proposed variants of LRP exploit the inherent low-rank structure information present in UTIm within TOA (or TDOA) measurements. The motivation behind using these variants is to draw from a richer pool of low-rank structural information compared to traditional LRP methods. This richer set of information allows us to better estimate UTIm, leading to a global solution rather than getting stuck in sub-optimal or local minimum values caused by the randomness of UTIm initialization.

In practical terms, the procedure consists of two main steps. First, based on the four LRP formulations discussed above, we derive four linear constraints. These constraints are then combined into a single objective function. To solve this objective function, we apply the Gaussian-Newton method [42], which helps mitigate the issues of local minima by improving convergence rates and recovery performance, even in the presence of noisy measurements and random initialization.

6.1. Linear constraint based on LRP

From Eqs. (9) and (11), denote

$$\begin{cases}\mathbf{D} = \begin{bmatrix}\mathbf{A} & \mathbf{B}\end{bmatrix} \\ \mathbf{U} = \begin{bmatrix}\mathbf{F} & \mathbf{G}\end{bmatrix}\end{cases}, \tag{18}$$

where the sub-matrices in $\mathbf{D}$ and $\mathbf{U}$ are $\mathbf{A} \in \mathbb{R}^{(M-1)\times 3}$, $\mathbf{B} \in \mathbb{R}^{(M-1)\times(N-1-3)}$, $\mathbf{F} \in \mathbb{R}^{(M-1)\times 3}$ and $\mathbf{G} \in \mathbb{R}^{(M-1)\times(N-1-3)}$. Then we have

$$rank(\begin{bmatrix}\mathbf{A}+\mathbf{F} & \mathbf{B}+\mathbf{G}\end{bmatrix}) = rank(\mathbf{A}+\mathbf{F}) \le 3. \tag{19}$$

From Eq. (19), we can assume there is a matrix $\mathbf{X}$ that enables

$$(\mathbf{A}+\mathbf{F})\mathbf{X} = \mathbf{B}+\mathbf{G}, \tag{20}$$

where $\mathbf{X} \in \mathbb{R}^{3\times(N-1-3)}$ is unknown and will be estimated in Section 6.5.

6.2. Linear constraint based on LRPV1

From Eqs. (12) and (15), we split $\mathbf{T}_1^*$ as $\mathbf{T}_1^* = \begin{bmatrix}\mathbf{T}_{11}^* & \mathbf{T}_{12}^*\end{bmatrix}$, where $\mathbf{T}_{11}^* \in \mathbb{R}^{(M-1)\times(N-1+3)}$ and $\mathbf{T}_{12}^* \in \mathbb{R}^{(M-1)\times(N-1-3)}$, then we have

$$rank(\begin{bmatrix}\mathbf{T}_{11}^* & \mathbf{T}_{12}^*\end{bmatrix}) = rank(\mathbf{T}_{11}^*) \le N-1+3. \tag{21}$$

From Eq. (21), we can assume there is a matrix $\mathbf{Z}$ that enables

$$\mathbf{T}_{11}^*\mathbf{Z} = \mathbf{T}_{12}^*, \tag{22}$$

where $\mathbf{Z} \in \mathbb{R}^{(N-1+3)\times(N-1-3)}$ is unknown and will be estimated in Section 6.5.

6.3. Linear constraint based on LRPV2

From Eqs. (13) and (16), we can split $\mathbf{T}_2^*$ as $\mathbf{T}_2^* = \begin{bmatrix}\mathbf{T}_{21}^* & \mathbf{T}_{22}^*\end{bmatrix}$ where $\mathbf{T}_{21}^* \in \mathbb{R}^{(N-1)\times(M-1+3)}$ and $\mathbf{T}_{22}^* \in \mathbb{R}^{(N-1)\times(M-1-3)}$. Then we have

$$rank(\begin{bmatrix}\mathbf{T}_{21}^* & \mathbf{T}_{22}^*\end{bmatrix}) = rank(\mathbf{T}_{21}^*) \le M-1+3. \tag{23}$$

From Eq. (23), we can assume there is a matrix $\mathbf{W}$ that enables

$$\mathbf{T}_{21}^*\mathbf{W} = \mathbf{T}_{22}^*, \tag{24}$$

where $\mathbf{W} \in \mathbb{R}^{(M-1+3)\times(M-1-3)}$ is unknown and will be estimated in Section 6.5.

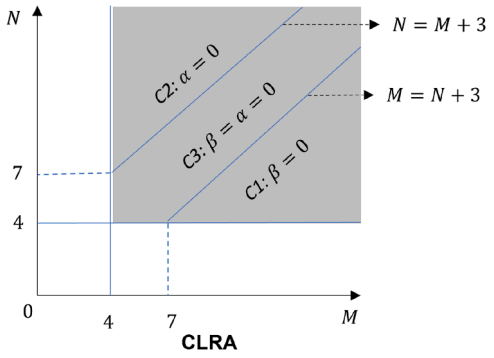


Fig. 1. The illustration for combinations of four low-rank properties of proposed CLRA method with different number of microphones M and sources N (α for LRPV1, β for LRPV2, γ for LRPV3; $\alpha = \beta = \gamma = 0$ in the shadow area for STLS [23]).

6.4. Linear constraint based on LRPV3

From Eqs. (14) and (17), we define M_N as $min(N-1+3, M-1+3)$ and split $\mathbf{T}_3^*$ as $\mathbf{T}_3^* = \begin{bmatrix}\mathbf{T}_{31}^* & \mathbf{T}_{32}^*\end{bmatrix}$ where $\mathbf{T}_{31}^* \in \mathbb{R}^{2(M-1)\times M_N}$ and $\mathbf{T}_{32}^* \in \mathbb{R}^{2(M-1)\times(2(N-1)-M_N)}$. Then we have

$$rank(\begin{bmatrix}\mathbf{T}_{31}^* & \mathbf{T}_{32}^*\end{bmatrix}) = rank(\mathbf{T}_{31}^*) \le M_N. \tag{25}$$

From Eq. (25), we can assume there is a matrix $\mathbf{Y}$ that enables

$$\mathbf{T}_{31}^*\mathbf{Y} = \mathbf{T}_{32}^*, \tag{26}$$

where $\mathbf{Y} \in \mathbb{R}^{M_N\times(2(N-1)-M_N)}$ is unknown and will be estimated in Section 6.5.

6.5. Algorithm

The STLS [23,38,40] utilizes the LRP only for UTIm estimation, so that the solution of UTIm is easy to stuck into local minima, this limits both the recovery and convergence rate for UTIm estimation, and the estimation errors of UTIm tends to be large in noise environments. To find global optimal solution for UTIm, based on the three variants of LRP we proposed and the linear constraints formulated from those four low-rank properties, LRP, LRPV1, LRPV2 and LRPV3, we have the objective function:

$$\begin{aligned} f(\dot{\boldsymbol{\delta}}, \dot{\boldsymbol{\eta}}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}) &= \|\mathbf{U}\|_F^2 + \lambda^2\|(\mathbf{A}+\mathbf{F})\mathbf{X} - (\mathbf{B}+\mathbf{G})\|_F^2 \\ &+ \alpha^2\|\mathbf{T}_{11}^*\mathbf{Z} - \mathbf{T}_{12}^*\|_F^2 + \beta^2\|\mathbf{T}_{21}^*\mathbf{W} - \mathbf{T}_{22}^*\|_F^2 + \gamma^2\|\mathbf{T}_{31}^*\mathbf{Y} - \mathbf{T}_{32}^*\|_F^2, \end{aligned} \tag{27}$$

where $\|\bullet\|_F$ is Frobenius norm, $\|\mathbf{U}\|_F^2$ is regularization term [23,38], and λ, α, β and γ are penalty parameters for the four low-rank properties, respectively (λ for LRP; α for LRPV1; β for LRPV2; γ for LRPV3). Upon inspection of Eq. (27), We can see that when $\gamma = \alpha = \beta = 0$, proposed CLRA method degenerates to STLS that uses the LRP only for UTIm estimation [23,38].

In addition, for any given number of microphones M and any given number of sources N, we can distinguish three cases, i.e., 1) Case 1 (C1): $M - N > 3$; 2) Case 2 (C2): $N - M > 3$; 3) Case 3 (C3): $|M - N| \le 3$. From $M - 1 > N - 1 + 3$ in Eq. (15) and $N - 1 > M - 1 + 3$ in Eq. (16), we can see that LRPV1 and LRPV2 have low-rank properties only in C1 and C2, respectively, then we can summarize the combination way for four low-rank properties as

(1) $\beta = 0$ in C1;

(2) $\alpha = 0$ in C2; and

(3) $\alpha = 0$ and $\beta = 0$ in C3. Fig. 1 illustrates the combination way of four low-rank properties for CLRA.

Next, we present a general and complete solution for the minimization of above objective function. We use the column-wise matrix vectorization, $\upsilon(\cdot)$ (i.e., $\upsilon(\mathbf{X}) = [\mathbf{X}_{1,1}, \ldots, \mathbf{X}_{3,1}, \mathbf{X}_{1,2}, \ldots, \mathbf{X}_{3,2}, \ldots, \mathbf{X}_{1,(N-1-3)}, \ldots, \mathbf{X}_{3,(N-1-3)}]^T$), to define

$$\mathbf{p} = \begin{bmatrix}\dot{\boldsymbol{\delta}}^T & \dot{\boldsymbol{\eta}}^T & \upsilon(\mathbf{X})^T & \upsilon(\mathbf{Y})^T & \upsilon(\mathbf{Z})^T & \upsilon(\mathbf{W})^T\end{bmatrix}^T,$$

and

$$\mathbf{q} = \begin{bmatrix}\mathbf{f_A}^T & \lambda\mathbf{f_B}^T & \gamma\mathbf{f_C}^T & \alpha\mathbf{f_D}^T & \beta\mathbf{f_E}^T\end{bmatrix}^T,$$

where

$$\begin{cases} \mathbf{f_A} = \upsilon(\mathbf{U}) \\ \mathbf{f_B} = \upsilon((\mathbf{A}+\mathbf{F})\mathbf{X} - \mathbf{B} - \mathbf{G}) \\ \mathbf{f_C} = \upsilon(\mathbf{T}_{31}^*\mathbf{Y} - \mathbf{T}_{32}^*) \\ \mathbf{f_D} = \upsilon(\mathbf{T}_{11}^*\mathbf{Z} - \mathbf{T}_{12}^*) \\ \mathbf{f_E} = \upsilon(\mathbf{T}_{21}^*\mathbf{W} - \mathbf{T}_{22}^*) \end{cases}, \tag{28}$$

and then re-write the objective function Eq. (27) as $f(\mathbf{p}) = \|\mathbf{q}\|_2^2$, where the dimension of the vectors $\mathbf{q}$ and $\mathbf{p}$ are $Q = (M-1)(8(N-1) - 2M_N - 6) - 3(N-1)$ and $P = M + N - 1 + 3(N-1-3) + M_N(2(N-1) - M_N) + (N-1+3)(N-1-3) + (M-1+3)(M-1-3)$, respectively (Algorithm 6.5).

Finally, we solve the nonlinear least square problem by minimizing $f(\mathbf{p})$ with the Gauss-Newton algorithm [42]. Therefore we compute the Jacobian matrix

$$\mathbf{J} = \partial\mathbf{q}/\partial\mathbf{p} \in \mathbb{R}^{Q\times P}, \tag{29}$$

which is the derivation of vector $\mathbf{q}$ with respect to vector $\mathbf{p}$ (see details in *Section A4 of Supplementary Material*), and we update $\mathbf{p}$ in an iterative way as

$$\mathbf{p}^{m+1} = \mathbf{p}^m - (\mathbf{J}^{mT}\mathbf{J}^m)^{-1}\mathbf{J}^{mT}\mathbf{q}^m, \tag{30}$$

Algorithm 1 Proposed CLRA method for estimation of UTIm

Input: D, U, M, N;
Output: $\rho = [\dot{\delta}; \dot{\eta}; x; y; z; w]$;
Parameters: ω^*: threshold for divergence;
d_ρ: stopping threshold for Gaussian-Newton iterations;
m_2: maximum number of iterations;
λ, α, β and γ: penalty constant in (27)
  if $|M - N| < 3$ **then**
    $\alpha = 0$ and $\beta = 0$;
  elseif $M > N + 3$ **then**
    $\beta = 0$;
  else
    $\alpha = 0$;
  end if
Initialize $\delta^{(0)}$, $\eta^{(0)}$
Compute D, A, B, $X^{(0)} = (A+F)^\dagger(B+G)$; (9, 18, 20)
  $Z^{(0)} = T_{11}^{*\dagger}T_{12}^*$, $W^{(0)} = T_{21}^{*\dagger}T_{22}^*$, $Y^{(0)} = T_{31}^{*\dagger}T_{32}^*$ (22,24,26)
  (†: Moore-Penrose generalized inverses [40])
while $g == 0$ **do**
Step 1: Compute U^{m_1}, F^{m_1}, G^{m_1}; (9)(18)
Step 2: Compute $T_1^{*m_1}$, $T_{11}^{*m_1}$, $T_{12}^{*m_1}$ (22)
Step 3: Compute $T_2^{*m_1}$, $T_{21}^{*m_1}$, $T_{22}^{*m_1}$ (24)
Step 4: Compute $T_3^{*m_1}$, $T_{31}^{*m_1}$, $T_{32}^{*m_1}$ (26)
Step 5:
Compute $f^m = [f_A^{Tm_1}\ \lambda f_B^{Tm_1}\ \gamma f_C^{Tm_1}\ \lambda f_D^{Tm_1}\ \beta f_E^{Tm_1}]^T$ (28)
Step 6:Compute J^{m_1} (29)
Step 7: Update $\rho^{(m_1+1)}$ (30)
  if $\left(\|f_B^{m_1}\|_2^2 + \|f_C^{m_1}\|_2^2 + \|f_D^{m_1}\|_2^2 + \|f_E^{m_1}\|_2^2\right) > \omega^*$
    OR $\left(\|p^{(m_1+1)} - p^{(m_1)}\|\right) < d_\rho$
    OR $m_1 = m_2$; **then**
      $g = 1$;
  end if
  $m_1 = m_1 + 1$;
end while

where m denotes m^{th} iteration and the pseudo code of proposed CLRA can be seen from ***Algorithm***.

7. Experimental results

The experimental settings are described in Section 7.1, and the influence of the parameters in the three proposed LRP variants on the CLRA method is analyzed in Section 7.2. Sections 7.3 and 7.4 present the performance of the proposed CLRA method. Finally, Section 7.5 provides a consolidated evaluation, including performance comparisons with STLS and the auxiliary function-based algorithm [12] (Ono), as well as a robustness analysis under additive noise on both the realistic simulation dataset and the real dataset; in this part, we also report results for SDP (semidefinite programming) and SDP + LM (Levenberg-Marquardt) [31], where LM is initialized by SDR.

7.1. Setup

The simulation data for microphone and source locations, start times, and emission times were randomly generated using MATLAB R2019a on a computer with a 3.7-GHz CPU, six cores, and 16 GB RAM. Microphone and source locations followed a uniform distribution within 10m × 10m × 3m, which is suitable for real applications [31]. Start times ($\boldsymbol{\delta}$) and emission times ($\boldsymbol{\eta}$) were uniformly distributed within $[-1, 1]$s [31], and the speed of sound was set to $c = 340$m/s. We also use both realistic simulation data and real data for validating proposed methods (see Section 7.5). The parameters of the proposed CLRA were set as $w^* = 10^{30}$, $d_p = 10^{-9}$, and $m_2 = 100$. The process was terminated if the objective function exceeded 10^{30}, the variable $\mathbf{p}$ changed by less than 10^{-9}, or the iteration count exceeded 100.

Multiple initializations were tested for each configuration to evaluate CLRA's robustness against initialization variations. For a specific configuration, only one initialization for the microphones' start times and sources' emission times is required. However, some random initializations may not achieve the globally optimal solution for UTIm. To demonstrate the performance, the recovery rate ($Rr(M, N)$) was calculated as:

$$Rr(M,N) = \frac{\sum_{i=1}^{Nc(M,N)} Ne_i(M,N)}{I_n(M,N)Nc(M,N)}, \tag{31}$$

where $Ne_i(M, N)$ represents the number of globally optimal solutions for the ith configuration, $Nc(M, N)$ is the total number of configurations, and $I_n(M, N)$ is the total initializations for each configuration.

Microphone sampling rates, typically below 100 kHz, introduce errors in TOA (or TDOA) of approximately 10^{-5}s when using GCC-PHAT [10]. For example, a sampling rate of 44.1 kHz can introduce an error of 1.2798×10^{-5}s. Based on this principle, UTIm recovery is deemed successful if the difference between the estimation and ground truth is less than 10^{-4}s. The error metric (er) is defined as:

$$er = \frac{\sum_{i=1}^{M} \|\dot{\boldsymbol{\delta}}_i - \hat{\dot{\boldsymbol{\delta}}}_i\|}{M} + \frac{\sum_{j=1}^{N} \|\dot{\boldsymbol{\eta}}_j - \hat{\dot{\boldsymbol{\eta}}}_j\|}{N}, \tag{32}$$

where $\dot{\boldsymbol{\delta}}_i$ and $\dot{\boldsymbol{\eta}}_j$ are the ground truth values, and $\hat{\dot{\boldsymbol{\delta}}}_i$ and $\hat{\dot{\boldsymbol{\eta}}}_j$ are the estimated values.

To assess robustness across configurations, the convergence rate ($Cr(M, N)$) is defined as:

$$Cr(M,N) = \frac{Ce(M,N)}{Nc(M,N)}, \tag{33}$$

where $Ce(M, N) = \sum_{i=1}^{Nc(M,N)} sr_i$ counts the number of successful recoveries, and $sr_i = 1$ at least one initialization succeeds in the ith configuration, otherwise, $sr_i = 0$. The convergence rate approaches 100% when the number of microphones or sources is sufficient, demonstrating the robustness of the proposed CLRA even under suboptimal configurations.

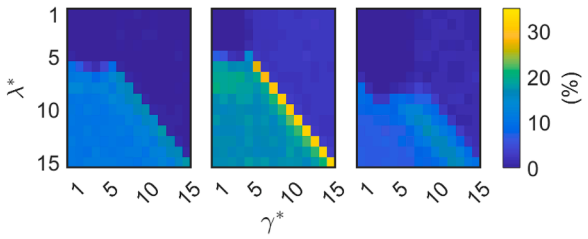


Fig. 2. The parameters impact (λ^* for LRP and γ^* for LRPV3) on proposed CLAR1 in terms of recovery rate with three different cases (Left: $M = 10$, $N = 10$; Middle: $M = 15$, $N = 8$; Right: $M = 8$, $N = 15$).

7.2. Parameters analysis

This subsection analyzes the parameter settings of the three proposed LRP variants within CLRA. We denote the variants as: (1) CLRA1 when $\alpha = \beta = 0$ (LRP + LRPV3), (2) CLRA2 when $\alpha = \gamma = 0$ (LRP + LRPV2), and (3) CLRA3 when $\beta = \gamma = 0$ (LRP + LRPV1). We set $N_c(M, N) = 10$ configurations and $I_n(M, N) = 100$ initializations per configuration, yielding 1000 implementations for fixed M and N to assess parameter effects.

As shown in Fig. 1, $\beta = 0$ in C1 and $\alpha = 0$ in C2; in C3, both α and β are zero. Even when $\alpha = \beta = 0$, CLRA applies to different (M, N) because LRPV3 remains low-rank in C1-C3. Accordingly, we evaluate CLRA1 by varying λ and γ (Eq. (27)) under (C1) $(M, N) = (15, 8)$, (C2) $(8, 15)$, and (C3) $(10, 10)$. We then analyze β for CLRA2 with $(M, N) = (15, 8)$, and α for CLRA3 with $(M, N) = (8, 15)$.

(1) Parameters λ and γ analysis for proposed CLRA1: We set parameters $\lambda = 10^{\lambda^*}$ and $\gamma = 10^{\gamma^*}$ for proposed CLRA1 in Eq. (27) and vary both λ^* and γ^* from 1 to 15. Fig. 2 shows the effect of λ^* and γ^* on proposed CLRA1 with different number of microphones and sources. As can be seen from Fig. 2 (left and middle sub-figures), the effects of parameters λ^* and γ^* on the recovery rate of CLRA1 is similar, when both λ^* and γ^* are small (less than 5), the recovery rate is almost 0%, and when $\lambda^* > 5$ and $\gamma^* \leq \lambda^*$, the recovery rate is larger than 0%. This indicates that λ^* and γ^* have a joint effect on proposed CLRA1. In addition, we can see that when $\lambda^* = \gamma^*$, the peaks regarding recovery rate are achieved.

Besides, from right subfigure in Fig. 2, we can see that when λ^* is less than 8, the recovery rate for proposed CLRA1 is almost 0%, and when $\lambda^* > 8$ and $\gamma^* \leq \lambda^* + 2$, the recovery rate is larger than 0%. In addition, the peaks regarding recovery rate are obtained when $\lambda^* = \gamma^* + 3$. Thus, if there is no special mention, we set the following parameters for proposed CLRA1, i.e., $\lambda^* = 10$ and $\gamma^* = 10$ in both C1 and C3, $\lambda^* = 12$ and $\gamma^* = 9$ in C2.

(2) Parameter β analysis for proposed CLRA2: We set parameter $\lambda = 10^{10}$ and analyze the effect of parameter β on CLRA2. Denote $\beta = 10^{\beta^*}$ and we vary β^* from 1 to 15, then Fig. 3 plots the recovery rate of proposed CLRA2 when β^* varies. It can be seen that β^* has a big effect on the performance of proposed CLRA2. In details, when β^* is small (less than 8), the recovery rate is stable (about 8%) and when β^* reaches to 11, the proposed CLRA2 can achieve about 28% recovery rate, then the recovery rate decreases once β^* increases further. Thus, we set $\beta^* = 11$ for the proposed CLRA2 if there is no special mention.

(3) Parameter α analysis for proposed CLRA3: We set parameter $\lambda = 10^{10}$ and analyze the effect of parameter α on CLRA3. Denote $\alpha = 10^{\alpha^*}$ and we vary α^* from 1 to 15. It can be seen from Fig. 3 that the parameter α^* also has a big effect on proposed CLRA3. In details, when $\alpha^* < 9$, the recovery rate of proposed CLRA3 is stable (about 17%), then the recovery rate increases once the value of α^* continues to increase, and we can see that the proposed CLRA3 can reach about 37% recovery rate

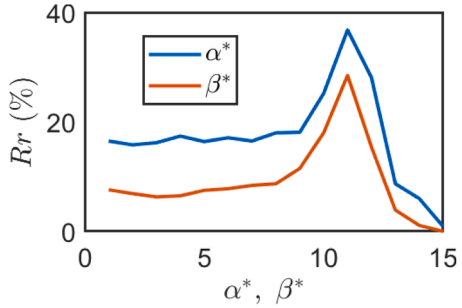


Fig. 3. The effect of β^* on CLRA2 with $M = 8$, $N = 15$ and α^* on CLRA3 with $M = 15$, $N = 8$.

when $\alpha^* = 11$, thus, we choose $\alpha^* = 11$ for proposed CLRA3 if no special mention.

7.3. Comparison of the performance for low-rank properties

This subsection shows the performance of the proposed CLRA methods in comparison with the STLS [23,37,38].

Upon examining the linear constraints in Eq. (20) formulated by LRP, it is evident that the number of equations, given by $(M-1)(N-1-3)$, must be greater than or equal to the number of unknowns, expressed as $3(N-1-3)+M+N-1$. This requirement leads to the condition $(M-5)(N-5) \geq 8$, which indicates that the minimal number of microphones required for UTIm is 6, 7, or 8, while the minimal number of sources is 13, 9, or 8, respectively, depending on the specific configuration. Similarly, the reverse is also true. Since proposed three variants of LRP are the additional LRP for UTIm based on LRP, thus, they should also follow the lower boundary of M and N for UTIm estimation with their corresponding additional constraints.

The parameter $\lambda = 10^{10}$ for STLS in Eq. (27) ($\gamma = \alpha = \beta = 0$). For the proposed CLAR method that combines one proposed low-rank property with LRP, i.e., CLRA1, CLRA2 and CLRA3, the parameters are illustrated already in Section 7.2. For the proposed CLAR method that utilizes all the four low-rank properties, we categorize them as three cases (see Fig. 1), (1) C3: the parameters λ and γ of proposed CLRA in Eq. (27) ($\alpha = \beta = 0$) are set to be 10^{10} and 10^{10}, respectively. (2) C1: we set the parameters $\lambda = 10^{10}$, $\gamma = 10^{10}$ and $\alpha = 10^{11}$ for proposed CLRA in Eq. (27) ($\beta = 0$). (3) C2: we set the parameters $\lambda = 10^{12}$, $\gamma = 10^{9}$ and $\beta = 10^{13}$ for proposed CLRA in Eq. (27) ($\alpha = 0$). Besides, we vary both M and N from 1 to 15, and for each pair of fixed M and N, we the number of configuration $N_c(M,N) = 200$ and the number of initializations $I_n(M,N) = 100$ for each configuration.

7.3.1. Comparison of recovery rate

Fig. 4 compares the recovery rates of the proposed CLRA methods with STLS.

For STLS, the recovery rate is about 0% when $M \leq 6$ or $N \leq 6$. When $M \geq 7$ and $N \geq 7$, it ranges from 0% to 28%.

For CLRA1, the recovery rate matches STLS when $M \leq 6$ or $N \leq 5$. However, for $M = 7, N \geq 10$, it ranges from 0% to 8%, outperforming STLS. Additionally, when $M \geq 14, N = 6$, CLRA1 achieves about 1%, while STLS remains at 0%. For $M > 12, N > 8$, CLRA1 outperforms STLS by 10%–20%.

For CLRA2, which applies when $N > M+3$ and $M > 4$, it achieves 2%–4% recovery rates for $M = 6, N > 13$, exceeding STLS. When $M \geq 7$ and $N > M+3$, particularly for $7 \leq M \leq 10, N = 15$, CLRA2 surpasses STLS by 18%–27%.

For CLRA3, valid when $M > N+3$ and $N > 4$, it achieves 2%–4% recovery rates for $N = 6, M > 13$, exceeding STLS. When $N \geq 7, M > N+3$, particularly for $M = 15, 7 \leq N \leq 10$, CLRA3 outperforms STLS by 12%–19%.

Finally, the combined CLRA method (LRP with LRPV1, LRPV2, and LRPV3) consistently achieves higher recovery rates than STLS for $M > 5, N > 5$. Combining all three variants further improves performance compared to using a single variant. These results validate the effectiveness of the proposed LRPV1, LRPV2, and LRPV3.

7.3.2. Comparison of convergence rate

Fig. 5(a) illustrates the convergence rate for STLS, while Fig. 5(b) shows the percentage point differences between STLS and the proposed CLRA methods.

For STLS, the convergence rate is 0% when $M < 7$ or $N < 6$, but increases with larger M and N. Notably, when $M \geq 13$ and $N \geq 12$, the convergence rate exceeds 90%.

For CLRA1, when $M < 6$ or $N < 6$, the convergence rate matches STLS at 0%. However, for $M > 10$ and $N > 9$, CLRA1 surpasses STLS by 2%–18%. In the range $6 \leq M \leq 9$ or $5 \leq N \leq 8$, CLRA1 achieves 6%–58% higher rates, validating the effectiveness of LRPV3 and CLRA1.

For CLRA2, when $M = 6, N > 12$, the convergence rate is 28%–63% higher than STLS. For $M \geq 7$ and $N > M+3$, particularly when $M = 7, N > 10$, CLRA2 exceeds STLS by 11%–50%. These results validate LRPV2 and CLRA2.

For CLRA3, when $N = 6, M > 12$, the convergence rate is 32%–58% higher than STLS. Similarly, for $N \geq 7$ and $M > N+3$, especially when $N = 7, M \geq 11$, it exceeds STLS by 32%–58%. For $N \geq 8, M > N+3$, the improvement ranges from 14%–19%. These results validate LRPV1 and CLRA3.

Finally, the combined CLRA method (LRP with LRPV1, LRPV2, and LRPV3) achieves better convergence rates than STLS for $M > 5, N > 5$. Furthermore, combining all three low-rank properties with LRP outperforms combining a single property, further validating LRPV1, LRPV2, and LRPV3.

7.4. Computational complexity analysis

In this subsection, we analyze the computational complexity of the proposed CLRA method and the STLS method (using only LRP) [23,37,38].

Both methods rely on Gauss-Newton optimization. The most computationally intensive part is the update of the variable $\mathbf{p}$ (see Eq. (30) and Algorithm). The operations in Eq. (30) involve: 1) Computing the product of two Jacobian matrices $\mathbf{J}^{(m)^T}\mathbf{J}^{(m)}$; 2) Inverting the matrix $(\mathbf{J}^{(m)^T}\mathbf{J}^{(m)})^{-1}$; 3) Multiplying $(\mathbf{J}^{(m)^T}\mathbf{J}^{(m)})^{-1}$ with $\mathbf{J}^{(m)^T}$.

For LRP only, the variable $\mathbf{p}$ has three sub-variables: $\dot{\boldsymbol{\delta}}$, $\dot{\boldsymbol{\eta}}$, and $\mathbf{X}$. The Jacobian size is: - Rows: $\hat{J}_{1,r} = (M-1)(2(N-1)-3)$ - Columns: $\hat{J}_{1,c} = M - 3^2 + (3+1)(N-1)$. The complexity is: $\mathcal{O}(\hat{J}^2_{1,c}\hat{J}_{1,r})$ for matrix multiplication, $\mathcal{O}(\hat{J}^3_{1,c})$ for matrix inversion, and $\mathcal{O}(\hat{J}^2_{1,c}\hat{J}_{1,r})$ for final multiplication. Thus, LRP's complexity is: $\mathcal{O}(\min(\hat{J}^2_{1,c}\hat{J}_{1,r}, \hat{J}^3_{1,c}))$.

For the CLRA method, the complexity depends on the number of microphones and sources, divided into three cases: C1, C2, and C3.

Case C3: Using LRP and LRPV3, the variable $\mathbf{p}$ has four sub-variables: $\dot{\boldsymbol{\delta}}$, $\dot{\boldsymbol{\eta}}$, $\mathbf{X}$, and $\mathbf{Y}$. The Jacobian size is: - Rows: $\hat{J}_{1,4,r} = (M-1)(6(N-1)-3-M_N)$ - Columns: $\hat{J}_{1,4,c} = M - 3^2 - M_N^2 + (3+1+2M_N)(N-1)$. The complexity is: $\mathcal{O}(\min(\hat{J}^2_{1,4,c}\hat{J}_{1,4,r}, \hat{J}^3_{1,4,c}))$.

Case C1: Using LRP, LRPV1, and LRPV3, $\mathbf{p}$ has five sub-variables: $\dot{\boldsymbol{\delta}}$, $\dot{\boldsymbol{\eta}}$, $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$. The Jacobian size is: - Rows: $\hat{J}_{1,2,4,r} = (M-1)(7(N-1)-23-M_N)$ - Columns: $\hat{J}_{1,2,4,c} = M - 23^2 - M_N^2 + (3+N+2M_N)(N-1)$. The complexity is: $\mathcal{O}(\min(\hat{J}^2_{1,2,4,c}\hat{J}_{1,2,4,r}, \hat{J}^3_{1,2,4,c}))$.

Case C2: Using LRP, LRPV2, and LRPV3, $\mathbf{p}$ has five sub-variables: $\dot{\boldsymbol{\delta}}$, $\dot{\boldsymbol{\eta}}$, $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{W}$. The Jacobian size is: - Rows: $\hat{J}_{1,2,4,r} = (M-1)(7(N-1)-3-M_N)-(N-1)3$ - Columns: $\hat{J}_{1,2,4,c} = M - 3^2 - M_N^2 + (3+1+2M_N)(N-1) + (M-1-3)(M-1+3)$. The complexity is: $\mathcal{O}(\min(\hat{J}^2_{1,2,4,c}\hat{J}_{1,3,4,r}, \hat{J}^3_{1,3,4,c}))$.

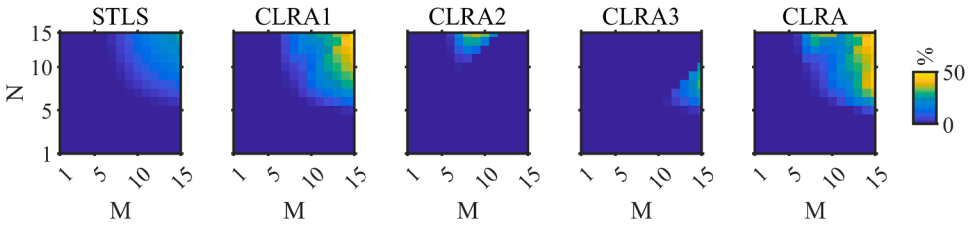


Fig. 4. The performance comparison for STLS and proposed CLRA (CLAR1: $\alpha = \beta = 0$; CLAR2: $\alpha = \gamma = 0$; CLAR3:$\beta = \gamma = 0$) in terms of recovery rate.

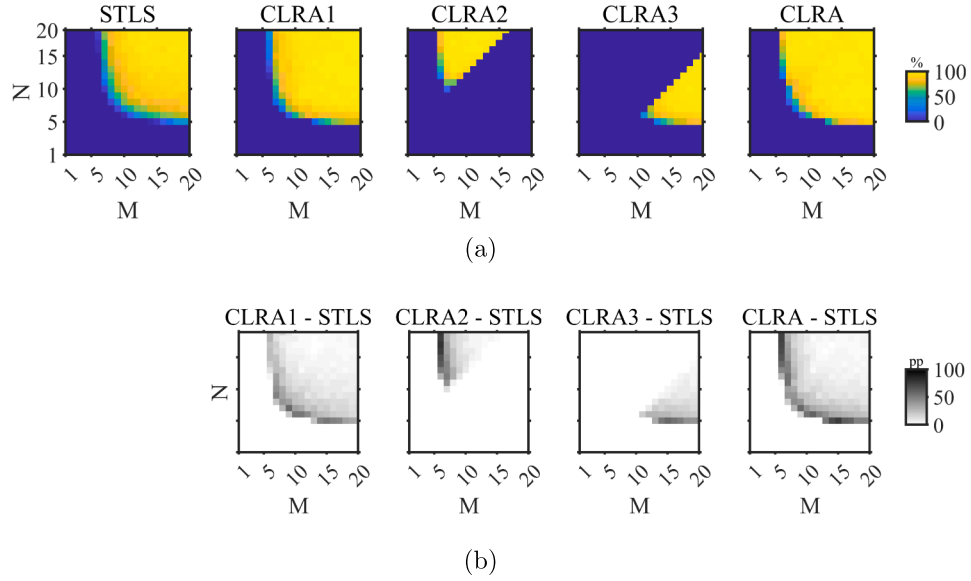


Fig. 5. The performance comparison for STLS and proposed CLRA in terms of (a) convergence rate; and (b) percentage point of convergence rate, where CLAR1: $\alpha = \beta = 0$, CLAR2: $\alpha = \gamma = 0$, CLAR3: $\beta = \gamma = 0$, pp: percentage point.

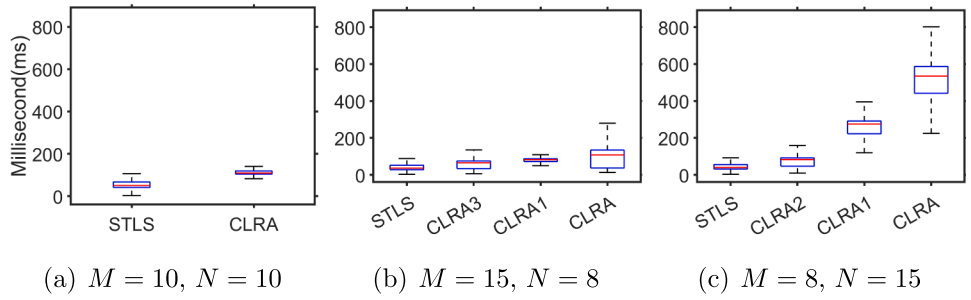


Fig. 6. The running time for proposed CLRA methods in comparison with STLS.

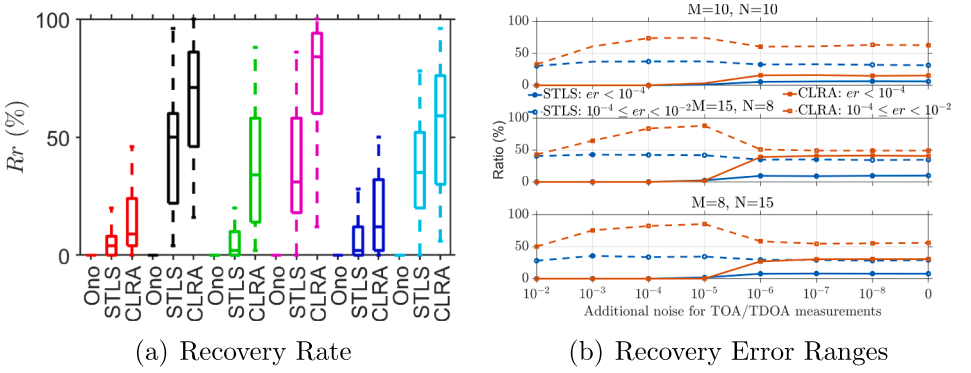
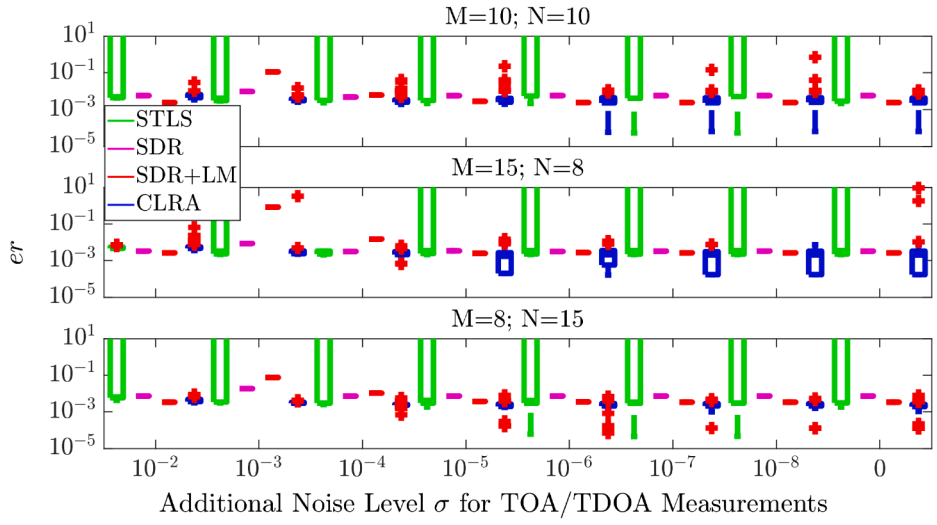
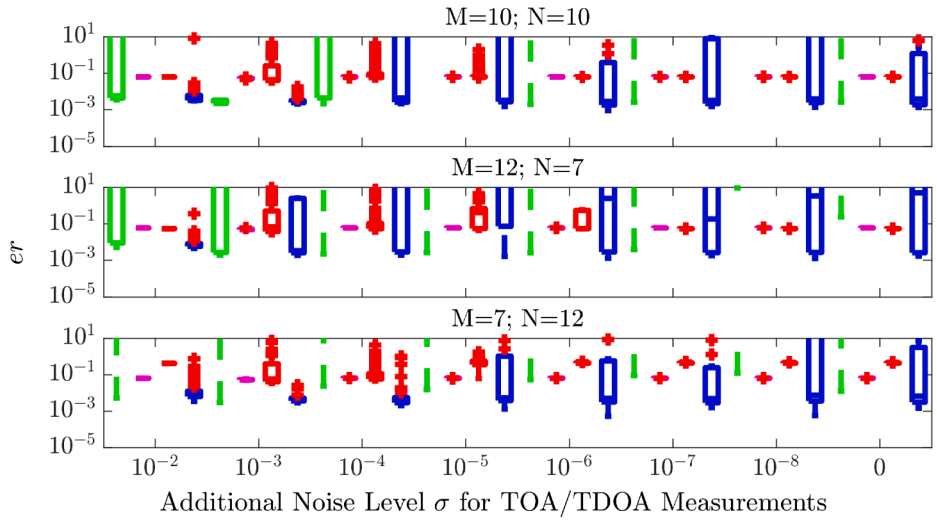


Fig. 7. (a) The performance comparison in terms of recovery rate of each configuration with C3 ($M = N = 10$ (■) and $M = N = 18$(■)), C1 ($M = 15$, $N = 8$ (■) and $M = 20$, $N = 14$ (■)) and C2 ($M = 8$, $N = 15$ (■) and $M = 14$, $N = 20$ (■)); (b)Two different recovery error ranges for UTIm ($er < 10^{-4}s$ and $10^{-4}s \le er < 10^{-2}s$) achieved by STLS and proposed CLRA.



(a) Realistic Simulation



(b) Office Real Data

Fig. 8. Estimation errors for UTIm achieved by STLS, SDP, SDP + LM and proposed CLRA with different additional noises in TOA (or TDOA) measurements.

The differences in these four complexities are:

$$\begin{cases} \hat{J}_{1,4,r} - \hat{J}_{1,r} = 2(M-1)(2(N-1) - M_N) \\ \hat{J}_{1,2,4,r} - \hat{J}_{1,4,r} = (M-1)(N-1-3) \\ \hat{J}_{1,3,4,r} - \hat{J}_{1,4,r} = (N-1)(M-1-3) \end{cases}$$

and

$$\begin{cases} \hat{J}_{1,4,c} - \hat{J}_{1,c} = M_N(2(N-1) - M_N) \\ \hat{J}_{1,2,4,c} - \hat{J}_{1,4,c} = (N-1+3)(N-1-3) \\ \hat{J}_{1,3,4,c} - \hat{J}_{1,4,c} = (M-1+3)(M-1-3) \end{cases}$$

With one additional proposed rank properties (such as LRPV1, LRPV2, or LRPV3), the computational complexity also increases due to the growing size of the Jacobian matrix. Thus, we evaluate the running time of the proposed CLRA methods and compare them with STLS [23,38] across three scenarios (C1, C2, C3). For M microphones and N sources, we use $N_c(M,N) = 50$ configurations and $I_n(M,N) = 100$ initializations per configuration, totaling 5000 implementations.

Fig. 6(a) shows that for M=10, N=10 (C3), CLRA (LRP + LRPV3) is slightly slower than STLS (3–110 ms) but remains below 150 ms per implementation. For M=15, N=8 (C1), Fig. 6(b) indicates that STLS stays fast (3–100 ms), while CLRA variants using LRPV1/LRPV3 run longer (CLRA1 median > CLRA3), and CLRA is below 300 ms. Fig. 6(c) presents the M=8, N=15 (C2) case (LRPV2/LRPV3), with runtimes comparable to (b) and all methods finishing within 900 ms.

Overall, adding LRPV1/LRPV2/LRPV3 increases the Jacobian size relative to STLS (LRP only), but the measured runtimes remain practical (below 150/300/900 ms for C3/C1/C2). This extra cost yields improved accuracy and robustness (Section 7.5); thus, STLS suits resource-limited or coarse estimation, whereas CLRA is preferable under asynchronous and noisy conditions.

7.5. Comparison with other methods and robustness analysis

In this subsection, we present a unified evaluation that includes: (1) a performance comparison between the proposed CLRA method and the auxiliary function-based algorithm [12] (hereafter referred to as Ono) on the simulation dataset; and (2) a robustness analysis in which additive noise is injected into the TOA (or TDOA) measurements on both the realistic simulation dataset and the real dataset, with additional comparisons against SDP and SDP + LM [31]. Specifically, (1) **Realistic Simulation Data:** microphone and source locations were randomly generated

in a $5\,\mathrm{m} \times 5\,\mathrm{m} \times 3\,\mathrm{m}$ room. Audio signals (sampling rate 48 kHz; speed of sound 340 m/s) were generated in MATLAB [10], and TOA (or TDOA) measurements were computed using GCC-PHAT, resulting in an average error of 5×10^{-6} s. (2) **Real Data:** this dataset [43] was collected in a $5\,\mathrm{m} \times 5\,\mathrm{m} \times 3\,\mathrm{m}$ office using 12 fixed microphones (sampling rate 96 kHz) and a loudspeaker, yielding an average TOA (or TDOA) error of 1×10^{-4} s.

7.5.1. Comparison with ono

The experiments are conducted across three different cases, namely C1, C2, and C3. For each configuration, given M microphones and N sources, we set the number of configurations $N_c(M, N) = 50$, and the number of initializations $I_n(M, N) = 50$, resulting in a total of 2500 implementations for each pair of M and N. This experimental design allows for a comprehensive evaluation of the method's performance under random initializations. Additionally, for the Ono algorithm [12], the maximum iteration number of each implementation is set to 2×10^5.

Fig. 7(a) presents the recovery rates obtained from 50 random initializations for each configuration. As shown in the figure, in all three cases (C1: $M = 15$, $N = 8$ and $M = 20$, $N = 14$; C2: $M = 8$, $N = 15$ and $M = 14$, $N = 20$; C3: $M = 10$, $N = 10$ and $M = 18$, $N = 18$), the recovery rate achieved by the Ono algorithm is always zero, indicating its inability to recover the global optima under these conditions. In contrast, both STLS and CLRA show improving recovery rates as M and N increase. Moreover, the recovery rate achieved by STLS is consistently higher than that of Ono.

Notably, for fixed values of M and N, the CLRA method outperforms STLS in terms of minimal, maximal, and median recovery rates. These results underscore the robustness of CLRA, as it consistently achieves higher recovery rates across all trials, even under random initializations. This demonstrates the effectiveness of the proposed CLRA method in overcoming the challenges posed by random initialization, further verifying its potential to achieve global optima.

7.5.2. Robustness analysis

To evaluate the robustness of the proposed CLRA, Gaussian noise with zero mean and standard deviations $\sigma = \{10^{-2}, 10^{-3}, \ldots, 10^{-8}\}$ was added to TOA (or TDOA) measurements. Experimental setups from Section 7.1 were used, with $N_c(M, N) = 50$ configurations and $I_n(M, N) = 50$ initializations, resulting in 2500 implementations for each M, N. Estimation errors (er) were categorized as $er < 10^{-4}s$ (corresponding to $0.034m$ distance errors, acceptable for localization) and $10^{-4}s \leq er < 10^{-2}s$ (3.4m distance errors, significantly affecting localization).

Fig. 7(b) shows that for $\sigma > 10^{-6}$, both STLS and CLRA achieve no results with $er < 10^{-4}s$. For $\sigma \leq 10^{-6}$, CLRA significantly outperforms STLS. For example, when $M = 15$, $N = 8$, CLRA achieves 40% accuracy for $er < 10^{-4}s$, compared to 6% for STLS. Moreover, CLRA consistently achieves better results for $10^{-4}s \leq er < 10^{-2}s$, indicating fewer instances where $er > 10^{-2}s$.

In addition to simulation data, robust analyses were conducted using two additional datasets. For both datasets, $I_n(M, N) = 100$. Fig. 8 shows that due to larger noise ($\sigma > 10^{-6}s$), most $er > 10^{-4}s$, causing significant localization errors. However, CLRA consistently achieves lower errors than STLS, highlighting its potential for real-world applications.

With the added baselines SDP and SDP + LM [31], CLRA remains the best performer overall, yielding the smallest median errors and the tightest dispersion across noise levels. The advantage is most evident at higher θ, where CLRA produces fewer large-error outliers than STLS and SDR-based methods. These results validate CLRA's applicability in domains such as smart homes, where precise synchronization enhances voice-activated systems, and emergency response, aiding in sound source localization under challenging conditions.

While our robust analysis incorporates realistic simulation and real-world datasets, there remain certain limitations in replicating the full complexity of dynamic acoustic environments. Future work will extend this study to include: (1) Validation in dynamic scenarios, such as moving sound sources or microphones. (2) Testing in more diverse environments, such as outdoor or industrial settings, to evaluate the impact of varying reverberation and noise levels. (3) Larger-scale experiments with extended microphone arrays to explore scalability and robustness in more complex networks.

8. Conclusion

In this paper, the main focus is to estimate the UTIm in TOA (or TDOA) measurements, used for synchronizing microphones and sources in ad-hoc WASN. By studying basic LRP of UTIm in TOA (or TDOA), three new variants of LRP were proposed to exploit more low-rank structure rank information of UTIm in TOA (or TDOA), and a proof was given for proposed three variants of LRP. Then, by utilizing the low-rank structure information revealed by proposed three variants of LRP together with LPR for allowing to constrain the UTIm, we proposed combined low-rank approximate method to estimate UTIm. Experimental results showed better performance of proposed combined low-rank approximate method than state-of-the-arts in terms of both the recovery and convergence rate as well as estimation errors, this verified low-rank structure information exploited by proposed three variants of LRP for estimation of UTIm. In addition, the proposed three variants of LRP can also be applied to the other asynchronous applications of wireless sensor network if the audio information is replaced by radio signals.

CRediT authorship contribution statement

Faxian Cao: Conceptualization, Methodology, Software, Writing – original draft; **Yongqiang Cheng:** Resources, Writing – review & editing, Supervision, Funding acquisition, Project administration, Formal analysis; **Adil Mehmood Khan:** Formal analysis, Writing – review & editing, Supervision; **Zhijing Yang:** Supervision, Writing – review & editing, Supervision, Validation; **S. M. Ahsan Kazmi:** Writing – review & editing, Validation; **Zhao Huang:** Writing – review & editing, Validation; **Yingxiu Chang:** Writing – review & editing, Validation.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Faxian Cao reports financial support was provided by China Scholarship Council. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research is supported in part by Natural Science Foundation of Guangdong Province (no. 2025A1515010454) and China Scholarship Council (no. 201908440403).

Supplementary material

Supplementary material associated with this article can be found in the online version at 10.1016/j.dsp.2026.106106.

References

[1] J. Li, S. Chu, Z. Huang, X. Cao, Z. Yao, F. Yang, C. Song, DWSen: dual-path wavelet-attention KAN for joint activity and indoor location sensing, IEEE Trans. Mob. Comput. (2026).

[2] Y. Gu, M. Xu, J. Li, Q. Li, Z. Huang, M. Li, L. Guan, M. Valkama, WiKAN: lightweight Kolmogorov–Arnold networks for accurate indoor WiFi localization, Pervasive Mob. Comput. (2025) 102121.
[3] G. Acar, A.E. Adams, ACMENet: an underwater acoustic sensor network protocol for real-time environmental monitoring in coastal areas, IEE Proc. Radar Sonar Navig. 153 (4) (2006) 365–380.
[4] A. Pastor-Aparicio, J. Segura-Garcia, J. Lopez-Ballester, S. Felici-Castell, M. Garcia-Pineda, J.J. Pérez-Solano, Psychoacoustic annoyance implementation with wireless acoustic sensor networks for monitoring in smart cities, IEEE Internet Things J. 7 (1) (2019) 128–136.
[5] Z. Sheng, S. Pfersich, A. Eldridge, J. Zhou, D. Tian, V.C. Leung, Wireless acoustic sensor networks and edge computing for rapid acoustic monitoring, IEEE-CAA J. Automatic 6 (1) (2019) 64–74.
[6] J. Svatos, J. Holub, Impulse acoustic event detection, classification, and localization system, IEEE Trans. Instrum. Meas. 72 (2023) 1–15.
[7] A. Griffin, A. Alexandridis, D. Pavlidi, Y. Mastorakis, A. Mouchtaris, Localizing multiple audio sources in a wireless acoustic sensor network, Signal Process. 107 (2015) 54–67.
[8] T. Jia, H. Wang, X. Shen, Z. Jiang, K. He, Target localization based on structured total least squares with hybrid TDOA-AOA measurements, Signal Process. 143 (2018) 211–221.
[9] M. Crocco, A. Del Bue, V. Murino, A bilinear approach to the position self-calibration of multiple microphones, IEEE Trans. Signal Process. 60 (2) (2011) 660–673.
[10] M.S. Brandstein, H.F. Silverman, A robust method for speech signal time-delay estimation in reverberant rooms, Proc. IEEE Int. Conf. Acoust. Speech, Signal Process. 1 (1997) 375–378.
[11] P. Pertilä, M. Mieskolainen, M.S. Hämäläinen, Closed-form self-localization of asynchronous microphone arrays, Proc. IEEE HSCM (2011) 139–144.
[12] N. Ono, H. Kohno, N. Ito, S. Sagayama, Blind alignment of asynchronously recorded signals for distributed microphone array, Proc. WASPAA (2009) 161–164.
[13] A. Contini, A. Canclini, F. Antonacci, M. Compagnoni, A. Sarti, S. Tubaro, Self-calibration of microphone arrays from measurement of times of arrival of acoustic signals, Proc. IEEE ISCCSP (2012) 1–6.
[14] I. McCowan, M. Lincoln, I. Himawan, Microphone array shape calibration in diffuse noise fields, IEEE/ACM Trans. Audio, Speech, Language Process. 16 (3) (2008) 666–670.
[15] B. Laufer-Goldshtein, R. Talmon, S. Gannot, A hybrid approach for speaker tracking based on TDOA and data-driven models, IEEE/ACM Trans. Audio Speech Language Process. 26 (4) (2018) 725–773.
[16] D. Bechler, M. Grimm, K. Kroschel, Speaker tracking with a microphone array using Kalman filtering, Adv. Radio Sci. 1 (2003) 113–117.
[17] T.K. Le, N. Ono, Closed-form and near closed-form solutions for TOA-based joint source and microphone localization, IEEE Trans. Signal Process. 64 (18) (2016) 4751–4766.
[18] T.K. Le, K.C. Ho, Algebraic complete solution for joint source and microphone localization using time of flight measurements, IEEE Trans. Signal Process. 68 (2020) 1853–1869.
[19] T.K. Le, N. Ono, Closed-form and near closed-form solutions for TDOA-based joint source and microphone localization, IEEE Trans. Signal Process. 65 (5) (2016) 1207–1221.
[20] S.T. Birchfield, A. Subramanya, Microphone array position calibration by basis-point classical multidimensional scaling, IEEE Trans. Speech Audio Process. 13 (5) (2005) 1025–1034.
[21] C. Evers, P.A. Naylor, Acoustic SLAM, IEEE/ACM Trans. Audio Speech Language Process. 26 (9) (2018) 1484–1498.
[22] M. Crocco, A. Trucco, A. Del Bue, Uncalibrated 3D room geometry estimation from sound impulse responses, J. Franklin Inst. 354 (18) (2017) 8678–8709.
[23] J. Zhang, R.C. Hendriks, R. Heusdens, Structured total least squares based internal delay estimation for distributed microphone auto-localization, Proc. Int. Workshop Acoustic Signal Enhancement (2016) 1–5.
[24] S. Woźniak, K. Kowalczyk, Passive joint localization and synchronization of distributed microphone arrays, IEEE Signal Process. Lett. 26 (2) (2018) 292–296.
[25] D. Hu, Z. Chen, F. Yin, Passive geometry calibration for microphone arrays based on distributed damped Newton optimization, IEEE/ACM Trans. Audio Speech Language Process. 29 (2020) 118–131.
[26] T.K. Le, K.C. Ho, Joint source and microphone localization by angles of arrival, IEEE Trans. Signal Process. 68 (2020) 6521–6534.
[27] R. Heusdens, N. Gaubitch, Time-delay estimation for TOA-based localization of multiple sensors, Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (2014) 609–613.
[28] P.H. Schönemann, On metric multidimensional unfolding, Psychometrika 35 (3) (1970) 349–366.
[29] F. Cao, Y. Cheng, A.M. Khan, Z. Yang, Z. Guo, Are microphone signals alone sufficient for TOA self-positioning?, In 2024 29th International Conference on Automation and Computing (2024) 1–6.
[30] R. Biswas, S. Thrun, A passive approach to microphone network localization, Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (2004) 1544–1549.
[31] D.E. Badawy, V. Larsson, M. Pollefeys, I. Dokmanic, Localizing unsynchronized sensors with unknown sources, IEEE Trans. Signal Process. 71 (2023) 641–654.
[32] V.C. Raykar, V.C. Kozintsev, R. Lienhart, Position calibration of microphones and loudsources in distributed computing platforms, IEEE Trans. Speech Audio Process. 13 (1) (2004) 70–83.
[33] N.D. Gaubitch, W.B. Kleijn, R. Heusdens, Auto-localization in ad-hoc microphone arrays, Proc. IEEE Int. Conf. Acoust. Speech, Signal Process. (2013) 106-110.
[34] N.D. Gaubitch, W.B. Kleijn, R. Heusdens, Calibration of distributed sound acquisition systems using toa measurements from a moving acoustic source, Proc. IEEE Int. Conf. Acoust. Speech, Signal Process. (2014) 7455–7459.
[35] F. Jiang, Y. Kuang, Time delay estimation for tdoa self-calibration using truncated nuclear norm regularization, Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (2013) 3885–3889.
[36] Y. Kuang, K. Åström, Stratified microphone network self-calibration from tdoa measurements, Proc. EUSIPCO (2013) 1–5.
[37] L. Wang, T.K. Hon, J.D. Reiss, A. Cavallaro, Self-localization of ad-hoc arrays using time difference of arrivals, IEEE Trans. Signal Process. 64 (4) (2015) 1018–1033.
[38] R. Heusdens, N. Gaubitch, Time-delay estimation for toa-based localization of multiple microphones, Proc. IEEE Int. Conf. Acoust. Speech, Signal Process. (2014) 609–613.
[39] I. Markovsky, Structured low-rank approximation and its applications, Automatica 44 (4) (2002) 891–909.
[40] J.B. Rosen, H. Park, J. Glick, Total least norm formulation and solution for structured problems, SIAM J. Matrix Anal. Appl. 17 (1) (1996) 110–126.
[41] J. Liu, Z. Wang, J.H. Cui, S. Zhou, B. Yang, A joint time synchronization and localization design for mobile underwater sensor networks, IEEE Trans. Mob. Comput. 15 (3) (2015) 530–543.
[42] J. Nocedal, S. Wright, Numerical Optimization, Springer Science and Business Media (2006).
[43] K. Batstone, G. Flood, T. Beleyur, V. Larsson, H.R. Goerlitz, M. Oskarsson, K. Åström, Robust self-calibration of constant offset time-difference-of-arrival, Proc. IEEE Int. Conf. Acoust. Speech, Signal Process. (2019) 4410–4414.