TOA and TDOA Based Asynchronous Self-Localization: Three Stage Framework for Simultaneous Localization of Microphones and Audio Sources

PhD Thesis in Computer Science

Faxian Cao

Supervisors:
Profs. Yongqiang Cheng and Adil Khan

University of Hull
Cottingham Road
HU6 7RX, Hull, U.K.

w w w. h u l l. a c. u k

December 2024

Declaration

I, Faxian Cao, hereby confirm that the research works presented in this PhD thesis is my original work. All sources and contributions have been appropriately acknowledged. I have duly cited previously published materials where applicable.

I affirm that reasonable care has been taken to ensure the originality of work in this PhD thesis. To the best of my knowledge, this PhD thesis does not violate any UK laws, infringe on third-party copyrights or other intellectual property rights, nor contain any confidential material.

I acknowledge that the University of Hull reserves the right to use plagiarism detection software to verify the originality of the electronic version of this thesis.

I certify that this PhD thesis has not been previously submitted for the award of a degree at any other university.

Signature: Faxian Cao

Date: August 8, 2024

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my two PhD supervisors, Profs. Yongqiang Cheng at the University of Sunderland and Adil Khan at the University of Hull (UoH), for their invaluable supervision and ongoing encouragement throughout my PhD journey at the UoH. Their willingness to share their knowledge and experience, as well as their encouragement to explore challenging research problems and build self-confidence, has been instrumental to my success. Without their support in facilitating my transfer from Queen Mary University of London (QMUL) to the UoH during a difficult period, I might not have been able to continue my scientific research. Their guidance and constant feedback were crucial to my achievement of a PhD degree.

I would also like to express my heartfelt gratitude to my master's supervisor, Prof. Zhijing Yang at Guangdong University of Technology (GDUT). I vividly remember our first meeting in July 2016 at the No. 1 Engineering Building in GDUT, where Prof. Yang introduced me to the field of hyperspectral image classification with machine learning algorithms. This marked the beginning of my scientific research journey. Over the past eight years, from 2016 to 2024, Prof. Yang has continually encouraged me to pursue academic excellence. During challenging times, Prof. Yang discussed the problems with me and helped me find the best solutions. His help will remain in my heart forever.

I would like to extend my profound thanks to my wife, Yiling Jiang. From QMUL to the UoH, she witnessed my arduous PhD journey and stood by me unwaveringly. Her love and support are indescribable.

Finally, I would like to thank my daughter, Xiaoman Cao. Your birth brought immense joy at the end of my PhD journey. Seeing your face reminds me of the most precious happiness in this world. Your father is dedicated to providing you with a happy life, and you will always be the sunshine of your father's life.

Research work during PhD

Journal papers

[J1] F. Cao, Z. Yang, X. Hong, Y. Cheng, Y. Huang and J. Lv, ”Supervised dimensionality reduction of hyperspectral imagery via local and global sparse representation,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3860-3874, 2021.

[J2] Z. Yang, F. Cao (Corresponding Author), Y. Cheng, W. K. Ling and R. Hu, ”Locality regularized robust-PCRC: A novel simultaneous feature extraction and classification framework for hyperspectral images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 12, pp. 8567-8582, 2020.

[J3] X. Zhang, Z. Yang, F. Cao, J. Cao, M. Wang, and N. Cai, ”Conditioning optimization of extreme learning machine by multitask beetle antennae swarm algorithm,” *Memetic Computing*, vol. 12, pp. 151-164, 2020.

[J4] M. Liu, F. Cao, Z. Yang, X. Hong, and Y. Huang, ”Hyperspectral image denoising and classification using multi-scale weighted EMAPs and extreme learning machine,” *Electronics*, vol. 9, no. 12, pp. 2137, 2020.

[J5] F. Cao, Y. Cheng, A. M. Khan, Z. Yang, S. M. Ahsan Kazmi, and Y. Chang, ”Low Rank Properties for Synchronizing Microphones and Sources in Ad-Hoc Wireless Acoustic Sensor Network,” *Under Review*, *IEEE Transactions on Mobile Computing*, 2024.

[J6] F. Cao, Y. Cheng, A. M. Khan, Z. Yang, and S. M. Ahsan Kazmi, ”A New Numerical Method for Relaxing the Minimal Configurations of TOA-Based Joint Sensors and Sources Localization,” *Under Review*, *IEEE Transactions on Signal Processing*, 2024.

[J7] F. Cao, ”Concerns for Self-Localization of Ad-Hoc Arrays Using Time Difference of Arrivals,” Under Review, *IEEE Transactions on Signal Processing*, 2024.

[J8] F. Cao, Y. Cheng, A. M. Khan, Z. Yang, and S. M. Ahsan Kazmi, ”Echo SLAM: Closed Form and Iterative Solutions,” *Target Journal: IEEE Transactions on Signal Processing*.

[J9] F. Cao, Y. Cheng, A. M. Khan, Z. Yang, and S. M. Ahsan Kazmi, ”Can Maximal Coding Rate Reduction and Information Bottleneck Theory Open the Black Box of Deep Learning?,” *Target Journal: IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Conference papers

[C1] F. Cao, Y. Cheng, A. M. Khan, Z. Yang and Z. Guo ”Are Microphone Signals Alone Sufficient for TOA-Based Self-localization?,” Accepted at *29th International Conference on Automation and Computing (ICAC)*, 2024.

[C2] Y. Chang, Y. Cheng, J. Murray, U. Manzoor, M. Khalid and F. Cao, ”Applying and Exploring Supervised Contrasitive Regression for Vision-Based Steering Prediction,” *In 2023 28th International Conference on Automation and Computing (ICAC)*, pp. 1-6, IEEE.

Thesis supervisors Author
Profs. Yongqiang Cheng and Adil Khan **Faxian Cao**

TOA and TDOA Based Asynchronous Self-Localization: Three Stage Framework for Simultaneous Localization of Microphones and Audio Sources

Abstract

Self-localization, a pivotal aspect explored in this research, holds significant relevance across various applications, including human-robot interaction and surveillance for aging individuals. Traditional localization methods relying on GPS signals or visual information face limitations in poorly illuminated environments or areas with obstructed GPS signals. In these situations, audio signals emerge as a promising alternative. When localizing both devices/microphones and ambient objects using audio signals from sources, typically two types of information are used: time of arrival (TOA) and time difference of arrival (TDOA). TOA measures the distance between microphones and sources, while TDOA measures the range difference between pairs of microphones relative to the audio source. However, there are three challenges in localizing both microphones and sources with TOA and TDOA measurements, which limit the efficiency and accuracy of self-localization, regardless of whether the microphones and sources are synchronous or asynchronous.

In scenarios where both microphones and sources are asynchronous, both TOA and TDOA contain unknown timing information (UTIm). The unknown start time for the microphones and the emission time for the sources are embedded in the TOA measurements. Additionally, there is an unknown time offset between pairs of microphones in TDOA measurements. Under this scenario, there are at least two challenges for self-localization. Firstly, TOA requires estimation from both microphone and source signals, whereas TDOA requires estimation from microphone signals only. Even if the UTIm in TOA and TDOA is accurately estimated, asynchronous TOA provides range measurements between microphones and sources, while asynchronous TDOA only provides range differences. Range measurements contain richer and more efficient information than range difference measurements for self-localization, as range differences can be derived directly from range measurements. Therefore, when audio source signals are absent, it is crucial to find a way to use microphone signals alone for efficient self-localization before estimating UTIm. Secondly, UTIm in both TOA and TDOA pose significant challenges for self-localization. Traditional methods for estimating UTIm (synchronizing microphones and sources) in TOA/TDOA measurements often get stuck in local minima due to the randomness of UTIm, leading to inaccuracies in range measurements and substantial localization errors. Therefore, it's paramount to design a method to improve the accuracy of range measurements for self-localization.

The third challenge arises in scenarios where both microphones and sources are synchronized, and range measurements between them are available. Traditional methods require a minimum number of microphones and sources to achieve effective self-localization. Typically, at least six, five, or four microphones are required along with four, five, or six sources, respectively. This requirement is based on the principle that the number of equations (known range measurements) should be greater than or equal to the number of unknowns (location variables for microphones and sources). When the number of microphones and sources is below this minimum threshold, traditional state-of-the-art methods fail. Unfortunately, this issue has not been adequately explored, significantly limiting the efficiency of self-localization. This poses the third challenge to find a way to reduce the number of microphones and sources required for self-localization.

To address above three challenges, this PhD thesis proposes a three stage framework (TSF) designed to simultaneously localize both microphones and audio sources, improving both accuracy and efficiency for self-localization. The initial stage focuses on developing a mapping function that can transform between TOA and TDOA formulas, demonstrating their potential equivalence for the first time. This breakthrough reveals that microphone signals alone are adequate for self-localization, eliminating the need for source signal waveforms and providing richer information for localization once UTIm is estimated in asynchronous TOA/TDOA measurements. This advancement could revolutionize self-localization techniques, greatly expanding their use in challenging environments. Backed by solid mathematical proof and compelling experimental results, this research makes a significant contribution to the current discourse on audio self-localization. In the second stage, an innovative combined low-rank approximation (CLRA) technique aimed at estimating UTIm is introduced. This involves developing three novel low-rank property (LRP) variants, each of which is backed by mathematical proof, allowing UTIm to utilize a broader range of low-rank structural information. By leveraging this augmented low-rank information from both the LRP and the proposed variants, I formulate four linear constraints on UTIm. Employing the CLRA algorithm, global optimal solutions for UTIm based on these constraints are derived. Experimental results showcase proposed method's superior performance over current state-of-the-art approaches, as demonstrated by higher recovery numbers and lower estimation errors for UTIm. In the third stage, the proposed TSF relaxes the minimal configurations required for self-localization by presenting a novel numerical method. Based on the laws of cosine, the localization problem is transformed to estimate four unknown pairs of distances pertaining to one pair of microphones and three pairs of sources. Using the triangle inequality, both the lower and upper boundaries of these four unknown pairs of distances can be obtained, enabling the determination of the numerical method by searching for candidates within the corresponding boundaries. This approach shows that self-localization in 3D space is achievable with only four microphones and four sources, relaxing the minimal configurations required by traditional methods, improving the efficiency for self-localization. Both theory and simulation results validate the feasibility of this new numerical method.

In summary, the impact of the proposed TSF in this PhD thesis extends to providing a comprehensive understanding of self-localization, enhancing accuracy and efficiency in challenging environments. The proposed methodologies contribute to the advancement of signal and audio processing, paving the way for more intelligent and flexible solutions in real-world scenarios.

Table of Contents

List of figures

List of abbreviations

3D: 3 Dimensional

AOA: Angle of Arrival

CLRA: Combined Low-Rank Properties Algorithm

CLRA1: Sub-method of Combined Low-Rank Properties Algorithm

CLRA2: Sub-method of Combined Low-Rank Properties Algorithm

CLRA3: Sub-method of Combined Low-Rank Properties Algorithm

GPS: Global Positional System

GCC-PHAT: Generalized Cross-Correlation with Phase Transform

JMSL: Joint Microphones and Sources Localization

LRP: Low-Rank Property

LRPV1: Variant of Low-Rank Property

LRPV2: Variant of Low-Rank Property

LRPV3: Variant of Low-Rank Property

TOA: Time of Arrival

TDOA: Time Difference of Arrival

SVD: Singular Value Decomposition

STLS: Structure Total Least Square

TSF: Three Stage Gramework

UTIm: Unknown Timing Information

List of symbols

$\mathbf{r}$: location for microphone

$\mathbf{s}$: location for source

c: speed of sound

$\mathbf{t}$: time of arrival

τ: time difference of arrival

$\mathbf{t_p}$: pseudo time of arrival

$\tau_{\mathbf{p}}$: pseudo time difference of arrival

δ: microphone start time

η: source emission time

$\delta_{\mathbf{p}}$: pseudo microphone start time

$\eta_{\mathbf{p}}$: pseudo source emission time

δ': time offset of a pair of microphones

M: number of microphones

N: number of sources

$\mathbf{R}$: location matrix of microphones

$\mathbf{S}$: location matrix of sources

$\mathbf{D}$: time of arrival matrix

$\mathbf{U}$: matrix for time of arrival, microphone start time and source emission time

$\mathbf{D}^*$: matrix for range measurements between microphones and sources

$\mathbf{U}^*$: left singular matrix

$\mathbf{V}^*$: right singular matrix

$\mathbf{A}^*$: singular values matrix

$\mathbf{U_p^*}$: part of left singular matrix

$\mathbf{V_p^*}$: part of right singular matrix and singular values matrix

$\mathbb{R}$: field of real number

$\mathbf{d}$: range between microphone and source

$\mathbf{C}$: solutions for location matrices of microphones and sources

f: mapping function for time of arrival and time difference of arrival

Δf: difference between transformations of time of arrival and time difference of arrival

$\sum$: sum operator

$\|\bullet\|_2$: norm 2 operator

$\|\bullet\|_F$: norm F operator

$\mathbf{T_1^*}$: matrix for proposed variant of low-rank property

$(\mathbf{T_{11}^*}, \mathbf{T_{12}^*})$: sub-matrices for $\mathbf{T_1^*}$

$\mathbf{Z}$: coefficient matrix for LRPV1

$\mathbf{T_2^*}$: matrix for proposed variant of low-rank property

$(\mathbf{T_{21}^*}, \mathbf{T_{22}^*})$: sub-matrices for $\mathbf{T_2^*}$

$\mathbf{W}$: coefficient matrix for LRPV2

$\mathbf{T_3^*}$: matrix for proposed variant of low-rank property

$(\mathbf{T_{31}^*}, \mathbf{T_{32}^*})$: sub-matrices for $\mathbf{T_3^*}$

$\mathbf{Y}$: coefficient matrix for LRPV3

M_N: minimal values between $M-1+3$ and $N-1+3$

$min\{\}$: minimal value operator

$\mathbf{X}$: coefficient matrix for LRP

$(\alpha, \beta, \gamma, \eta)$: parameters for proposed combined low-rank properties algorithm

$\mathbf{v}$: column-wise matrix vectorization

$\mathbf{J}$: Jacobian matrix

Rr: recovery rate

C_r: convergency rate

er: errors for unknown timing information estimation

$(\alpha_{mic}, \beta_s, \gamma_s, \eta_s)$: distances for one pair of microphones and three pairs of sources

Er: principle for selecting numerical solution pertaining to locations of microphones and sources

EM: localization errors for microphones and sources

Chapter 1

Introduction

1.1 Motivation and Objectives

With the development of automation technology and intelligent systems [1], the demand for autonomously localize devices themselves and the ambient objects in an unknown environment is increasing [2]. Since the localization research become more and more power-efficient, robust and intelligent [2], many applications are dominated by this research topic, such as automatic speech recognition [3], fire fighting [4], human robot interaction (HRI) [5], surveillance for aging people [6] and rescue missions [7]. Traditional localization methods [8, 9] often rely on Global Positional System (GPS) or visual sensors, however, those methods may face limitations in certain scenarios [10], such as dark indoor environments, urban canyons, or area with obstructed GPS signals. Therefore, speech self-localization arises much attentions in the past decades. For example, by using the time difference in the sound source from three microphones, *Lee et al.* [11] presented an algorithm for real-time sound localization in human-robot interaction. In addition, *Nakamura et al.* [12] developed an intelligent human tracking system for robots using sound source localization by addressing noise robustness, selective listening, and tracking challenges through advanced GEVD-MUSIC, hierarchical Gaussian mixture models for sound source identification, and audio-visual particle filtering, achieving significant improvements in noise handling, selective listening, and tracking accuracy. Furthermore, *Korayem et al.* [13] presented a low-cost communication system for HRI, integrating a Scout robot and a robotic face with a hidden Markov model-based voice command detection system trained on a non-native Persian speaker database, achieving the improvement in true detection over systems trained on native English data. In this context, the interplay between audio self-localization and object localization becomes crucial, especially when the position of the emitting object is of interest. By integrating insights from both domains, a comprehensive understanding of the environment can be achieved, enhancing the potential for intelligent and adaptable solutions in various real-world scenarios above.

The measurements of time of arrival (TOA) and time difference of arrival (TDOA) [14, 15] are two typical types of information for self-localization, and both of them can be estimated by analyzing the received audio signals from microphones and emitted audio signals from sources [16, 17]. Specifically, if the waveform of the source signal is obtained beforehand, such as the corresponding amplitude, frequency, and duration, TOA between individual microphones and sources can be obtained through generalized cross-correlation with phase transform (GCC-PHAT) [18–20] or the mutual information function [21, 22]. Additionally, compared with TOA, when the source signal's waveform is unknown, the TDOA of a pair of microphones relative to a corresponding source can be determined by performing GCC-PHAT [18–20] or the mutual information function [21, 22] on the signals from the microphones pair [23]. When the recording start times of microphones and the emission times of audio sources are synchronized, the interval between the emission of an audio signal by the source and its reception by a microphone can be measured using TOA. This measurement captures the distance between the microphones and sources. Similarly, TDOA can be used to measure the delay between the reception of corresponding audio signals by a pair of microphones, reflecting the range difference relative to the audio source. TOA inherently includes the direct range between microphones and sources, while TDOA encapsulates the range difference between microphone pairs. Consequently, TDOA can be derived from TOA. Both measurements have been employed in various speech applications, including microphone array calibration [24–26], source localization [27–29], source tracking [30–32], joint microphones and sources localization (JMSL) [33–35], and simultaneous localization and mapping [36, 37]. However, there are three main challenges for localizing both microphones and sources, significantly limiting the efficiency and accuracy for localizing both them.

In practice, microphones and sources are often not synchronized. Human voices or other audio sources emit signals without being aware of the microphones' recording times, resulting in unknown emission times. Additionally, each microphone may start recording at its own independent and unknown time. This asynchrony causes TOA and TDOA to contain unknown timing information (UTIm) [38]. Specifically, TOA includes the unknown start times of microphones and emission times of sources, while TDOA incorporates the unknown time offset between pairs of recordings. When both microphones and audio sources are asynchronous, there are two main challenges for localizing them, which limit the efficiency and accuracy of self-localization. Firstly, even if the

UTIm in TOA/TDOA is accurately estimated, TDOA only provides the range difference between pairs of microphones relative to the corresponding audio source. In contrast, TOA contains the range measurements between microphones and sources, which include richer information since range differences can be directly derived from range measurements. Additionally, once both range measurements and angle of arrival (AOA) measurements [39] are available, the locations of both microphones and sources can be directly determined [40]. However, acquiring TOA requires both microphone and source signals, while TDOA can be obtained using only microphone signals. Therefore, before estimating UTIm, the urgent task is to unify TOA and TDOA to obtain range measurements between microphones and sources using only microphone signals. This unification would improve the efficiency of localizing both microphones and sources, as it allows the use of fewer resources (microphone signals only) to obtain more information (range measurements) for self-localization. Secondly, the randomness of UTIm poses significant challenges for the asynchronous localization of microphones and sources. Traditional methods often get stuck in local minima when estimating UTIm, leading to inaccuracies in synchronization and localization. Therefore, to accurately localize microphones and sources using TOA or TDOA, it is imperative to estimate the global solutions of UTIm contained within these measurements. This involves determining the unknown start and emission times to effectively utilize TOA or TDOA for accurate localization purposes.

Furthermore, even when both microphones and sources are synchronized and range measurements between them are available, the localization task remains challenging. Current state-of-the-art methods for localizing both microphones and sources can be divided into two categories: closed-form solutions [33, 35] and iterative methods [41]. These methods have achieved accurate localization of microphones and sources. However, due to the requirement that the number of equations (known range measurements) must be greater than or equal to the number of unknowns (location variables for microphones and sources), the state-of-the-art research indicates that the minimal number of microphones and sources needed for localization is six/five/four and four/five/six, respectively. These research efforts have aimed to meet this requirement by presenting either closed-form or iterative methods [33, 35, 41]. Unfortunately, these minimal configurations limit the flexibility of self-localization. If the number of microphones and sources is less than the required minimum, the task of self-localization fails. Therefore, an important question arises: is it possible to relax these minimal configuration requirements, thus enabling more flexible configurations for self-localization and enhancing its efficiency in more challenging scenarios?

The goal of this PhD thesis is to enhance the efficiency and accuracy of self-localization by introducing a three-stage framework (TSF) that addresses the three key challenges in localizing both microphones and sources. In the first stage of the proposed TSF, a novel mapping function [38] designed to transform both TOA and TDOA into an equivalent form is introduced. This unification allows TOA to be obtained using only microphone signals, enabling the acquisition of range measurements between individual microphones and sources once the UTIm in the unified TOA/TDOA is estimated. Experimental results validate the effectiveness of this mapping function. By implementing this mapping function, the self-localization process becomes more efficient and adaptable.

In the second stage of the proposed TSF, a combined low-rank approximation (CLRA) algorithm to estimate the solutions for UTIm in TOA/TDOA is introduced, enabling the determination of accurate range measurements between microphones and sources for localizing both. UTIm in TOA/TDOA present significant challenges in various applications. Traditional optimization methods that directly estimate UTIm often perform poorly compared to techniques that leverage low-rank property (LRP) [42]. LRP introduces an additional low-rank structure, enabling the formulation of linear constraints on UTIm and enhancing the development of related low-rank structural information. This approach facilitates the attainment of globally optimal solutions for UTIm, provided the initialization is appropriate. However, the initialization process often relies on randomness, leading to suboptimal local minima. To address this, the second stage of the proposed TSF introduces a novel CLRA method designed to mitigate the effects of random initialization on UTIm. By proposing three new LRP variants, supported by mathematical proof, UTIm can utilize a richer array of low-rank structural information. Leveraging this enhanced low-rank information from both LRP and the proposed variants, four linear constraints on UTIm is established. Using the CLRA algorithm, global optimal solutions for UTIm based on these constraints is derived. Experimental results demonstrate that proposed method outperforms existing state-of-the-art approaches, showing higher recovery rates and lower estimation errors for UTIm. With the proposed CLRA, more accurate distance measurements between microphones and sources can be achieved, significantly enhancing the final task of localizing both microphones and sources.

In the third stage of proposed TSF in this PhD thesis, my research aims to relax the minimal requirements for self-localization in 3D space using range measurements between microphones and sources, enhancing the efficiency of self-localization. Traditionally, state-of-the-art methods indicate that a minimum number of microphones and sources are necessary for localizing both microphones and sources: typically at least four, five, or six microphones and six, five, or four sources, respectively. However, these minimal configurations restrict the applicability of self-localization in challenging scenarios where the available number of microphones and/or sources falls below these requirements.

To address this limitation, I propose a new numerical method that diverges from the principle of requiring a strict balance between the number of valid equations (range measurements) and unknown variables (locations of microphones and sources). This method aims to reduce the dependency on having a large number of microphones and sources by introducing a novel approach to localization. By conceptualizing the self-localization task as involving several triangles and applying the laws of cosine, I transform the problem to derive four unknown distances related to one pair of microphones and three pairs of sources. Then, leveraging triangle inequalities, I establish lower and upper boundaries for these four unknowns based on known range measurements between microphones and sources. This enables us to use a numerical method focused on solving for these four unknowns within their respective boundaries. This approach shows that achieving self-localization in 3D space is feasible with just four microphones and four sources, while maintaining acceptable localization errors. This relaxes the stringent minimum configuration requirements imposed by state-of-the-art methods. Both theoretical analysis and simulation results validate the feasibility and effectiveness of proposed new numerical method in achieving accurate localization outcomes under reduced microphones and sources configurations.
In summary, the objectives of this PhD thesis are summarized as:

- By proposing a mapping function, I demonstrated that the asynchronous TOA and TDOA formulas can be transformed into an identical form. This finding indicates that microphone signals alone are sufficient for TOA-based localization of microphones and sources. Once the UTIm is estimated using asynchronous TOA or TDOA measurements, the range measurements between microphones and sources can be obtained, enhancing the efficiency of localizing both microphones and sources.

- By presenting a CLRA method, three additional LRP variants are introduced. This enables the inclusion of UTIm into a richer LRP pool, allowing more linear constraints to be imposed on UTIm. This approach mitigates the impact of initialization randomness on UTIm, leading to global solutions for UTIm, which are used for synchronizing both microphones and sources. Consequently, this results in more accurate range measurements for localizing both microphones and sources.

- By introducing a new numerical method, the minimum number of microphones and sources required for localizing both is reduced using range measurements. With proposed novel numerical method, the number of microphones and sources can be decreased to four each, relaxing the minimal configuration needed for the localization task and improving the efficiency of localizing both microphones and sources.

1.2 Challenges

Despite decades of research into audio self-localization, numerous challenges persist, particularly in underexplored areas, which hinder the efficiency and accuracy of self-localization. These challenges can be categorized as follows:

- 1. Equivalence between TOA and TDOA

  In scenarios where both microphones and sources operate asynchronously, state-of-the-art methods typically pursue similar approaches to derive range or range difference measurements for self-localization. TOA can be determined when both microphone and source signals are available, facilitating the acquisition of range measurements for localization, contingent upon accurate estimation of the unknown start times of microphones and emission times of sources. Conversely, when only microphone signals are accessible, TDOA can be computed, allowing for the determination of range differences between pairs of microphones relative to the sources, once the unknown time offset between the microphones is estimated. However, range measurements inherently offer richer information for self-localization compared to range difference measurements. Therefore, developing a mapping function that effectively unifies asynchronous TOA and TDOA measurements to render them equivalent is crucial yet remains an unexplored area in current research.

  (Contribution point 1 tackles this challenge)

- 2. Exploring the low-rank structure information between TOA/TDOA and UTIm

  Once TOA and TDOA are unified, exploiting LRP becomes viable for investigating the relationship between UTIm and TOA/TDOA. LRP facilitates the establishment of linear constraints that connect UTIm with TOA/TDOA, enabling the derivation of global solutions for UTIm based on these measurements, assuming proper initialization of UTIm. However, LRP is susceptible to local optima due to the stochastic nature of initialization, leading to inaccuracies in UTIm estimation and significant localization errors. Overcoming this challenge involves identifying additional low-rank structural information that can enhance the diversity of information utilized by UTIm, thereby reducing the impact of initialization randomness.

Despite its potential, this area remains unexplored in current research, presenting a substantial challenge for future investigation.

(Contribution point 2 tackles this challenge)

- 3. Relaxing the minimal configurations for self-localization with numerical solutions

  Over the past few decades, various methods have been developed to localize both microphones and sources using closed-form or iterative solutions, typically requiring a minimum of four/five/six microphones and six/five/four sources, respectively. However, these state-of-the-art methods encounter limitations when the number of microphones and/or sources is insufficient to meet these minimal requirements. Therefore, relaxing these minimal configurations specified in existing literature could enhance the efficiency of self-localization setups, thereby facilitating localization tasks in challenging environments. Nevertheless, the minimal configurations outlined in current literature are based on the fundamental principle that the number of valid equations (range measurements) must be greater than or equal to the number of unknowns (locations of microphones and sources). This principle is inherently sound, continuing to pose significant challenges in this field.

  (Contribution point 3 tackles this challenge)

1.3 Contributions

This PhD thesis makes several key contributions to the field of audio self-localization (Fig. 1.1 presents the flowchart of proposed TSF in this thesis), summarized as follows:

- Unified asynchronous TOA and TDOA measurements

  A mapping function that transforms asynchronous TOA and TDOA measurements into an equivalent form is proposed. This unification allows for the efficient use of microphone signals alone for TOA-based localization, streamlining the process of obtaining range measurements between microphones and sources once the UTIm is estimated.

  This leads to research work [C1].

- Enhanced low-rank properties for UTIm estimation

  By presenting a CLRA method, I introduced three additional LRP variants. These variants enable the UTIm to be integrated into a richer LRP pool, imposing more linear constraints on UTIm. This approach reduces the impact of initialization randomness, leading to more accurate solutions for UTIm and improving the synchronization and localization accuracy of both microphones and sources.

This leads to research works [J5] and [J7].

- Reduced minimal configuration requirements

  A new numerical method that relaxes the minimal configuration requirements for self-localization is developed. Proposed method reduces the number of required microphones and sources to four each, enhancing the efficiency of localization tasks in challenging environments.

  This leads to research works [J6] and [J7].

- Mathematical proofs and theoretical insights

  This thesis includes rigorous mathematical proofs and derivations that underpin the proposed methods. These proofs provide a solid theoretical foundation for the unification of TOA and TDOA measurements, the enhancements in LRP, and the reduction of minimal configuration requirements.

- Wider impact on signal processing and localization

  Beyond the specific contributions to audio self-localization, the techniques and insights developed in this thesis have broader applications in signal processing and localization. The proposed methods can be adapted and applied to other domains with similar localization challenges, such as radio signals, potentially benefiting a wide range of applications in technology and engineering.

By addressing these key challenges and providing robust theoretical and practical solutions, this thesis significantly advances the state-of-the-art in audio self-localization and offers valuable contributions to the wider field of signal processing and localization.

1.4 Organization of Thesis

The organization of this PhD thesis is as follows:
Chapter 1: Introduction
This chapter provides an overview of the motivation behind this PhD thesis, highlighting the challenges addressed. Additionally, it summarizes the contributions made by this research.

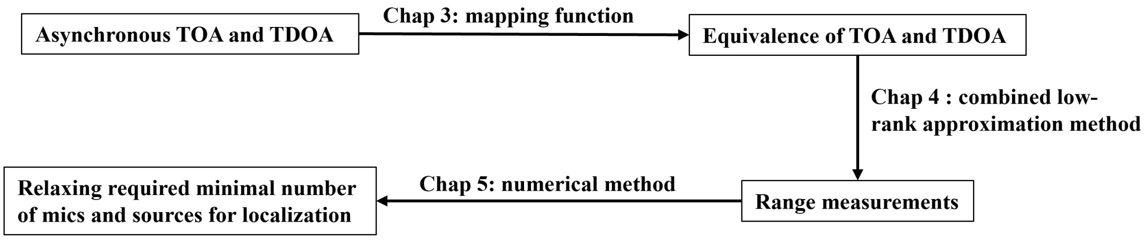


Figure 1.1: The illustration of proposed three stage framework for asynchronous self-localization (TOA: time of arrival; TDOA: time difference of arrival; Chap 3: Chapter 3; Chap 4: Chapter 4; Chap 5: Chapter 5).

Chapter 2: Literature Review and Problem Formulation

This chapter discusses the state-of-the-art literature concerning the synchronization and localization of both microphones and sources. It also introduces the problem formulation and necessary preliminaries for understanding the research context.

Chapter 3: Unified Mapping Function for TOA and TDOA

In this chapter, proposed mapping function designed to unify TOA and TDOA is presented. The methodology, mathematical underpinnings, and experimental results validating the proposed approach are detailed.

Chapter 4: Low-Rank Properties and CLRA Method for Synchronization of Microphones and Sources

This chapter focuses on proposed methods and theories for synchronizing microphones and sources. It includes the introduction of three new variants of LRP, supported by rigorous mathematical proofs, and the CLRA. Experimental results demonstrating the efficacy of the CLRA algorithm using simulation and real-world data are also presented.

Chapter 5: Relaxing Minimal Configurations for Self-Localization

This chapter explores research on relaxing the minimal configurations required for self-localization. It details the transformation of self-localization into estimating four pairs of unknown distances, establishing boundaries for these distances, and presenting numerical solutions for determining the locations of microphones and sources.

Chapter 6: Conclusion and Future Directions

This final chapter summarizes the achievements of this PhD thesis. It also provides insights into potential future research directions stemming from the findings and methodologies developed throughout the study.

Chapter 2

Background

This chapter begins by reviewing the state-of-the-art research on audio self-localization, detailed in Section 2.1. Section 2.2 presents the problem formulation for asynchronous self-localization. Section 2.3 provides an overview of the preliminaries related to LRP for UTIm and localization. Finally, Section 2.4 concludes this chapter by summarizing the key points discussed.

2.1 State-of-the-Arts

In this section, the state-of-the-art approaches for asynchronous audio self-localization are displayed. The methods for self-localization utilize various measurements such as TOA, TDOA, AOA, and energy-based techniques. Additionally, pairwise distance estimation methods have been proposed. For instance, *McCowan et al.* [43] and *Taghizadeh et al.* [44] assume a diffuse noise field, where the microphone pairwise distances are estimated by fitting the measured noise coherence to the noise field, treating localization as a multidimensional scaling problem [45–47]. However, this approach requires microphone synchronization and assumes a diffuse noise field, which may not always hold true. Another pairwise distance estimation method assumes minimum and maximum TDOA values from sources located at end-fire directions, calculating inter-device distances accordingly [48–51]. This method bypasses the need for unknown time offsets for self-localization but relies on the assumption of end-fire sources, which is not universally applicable. Some methods utilize energy information for self-localization [52–54], eliminating the need to obtain UTIm. However, these methods often assume that certain microphone-source pairs are co-located.

The AOA is another crucial measurement used for self-localization. *Le et al.* [39] introduced a rank property specific to AOA, which enables the recovery of both microphone and source locations.

Several methods have been proposed that utilize both TDOA and AOA measurements, leveraging techniques such as the least squares method [55, 56]. Unlike other methods, AOA measurements do not require the acquisition of UTIm. However, obtaining accurate AOA measurements poses several challenges. These challenges include dealing with multipath propagation, environmental noise, array design considerations, real-time processing requirements, frequency dependence, and the complexity of phase information.

Self-localization using TOA or TDOA information can be categorized into unsynchronized and synchronized scenarios based on whether microphones and sources share a common clock. In scenarios where both TOA and TDOA measurements lack a common clock (unsynchronized), methods for estimating UTIm and performing self-localization can be categorized into two main groups: optimization-based and LRP-based methods. Optimization-based approaches typically employ maximum likelihood estimation to estimate locations and source emission times. For instance, *Biswas et al.* [57] proposed a joint estimation method using TOA information, albeit without microphones' start times. *Raykar et al.* [58] derived closed-form solutions assuming co-located pairs using TOA or TDOA information with maximum likelihood estimation for locations and start times. *Ono et al.* [59] introduced an auxiliary function-based algorithm for TDOA-based estimation, showing improved convergence properties but is sensitive to initialization randomness, impacting the stability and optimality of microphone and source locations. *Badawy et al.* [41] proposed a method for UTIm estimation and self-localization using TOA information. They introduced a novel loss function that directly computes the locations of microphones and sources. Initially, this method eliminates UTIm in TOA/TDOA using a Gram matrix, followed by minimizing the corresponding loss function to determine microphone and source locations. However, the approach is prone to instability because the Gram matrix substitutes UTIm in TOA/TDOA with functions related to microphone and source locations, potentially introducing ambiguity in their determination. *Hu et al.* [56] utilized least squares with TDOA and AOA for UTIm estimation using the Hessian matrix and local gradients. *Wozniak et al.* [55] applied a Moore-Penrose pseudoinverse method combining TDOA and AOA, albeit requiring both measurements, which is not always feasible for self-localization. LRP-based methods leverage low-rank information to formulate linear constraints from UTIm information, categorized into alternating minimization [60, 61], nuclear truncation minimization [62, 63], and structure total least square (STLS) [64, 65]. Compared to optimization-based methods, LRP-based approaches demonstrate superior performance in UTIm estimation and self-localization by exploiting UTIm's low-rank structure. However, methods relying solely on LRP may encounter local optima, which can lead to inaccuracies in the range measurements between microphones and sources.

In synchronized scenarios, various methods have been developed. Localization methods in the current literature can be broadly categorized into two main groups: closed (or near-closed) form solutions [33,35] and iterative methods [66,67]. For closed-form solutions, when synchronized TDOA measurements and the positions of microphones are known, localization of sources can be achieved using methods such as spherical interpolation [68, 69], hyperbolic intersection [70, 71], or linear intersection techniques [72]. Applications assuming co-location of microphones and sources utilize multidimensional scaling methods [45, 58] based on the distances between microphones/sources. Extensive reviews and mathematical properties of these localization methods are documented in [73], with detailed references provided in [74] and [75]. *Crocco et al.* [23] propose a closed-form solution for JMSL using LRP [42] and synchronized TOA measurements for a pair of co-located microphone and source. Additionally, *Le et al.* [33] extend this approach by employing LRP and linear methods to solve polynomial equations [76,77], providing closed-form and near-closed solutions for JMSL when there are seven or more microphones and four or more sources with synchronized TOA measurements. With synchronized TDOA measurements, *Le et al.* [40] first derive synchronized TOA measurements from TDOA measurements using LRP. They then present closed-form and near-closed solutions for JMSL under conditions where there are seven or more microphones and five or more sources, excluding specific cases [35]: 1) seven microphones and five sources, 2) seven microphones and six sources, and 3) eight microphones and five sources. *Le et al.* [34] also propose an algebraic complete solution for self-localization with synchronized TOA measurements by transforming the problem into finding an upper triangular linear transformation matrix. Moreover, *Pollefeys et al.* [78] propose a method for JMSL using rank-5 factorization with synchronized TDOA measurements, applicable when there are ten microphones and five sources. In addition to closed-form solutions, some studies employ gradient descent or auxiliary function methods for nonlinear least-squares optimization to obtain iterative solutions for JMSL using synchronized TOA [41, 61, 66] or TDOA measurements [43, 53, 59, 63, 67].

By investigating rigid bipartite graphs, specifically synchronized TOA measurements for JMSL, *Bolker et al.* [79] demonstrate that the minimal configurations for JMSL should consist of at least three microphones and three sources when both microphones and sources are positioned in 2D space (refer to Theorems 11 and 14 in [79]; note that any three positions among microphones and sources must not lie on a single line). Additionally, for scenarios where both microphones and sources are in 3D space, Theorems 11 and 12 in [79] stipulate that the minimal configurations for localizing microphones and sources require six/five/four microphones and four/five/six sources (note that any three/four positions among microphones and sources should not lie on a single line/plane). The rigid mathematical properties defined in [79] establish these minimal configurations/cases as the consensus for JMSL. This consensus indicates that the number of valid equations (synchronized TOA or range measurements between microphones and sources) should at least equal the number of unknown variables (locations of microphones and sources). Numerous methodologies in the state-of-the-art aim to achieve these minimal configurations/cases for JMSL without imposing constraints on the geometric arrangements of microphones and sources. For instance, researchers at Lund University have proposed closed-form solutions using methods such as the Grobner basis method [80] and employing the Macaulay2 software to solve polynomial equations [81] (referencing works like [33, 82, 83]). Notably, *Kuang et al.* [82] have developed closed-form solutions for scenarios where both microphones and sources are located in three dimensions. *Burgess et al.* [83,84] have investigated closed-form solutions when the span of microphone/source locations has a higher dimension than the corresponding span of source/microphone locations. Moreover, *Zhayida et al.* [85] have examined cases involving four/five microphones and five/four sources, where two distances are provided between any pair of microphones and any pair of sources. However, these methods typically require a minimum number of microphones and sources—-four/five/six for microphones and six/five/four for sources—-limiting the efficiency in self-localization configurations when the required number of microphones and audio sources is not available.

2.2 Problems Formulation

In this part, the main focuses are about the problem formulation of this PhD thesis. First, the definition of TOA and TDOA will be displayed, then three challenges pertaining to the transformation of TOA and TDOA, synchronization of microphones and sources, localization of microphones and sources will be displayed.

In a 3D space, assuming there are M microphones and N sources located at unknown positions within a room. These positions are represented by $\begin{bmatrix}\mathbf{r}_1, & \mathbf{r}_2, & \cdots, & \mathbf{r}_M\end{bmatrix}_{3\times M}$ for the microphones and $\begin{bmatrix}\mathbf{s}_1, & \mathbf{s}_2, & \cdots, & \mathbf{s}_N\end{bmatrix}_{3\times N}$ for the sources, where $\mathbf{r}_i = \begin{bmatrix}\mathbf{r}_{1,i}, & \mathbf{r}_{2,i}, & \mathbf{r}_{3,i}\end{bmatrix}^T$ and $\mathbf{s}_j = \begin{bmatrix}\mathbf{s}_{1,j}, & \mathbf{s}_{2,j}, & \mathbf{s}_{3,j}\end{bmatrix}^T$ (i and j range from 1 to M and 1 to N, respectively). If there is a fundamental control center that synchronizes both the microphones and sources, it is evident that both the sources emission time and microphones start time can be known. Under these conditions, TOA between i^{th} microphone and j^{th} source can be determined using formula $\mathbf{t}_{i,j} = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|_2}{c}$ (c: speed of sound; $\|\bullet\|$: L2 norm). Additionally, the TDOA formula for a pair of microphones relative to the j^{th} audio source is displayed

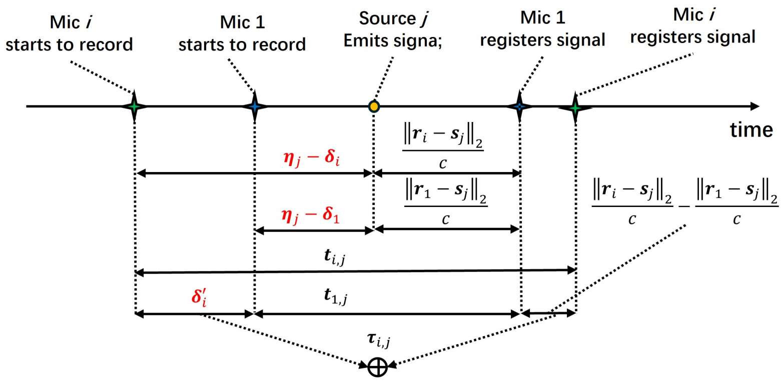


Figure 2.1: The illustration for time of arrival formulation in Eq. (2.1) and time difference of arrival formulation in Eq. (2.2) with j^{th} source and a pair of microphones ($\mathbf{r}_1$: location of 1^{st} mic; $\mathbf{r}_i$: location of i^{th} mic; $\mathbf{s}_j$: location of j^{th} source; η_j: emission time of j^{th} source; δ_1: start time of 1^{st} mic; δ_i: start time of i^{th} mic; δ_i': time offset between 1^{st} mic and i^{th} mic; $\mathbf{t}_{i,j}$: time of arrival between i^{th} microphone and j^{th} source; $\mathbf{t}_{1,j}$: time of arrival between 1^{st} microphone and j^{th} source; $\tau_{i,j}$: time difference of arrival between 1^{st} microphone and i^{th} microphone with respect to j^{th} source; c: speed of sound; $\| \bullet \|_2$: L2 norm).

as $\tau_{i,j} = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|_2 - \|\mathbf{r}_1 - \mathbf{s}_j\|_2}{c}$ using received microphone signals. Using the synchronous TDOA formula $\tau_{i,j}$, it is evident that this formula is determined by the signals from the 1^{st} and i^{th} microphones signals. In general, TDOA is defined as the time difference between any two microphones relative to the j^{th} source. However, for convenience, this thesis defines it specifically as the time difference between the 1^{st} and i^{th} microphones relative to j^{th} source.

However, in most practical scenarios, the microphones and sources are asynchronous, as noted in references [38, 64, 65], resulting in unknown emission times for the sources and unknown start times for the microphones. In such asynchronous settings, even the waveform of the source signal is acquired beforehand, due to the UTIm pertaining to microphones start time and sources emission time, the TOA measurements between individual microphones and sources remains incomplete, where those UTIm are represented as $\delta = \left[\delta_1, \delta_2, \cdots, \delta_M\right]^T$ and $\eta = \left[\eta_1, \eta_2, \cdots, \eta_N\right]^T$ (δ_i: i^{th} microphone start time; η_j: j^{th} source emission time). Thereby, the formula of asynchronous TOA measurements is expressed as [38, 41, 64]:

$$\mathbf{t}_{i,j} = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|_2}{c} + \eta_j - \delta_i. \tag{2.1}$$

Given the invariance of the geometry of microphones and sources to rotation, translation, and reflection, the locations of the first and second microphones and the first source can be defined as $\mathbf{r}_1 = \begin{bmatrix}0, & 0, & 0\end{bmatrix}^T$, $\mathbf{r}_2 = \begin{bmatrix}0, & 0, & \mathbf{r}_{3,2}\end{bmatrix}^T$ and $\mathbf{s}_1 = \begin{bmatrix}0, & \mathbf{s}_{2,1}, & \mathbf{s}_{3,1}\end{bmatrix}^T$, respectively, where $\mathbf{r}_{3,2} > 0$ and $\mathbf{s}_{2,1} > 0$. In addition, the emission time for the first source can be defined as zero, i.e. $\eta_1 = 0$ [60].

In the context of asynchronous microphones and sources, when the waveform of the source signal is unknown, TDOA can be utilized. The GCC-PHAT method [18] can be employed to estimate TDOA from the received signals at the microphones. Thereby, by denoting the time offset δ_i' pertaining to 1^{st} and i^{th} microphones' start time, the asynchronous TDOA formula is expressed as [38,41,64]:

$$\tau_{i,j} = \mathbf{t}_{i,j} - \mathbf{t}_{1,j} = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|_2 - \|\mathbf{r}_1 - \mathbf{s}_j\|_2}{c} + \delta_1 - \delta_i = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|_2}{c} - \frac{\|\mathbf{r}_1 - \mathbf{s}_j\|_2}{c} + \delta_i', \tag{2.2}$$

where Fig. 2.1 illustrates the signal model pertaining to the TOA formula (see Eq. (2.1)) and TDOA formula (see Eq. (2.2)).

Upon examining Eq. (2.2), it becomes evident the audio signal emitted from j^{th} source can be received by both 1^{st} and i^{th} microphones. Therefore, the unknown i^{th} microphone start time (δ_i), j^{th} source emission time and the range/distance measurements between i^{th} microphone and j^{th} source are included in the i^{th} microphone signal. Moreover, the unknown 1^{st} microphone start time (δ_1), j^{th} source emission time and the range/distance measurements between 1^{st} microphone and j^{th} source are included in the 1^{th} microphone signal. Thereby, the estimation of TDOA ($\tau_{i,j}$) can be conducted by GCC-PHAT [18] method with the audio signals from both 1^{st} and i^{th} microphones, resulting in the statement that the estimation of TDOA measurement is independent from audio signal from sources. Furthermore, TDOA is defined as the time difference between two microphones relative to the corresponding source signal. Therefore, it is clear that the TDOA for the j^{th} source in Eq. (2.2) can also be obtained using the i^{th} microphone signal along with any other remaining microphone signal.

By defining $\eta_{\mathbf{t}j} = -\frac{\|\mathbf{r}_1 - \mathbf{s}_j\|_2}{c}$ and $\delta_{\mathbf{t}i} = -\delta_i'$, TDOA formula in Eq. (2.2) can be rewritten as [41,64]

$$\tau_{i,j} = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|_2}{c} + \eta_{\mathbf{s}j} - \delta_{\mathbf{t}i}. \tag{2.3}$$

Next, the three problems addressed in this PhD thesis are outlined as follows:

- **Problem 1:** It can be observed that TDOA formula in Eq. (2.3) exhibits a structural similarity to the TOA formula in Eq. (2.1). Unfortunately, the precise relationship pertaining to TOA and TDOA formulas remains ambiguous (see Eqs. (2.1) and (2.2)), as existing literature has not yet clarified this connection. This lack of clarity extends to whether using microphone signals alone are sufficient for TOA-based self-localization or not. On the other hand, Once the UTIm is estimated using Eq. (2.1) and Eq. (2.2), range measurements are obtainable via Eq. (2.1), whereas Eq. (2.2) provides only range differences. The research aim is to assess the feasibility of relying solely on microphone signals for self-localization when source signal waveforms are unavailable. Additionally, this PhD thesis seeks to determine the range measurements between individual microphones and sources post-UTIm estimation in TOA/TDOA. This investigation could potentially challenge the prevailing notion that acquiring source signal waveforms is indispensable for TOA-based self-localization, making the task of self-localization more efficient, potentially broadening the applicability of self-localization techniques in challenging environments.

- **Problem 2:** By integrating the TOA and TDOA formulas in Eqs. (2.1) and (2.2), the low-rank structure information pertaining to the UTIm and TOA/TDOA can be exploited for establishing a linear constraint for UTIm. However, relying solely on LRP can cause the solutions for UTIm to become trapped in local minima, resulting in incorrect range measurements between microphones and sources. Consequently, my goal is to estimate the UTIm from TOA/TDOA measurements by incorporating additional linear constraints. This method enables UTIm to tap into a more comprehensive LRP pool, thereby improving the accuracy of range measurements between microphones and sources and enhancing the localization of microphones and sources.

- **Problem 3:** Once both microphones and audio sources are synchronous, the next step is to localize both microphones and sources using the range/distance measurements between them. From Eq. (2.1), it is evident that to localize all microphones and sources, the number of equations MN (the number of range measurements) must be at least equal to the number of unknown variables $3(M+N)$. By accounting for the invariance related to translation, rotation, and reflection in the geometry of microphones and sources, state-of-the-art research over the past decades has established that the following inequality must be satisfied for successful localization of both microphones and sources [33]:

$$MN \geq 3(M+N) - \frac{d(d+1)}{2}, \tag{2.4}$$

where $d = 3$ denotes 3D space and $\frac{d(d+1)}{2} = 6$ is the invariance pertaining to the translation, rotation and reflection for the geometry of microphones and sources. Therefore, by operating some simple derivations, it is obvious that Eq. (2.4) can be rewritten as [33]

$$(M-3)(N-3) \geq 3. \tag{2.5}$$

By examining Eq. (2.5), it can be observed that given any three or four positions among microphones and sources, as long as they do not lie on the same line or plane, the minimum configurations for the number of microphones and sources are: 1) $M = 4$ and $N = 6$, 2) $M = 5$ and $N = 5$ and 3) $M = 6$ and $N = 4$. This indicates that the total number of microphones and sources must be at least ten. Otherwise, it is impossible to localize both microphones and sources. State-of-the-arts during the past decades try to approach this theoretical minimal configurations by presenting either closed-form solutions or iterative methods.

However, when there are not sufficient number of microphones and sources for conducting the task of self-localization, resulting in the impossibility for localizing both microphones and sources, so that the traditional methods in state-of-the-arts are useless for task of self-localization. Therefore, my goal is not only to localize both microphones and sources but also to reduce the number of devices/microphones and/or audio sources required for self-localization, thereby relaxing the minimal configurations suggested by state-of-the-art methods. Specifically, this thesis aim to demonstrate that the locations of both microphones and sources can be determined even when the total number of microphones and sources is less than ten. This approach will make the localization process more efficient, requiring fewer devices and/or targets, thus enhancing the flexibility of self-localization configurations. Additionally, it will improve the utility of microphones in challenging environments and extend the capabilities of self-localization techniques.

2.3 Preliminaries

This part shows the preliminaries of LRP pertaining to the tasks of both synchronization and localization.
By squaring both sides in Eq. (2.1), it can be derived that:

$$\frac{\mathbf{r}_i^T \mathbf{r}_i + \mathbf{s}_j^T \mathbf{s}_j - 2\mathbf{r}_i^T \mathbf{s}_j}{c^2} = \mathbf{t}_{i,j}^2 + \eta_j^2 + \delta_i^2 - 2(\mathbf{t}_{i,j}\eta_j - \mathbf{t}_{i,j}\delta_i + \eta_j \delta_i), \tag{2.6}$$

where i and j range from 1 to M and 1 to N, respectively.
To develop the LRP, this thesis starts by subtracting the equation for $i = 1$ and $j = 1$ sequentially

from Eq. (2.6), and then add the equation for $i = j = 1$. Given $\eta_1 = 0$, this results in [64, 65]:

$$\frac{-2(\mathbf{r}_i - \mathbf{r}_1)^T(\mathbf{s}_j - \mathbf{s}_1)}{c^2}$$
$$= \mathbf{t}_{i,j}^2 - \mathbf{t}_{i,1}^2 - \mathbf{t}_{1,j}^2 + \mathbf{t}_{1,1}^2 + 2\delta_i(\mathbf{t}_{i,j} - \mathbf{t}_{i,1}) - 2\delta_1(\mathbf{t}_{1,j} - \mathbf{t}_{1,1}) - 2\eta_j(\mathbf{t}_{i,j} - \mathbf{t}_{1,j}) + 2\eta_j(\delta_1 - \delta_i), \quad (2.7)$$

for $i = 2, \cdots, M$ and $j = 2, \cdots, N$.
Then 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 $\mathbf{U} \in \mathbb{R}^{(M-1)\times(N-1)}$ for Eq. (2.7), where

$$\begin{aligned}
\mathbf{R}_{:,i-1} &= \mathbf{r}_i - \mathbf{r}_1, \\
\mathbf{S}_{:,j-1} &= \mathbf{s}_j - \mathbf{s}_1, \\
\mathbf{D}_{i-1,j-1} &= \mathbf{t}_{i,j}^2 - \mathbf{t}_{i,1}^2 - \mathbf{t}_{1,j}^2 + \mathbf{t}_{1,1}^2, \\
\mathbf{U}_{i-1,j-1} &= 2\delta_i(\mathbf{t}_{i,j} - \mathbf{t}_{i,1}) - 2\delta_1(\mathbf{t}_{1,j} - \mathbf{t}_{1,1}) \\
&\quad - 2\eta_j(\mathbf{t}_{i,j} - \mathbf{t}_{1,j}) + 2\eta_j(\delta_1 - \delta_i),
\end{aligned}$$

for $i = 2, \cdots, M$ and $j = 2, \cdots, N$, Eq. (2.7) can be expressed in matrix form as:

$$\frac{-2\mathbf{R}^T\mathbf{S}}{c^2} = \mathbf{D} + \mathbf{U}. \quad (2.8)$$

In Eq. (2.8), the left-hand side represents information related to the unknown positions of the microphones and sources, while the right-hand side contains information pertaining to TOA/TDOA and UTIm. This indicates that determining the UTIm is a prerequisite for estimating the locations of the microphones and sources. Next, the LRP [38, 42, 64, 65, 86] for both UTIm and localizing microphones and sources is outlined.

LRP for UTIm estimation: if

$$\begin{cases} M - 1 > 3 \\ N - 1 > 3 \end{cases}, \quad (2.9)$$

LRP can be stated as

$$rank(\mathbf{D} + \mathbf{U}) = rank(\mathbf{R}^T\mathbf{S}) \leq 3. \quad (2.10)$$

Upon inspecting Eq. (2.10), it is evident that UTIm and TOA/TDOA are included in matrix $\mathbf{U}$, while TOA/TDOA is encapsulated in matrix $\mathbf{D}$, thereby illustrating the inclusion of low-rank structure information between UTIm and known TOA/TDOA in Eq. (2.10). However, initializing UTIm randomly often leads to solutions becoming trapped in local minima, presenting several challenges. Firstly, with an insufficient number of microphones or sources—specifically fewer than seven microphones or six sources—the convergence rate of LRP, defined as the ratio of successfully recovered configurations to all configurations, approaches zero percent. Secondly, the recovery rate, which denotes the ratio of successful initializations to all initializations within a configuration, remains limited regardless of the number of microphones and sources. Thirdly, the estimation error of UTIm increases significantly in the presence of noise in TOA/TDOA measurements. Hence, my primary emphasis is to delve into further low-rank structural information pertaining to UTIm and TOA/TDOA. Through the establishment of supplementary linear constraints grounded on these observations, my aim is to augment the recovery and convergency rates of UTIm, while alleviating estimation inaccuracies within noisy environments.

LRP for localization: Upon obtaining UTIm, the range measurements between microphones and sources is acquired. Alternatively, when both microphones and source signals are available, if there is a fundamental control device for synchronizing the microphones and sources, we can also obtain the range measurements between them. This allows us to proceed to the next step of deriving the locations of both the microphones and the sources. From Eq. (2.8), we have [33],

$$-2\mathbf{R}^T\mathbf{S} = c^2(\mathbf{D}+\mathbf{U}) = \mathbf{D}^*, \tag{2.11}$$

where ${\mathbf{D}^*}_{i-1,j-1} = c^2(\mathbf{D}_{i-1,j-1}+\mathbf{U}_{i-1,j-1}) = \mathbf{d}_{i,j}^2 - \mathbf{d}_{i,1}^2 - \mathbf{d}_{1,j}^2 + \mathbf{d}_{1,1}^2$, $\mathbf{d}_{i,j}$ is the range measurement between i^{th} microphone and j^{th} sources, with i and j ranging from 2 to M and 2 to N, respectively. Then, applying SVD to matrix $\mathbf{D}^*$ in Eq. (2.11) yields [33],

$$\mathbf{D}^* = \mathbf{U}^*\mathbf{A}^*{\mathbf{V}^*}^T, \tag{2.12}$$

where $\mathbf{U}^* \in \mathbb{R}^{(M-1)\times(M-1)}$ and ${\mathbf{V}^*}^T \in \mathbb{R}^{(N-1)\times(N-1)}$ are the left and right singular matrices, respectively, and $\mathbf{A}^* \in \mathbb{R}^{(M-1)\times(N-1)}$ represents the corresponding singular values. With LRP asserting $rank(\mathbf{D}^*) \leq 3$ in Eq. (2.11), defining $\mathbf{U}_\mathbf{p}^* = {\mathbf{U}^*}_{:,1:3} \in \mathbb{R}^{(M-1)\times 3}$, $\mathbf{V}_\mathbf{p}^* = {\mathbf{A}^*}_{1:3,1:3}{\mathbf{V}^*}^T_{:,1:3} \in \mathbb{R}^{3\times(N-1)}$, Eq. (2.12) can be rewritten as [33]

$$\mathbf{D}^* = \mathbf{U}_\mathbf{p}^*\mathbf{V}_\mathbf{p}^*. \tag{2.13}$$

Finally, utilizing Eqs. (2.11) and (2.13), by defining an unknown matrix $\mathbf{C} \in \mathbb{R}^{3\times 3}$, it can be derived [33]:

$$\begin{cases} \mathbf{R}^T = \mathbf{U}_\mathbf{p}^* \mathbf{C}^{-1} \\ -2\mathbf{S} = \mathbf{C}\mathbf{V}_\mathbf{p}^* \end{cases}. \tag{2.14}$$

Upon inspection of Eq. (2.14), it is obvious that there are nine unknown variables in matrix $\mathbf{C}$. Furthermore, considering $\mathbf{R}_{:,i-1} = \mathbf{r}_i - \mathbf{r}_1$, $\mathbf{S}_{:,j-1} = \mathbf{s}_j - \mathbf{s}_1$, $\mathbf{r}_1 = \begin{bmatrix} 0, & 0, & 0 \end{bmatrix}^T$ and $\mathbf{s}_1 = \begin{bmatrix} 0, & \mathbf{s}_{2,1}, & \mathbf{s}_{3,1} \end{bmatrix}^T$, introduces two additional unknown variables in $\mathbf{s}_1$. Therefore, once the unknown eleven variables in both matrix $\mathbf{C}$ and vector $\mathbf{s}_1$ are confirmed, the locations of all microphones and sources can be achieved.

Over the decades, current state-of-the-arts methods typically require at least four/five/six microphones and six/five/four sources, respectively, to localize both microphones and sources. This configuration has been widely accepted as the minimal configurations, streaming from the principle that the number of equations (range measurements) should equal or exceed the number of unknowns (variables related to microphone and source locations). However, exploring the possibility of reducing the number of microphones and sources for self-localization has been overlooked. If these minimal requirements by reducing the number of required microphones and sources can be relaxed, the process of self-localization could be significantly enhanced. This reduction not only streamlines self-localization tasks but also allows for more flexible configurations, potentially revolutionizing self-localization techniques in various challenging environments.

2.4 Summary

In this chapter, this thesis first summarizes the state-of-the-arts for audio self-localization, including pairwise distance estimation methods, energy-based method, TOA, TDOA and AOA based methods. Then the three problems that we shall investigate in this thesis are formulated, including transformations of asynchronous TOA and TDOA, synchronization of microphones and sources, localization of microphones and source. Finally, the preliminaries are displayed, including LRP for synchronization and localization.

My contributions to the TOA and TDOA based asynchronous self-localization are described in the following chapters.

Chapter 3

Transformation for Asynchronous TOA and TDOA Measurements

3.1 Introduction

Accurately localizing both distributed microphones and sound sources is critical in various acoustic applications, including signal enhancement, and separation as well as noise reduction, [14, 15, 64]. Traditionally, TOA and TDOA measurements are employed for this purpose [23]. However, UTIm is embedded within TOA and TDOA measurements [86], where TOA contains information about the both unknown microphone start time and source emission time, while TDOA reflects the time offsets between pairs of microphones relative to audio source, presenting challenges for asynchronous TOA and TDOA based self-localization.

When source signal waveforms are available are accessible, providing details pertaining to frequency, amplitude, and duration, TOA measurements can be estimated using methods like the GCC-PHAT [87]. This has led to the evolution of various self-localization methodologies, including maximal likelihood estimation [58], probabilistic generative models [88], Gram matrix and semi-definite relaxation [89], and approaches utilizing the LRP [42] with STLS [65] and alternating minimization method [61, 86]. Conversely, in scenarios where source signal waveforms are challenging to obtain, self-localization methodologies focus on TDOA measurements. These can be estimated using audio signals from microphone pairs [87], prompting the development of techniques such as such as auxiliary function methods [62], maximum likelihood estimation [55, 58], LRP with nuclear truncation minimization [62, 90], and distributed damped Newton optimization [56, 91].

However, amid these advancements, a crucial question arises: Can microphone signals alone be adequate for TOA-based self-localization, thereby eliminating the requirement for source signals and providing more comprehensive information? Specifically, after estimating UTIm in TOA/TDOA measurements, the range measurements between individual microphones and sources for self-localization can be derived by using TOA measurements, whereas TDOA measurements offer range differences between microphone pairs relative to corresponding sources. However, range measurements inherently provide more comprehensive information than range differences. For instance, combining range measurements with AOA measurements [39] allows for direct localization of both microphones and sources [40]. Alternatively, utilizing range measurements alone enables closed-form solutions for JMSL [33]. On the contrary, TDOA measurements rely solely on microphone signals, while TOA measurements require signals from both microphones and sources. Therefore, proving that TOA and TDOA can be equivalent with only microphone signals is crucial. Addressing this question carries profound implications for the field, potentially streamlining and enhancing the efficiency of self-localization processes. Furthermore, this represents a timely advancement, particularly in light of the growing complexity of audio environments and the need for flexible localization methods.
This study marks a pioneering effort to address a fundamental question in localization. I introduce a novel mapping function that equates TOA and TDOA formulas, demonstrating their perfect correspondence. This finding highlights that microphone signals alone are sufficient for TOA-based self-localization tasks. Proposed innovative approach not only elucidates the relationship between TOA and TDOA but also challenges the conventional notion that TOA requires both microphone signals and source signal waveforms. This breakthrough simplifies self-localization by eliminating the need for additional source signal information. Furthermore, proposed approach broadens the scope of properties typically linked to TOA-based localization, such as rank 3 [33] and rank 5 [78], making them applicable to TDOA-based methods. This advancement has the potential to redefine self-localization techniques in asynchronous environments and drive further advancements in signal and audio processing.

3.2 Proposed Mapping Function

When the waveform of the source signals is missing, TOA measurements become unavailable, leaving only TDOA measurements for self-localization of both microphones and sources. Previous works have not explored the relationships between asynchronous TOA and TDOA measurements. To address this gap, a novel general form of mapping function to demonstrate that using microphone signals alone is sufficient is introduced. In detail, we propose a novel mapping function to transform both TOA and TDOA formulas into alternative forms. By introducing new variables to represent the sub-components of the transformed TOA and TDOA formulas, we demonstrate that all sub-variables in the transformed formulas are identical. This establishes a clear relationship between the TOA and TDOA formulas, providing evidence that microphone signals alone are sufficient for TOA-based self-localization. The proposed general form of the mapping function, $f(\bullet)$, for both TOA and TDOA formulas (see Eqs. (2.1) and (2.2)), is defined as:

$$f(\mathbf{t}_{i,j}) = \mathbf{t}_{i,j} - \mathbf{t}_{i,1} - \frac{\sum_{i=1}^{M}(\mathbf{t}_{i,j} - \mathbf{t}_{i,1})}{M}, \tag{3.1}$$

and

$$f(\tau_{i,j}) = \tau_{i,j} - \tau_{i,1} - \frac{\sum_{i=1}^{M}(\tau_{i,j} - \tau_{i,1})}{M}. \tag{3.2}$$

By applying the mapping function, $f(\bullet)$, to the TOA and TDOA formulas in Eqs. (2.1) and (2.2), and introducing the following variables:

$$\begin{cases} \delta_{\mathbf{p}_i} = \frac{\|\mathbf{r}_i - \mathbf{s}_1\|_2}{c} \\ \eta_{\mathbf{p}_j} = \frac{\sum_{i=1}^{M}(\|\mathbf{r}_i - \mathbf{s}_1\|_2 - \|\mathbf{r}_i - \mathbf{s}_j\|_2)}{cM} \end{cases}, \tag{3.3}$$

the relationship is formulated as:

$$f(\mathbf{t}_{i,j}) = f(\tau_{i,j}) = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|_2}{c} - \delta_{\mathbf{p}_i} + \eta_{\mathbf{p}_j}, \tag{3.4}$$

where i and j range from 1 to M and 1 to N, respectively. From Eq. (3.4), it is clear that this relationship mirrors the structure of the TOA formula in Eq. (2.1). This implies that the locations of both microphones and sources can be determined using $f(\bullet)$ in a manner similar to TOA-based self-localization methods. Crucially, this finding challenges the long-held assumption that TOA requires both microphone signals and the waveform of source signals by demonstrating the sufficiency of using microphone signals alone. Furthermore, once the UTIm variables $\delta_{\mathbf{p}_i}$ and $\eta_{\mathbf{p}_j}$ are estimated, the range measurements between microphones and sources can be directly obtained. This innovation simplifies the self-localization process, making it more flexible and efficient, and enhancing its effectiveness in complex environments.

3.2.1 Proof for mapping function $f(\bullet)$

By developing the transformation for the TOA formula, as described in Eq. (3.1). Following that, the transformation for the TDOA formula is derived, as shown in Eq. (3.2). Ultimately, the accuracy of Eq. (3.4) is verified by comparing the transformed TOA formula in Eq. (3.1) with the transformed TDOA formula in Eq. (3.2).

Mapping function for TOA formula

Starting with the TOA formula given in Eq. (2.1):

$$\mathbf{t}_{i,1} = \frac{\|\mathbf{r}_i - \mathbf{s}_1\|_2}{c} + \eta_1 - \delta_i, \tag{3.5}$$

then using Eqs. (2.1) and (3.5), the difference between $\mathbf{t}_{i,j}$ and $\mathbf{t}_{i,1}$ is expressed as:

$$\mathbf{t}_{i,j} - \mathbf{t}_{i,1} = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|_2}{c} - \frac{\|\mathbf{r}_i - \mathbf{s}_1\|_2}{c} + \eta_j - \eta_1. \tag{3.6}$$

From Eq. (3.6), the average value of $\mathbf{t}_{i,j} - \mathbf{t}_{i,1}$ over index i can be calculated as:

$$\frac{\sum_{i=1}^{M}(\mathbf{t}_{i,j} - \mathbf{t}_{i,1})}{M} = \frac{\sum_{i=1}^{M}(\|\mathbf{r}_i - \mathbf{s}_j\|_2 - \|\mathbf{r}_i - \mathbf{s}_1\|_2)}{cM} + \eta_j - \eta_1. \tag{3.7}$$

Finally, using Eqs. (3.6) and (3.7), the mapping function for the TOA formula, $f(\mathbf{t}_{i,j})$, as defined in Eq. (3.1), is:

$$f(\mathbf{t}_{i,j}) = \mathbf{t}_{i,j} - \mathbf{t}_{i,1} - \frac{\sum_{i=1}^{M}(\mathbf{t}_{i,j} - \mathbf{t}_{i,1})}{M} = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|_2}{c} - \frac{\|\mathbf{r}_i - \mathbf{s}_1\|_2}{c} + \frac{\sum_{i=1}^{M}(\|\mathbf{r}_i - \mathbf{s}_1\|_2 - \|\mathbf{r}_i - \mathbf{s}_j\|_2)}{cM}, \tag{3.8}$$

where i and j range from 1 to M and 1 to N, respectively.

Mapping function for TDOA formula

Using TDOA formula from Eq. (2.2), the difference between $\tau_{i,j}$ and $\tau_{i,1}$ is derived as follows:

$$\tau_{i,j} - \tau_{i,1} = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|_2}{c} - \frac{\|\mathbf{r}_1 - \mathbf{s}_j\|_2}{c} - \frac{\|\mathbf{r}_i - \mathbf{s}_1\|_2}{c} + \frac{\|\mathbf{r}_1 - \mathbf{s}_1\|_2}{c}. \tag{3.9}$$

From Eq. (3.9), it can be observed that the mean value of $\tau_{i,j} - \tau_{i,1}$ over index i is displayed as

$$\frac{\sum_{i=1}^{M}(\tau_{i,j} - \tau_{i,1})}{M} = \frac{\sum_{i=1}^{M}(\|\mathbf{r}_i - \mathbf{s}_j\|_2 - \|\mathbf{r}_i - \mathbf{s}_1\|_2)}{cM} - \frac{\|\mathbf{r}_1 - \mathbf{s}_j\|_2}{c} + \frac{\|\mathbf{r}_1 - \mathbf{s}_1\|_2}{c}. \tag{3.10}$$

Finally, by combining Eqs. (3.9) and (3.10), the mapping function for the transformation of the TDOA formula, $f(\tau_{i,j})$, is derived as follows:

$$f(\tau_{i,j}) = \tau_{i,j} - \tau_{i,1} - \frac{\sum_{i=1}^{M}(\tau_{i,j} - \tau_{i,1})}{M} = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|_2}{c} - \frac{\|\mathbf{r}_i - \mathbf{s}_1\|_2}{c} + \sum_{i=1}^{M} \frac{(\|\mathbf{r}_i - \mathbf{s}_1\|_2 - \|\mathbf{r}_i - \mathbf{s}_j\|_2)}{cM} \tag{3.11}$$

where i and j range from 1 to M and 1 to N, respectively.

Validation of equivalence

Building on the definitions of $\delta_{\mathbf{p}_i}$ and $\eta_{\mathbf{p}_j}$ in Eq. (3.3), along with the transformations of the TOA and TDOA formulas from Eqs. (3.8) and (3.11), the equivalence of these equations can be confirmed. This equivalence verifies the mapping function $f(\bullet)$ as defined in Eq. (3.4).

With this verification in Eq. (3.4), it becomes clear that the transformations of the TOA and TDOA formulas are indeed equivalent, which proves that microphone signals alone are sufficient for both TOA and TDOA-based self-localization. This finding challenges the traditional notion that TOA requires both microphone signals and the waveform of source signals for localization. Furthermore, once the UTIm $\delta_{\mathbf{p}_i}$ and $\eta_{\mathbf{p}_j}$ are estimated as shown in Eq. (3.4), range measurements between microphones and sources can be directly determined using only the microphone signals. Additionally, the equivalence demonstrated in Eq. (3.4) suggests that properties typically associated with TOA-based localization, such as rank 3 [33] and rank 5 [78], can also be applied to TDOA-based localization. This makes the process of self-localization more efficient and adaptable. By eliminating the need for source signal information, this approach facilitates a range of applications, including noise reduction, source signal enhancement, and separation [14, 15, 64], all of which depend on effective self-localization.

3.2.2 Impact of noise on mapping function $f(\bullet)$

In this part, the impact of noise in TOA and TDOA on the output of proposed mapping function is analyzed. Let the noise intensities in TOA of Eq. (2.1) and TDOA of Eq. (2.2) be denoted as $\sigma_{\mathbf{t}i,j}$ and $\sigma_{\mathbf{td}i,j}$, respectively, where i and j range from 1 to M and j to N, respectively. Based on these assumptions, the TOA formula in Eq. (2.1) and TDOA formula in Eq. (2.2) can be reformulated as follows:

$$\begin{cases} \mathbf{t}_{i,j} = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|_2}{c} + \eta_j - \delta_i + \sigma_{\mathbf{t}i,j} \\ \tau_{i,j} == \frac{\|\mathbf{r}_i - \mathbf{s}_j\|_2}{c} - \frac{\|\mathbf{r}_1 - \mathbf{s}_j\|_2}{c} + \delta_i' + \sigma_{\mathbf{td}i,j} \end{cases} . \tag{3.12}$$

Upon inspection of Eq. (3.12), by applying the proposed mapping function to TOA and TDOA formulas in Eq. (3.12), it implies:

$$\begin{cases} f(\mathbf{t}_{i,j}) = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|_2}{c} - \delta_{\mathbf{p}_i} + \eta_{\mathbf{p}_j} + \sigma_{\mathbf{t}i,j} - \sigma_{\mathbf{t}i,1} - \frac{\sum_{i=1}^{M}(\sigma_{\mathbf{t}i,j} - \sigma_{\mathbf{t}i,1})}{M} \\ f(\tau_{i,j}) = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|_2}{c} - \delta_{\mathbf{p}_i} + \eta_{\mathbf{p}_j} + +\sigma_{\mathbf{td}i,j} - \sigma_{\mathbf{td}i,1} - \frac{\sum_{i=1}^{M}(\sigma_{\mathbf{td}i,j} - \sigma_{\mathbf{td}i,1})}{M} \end{cases} . \tag{3.13}$$

Based on the mapping function defined in Eqs. (3.1) and (3.2), Eq. (3.13) can be reformulated as:

$$\begin{cases} f(\mathbf{t}_{i,j}) = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|_2}{c} - \delta_{\mathbf{p}_i} + \eta_{\mathbf{p}_j} + f(\sigma_{\mathbf{t}i,j}) \\ f(\tau_{i,j}) = \frac{\|\mathbf{r}_i - \mathbf{s}_j\|_2}{c} - \delta_{\mathbf{p}_i} + \eta_{\mathbf{p}_j} + +f(\sigma_{\mathbf{td}i,j}) \end{cases} \tag{3.14}$$

Comparing Eqs. (3.4) with (3.14), it is evident that the proposed mapping function also transforms the noise into the same form as the TOA and TDOA in Eqs. (3.1) and (3.2). Thus, the output of the proposed mapping function is determined by the noise intensity $\sigma_{\mathbf{t}i,j}$ in TOA and $\sigma_{\mathbf{td}i,j}$ in TDOA.

3.3 Experimental Validation

The experimental results are presented to validate the proposed mapping function for showing the equivalence of asynchronous TOA and TDOA formulas. First, the experimental setups are described in Section 3.3.1. Next, Section 3.3.2 defines the evaluation metric and demonstrates the validations of the proposed mapping function.

3.3.1 Setups

Simulation data

The simulation data is generated randomly in MATLAB using a uniform distribution. The start time for the microphones and the emission time for the sources are within the range $[-1,\ 1]s$. The positions of the microphones and sources are distributed within a room measuring $10 \times 10 \times 3\ m^3$ [41] and the speed of sound is set to 340 m/s. Both the number of microphones (M) and the number of sources (N) are set to 20, and 1000 different configurations are simulated. This results in a total of

400,000 data points for the simulation. In more detail, by applying the ground truth for the locations of microphones and sources, as well as the microphones' start times and the sources' emission times, to the TOA and TDOA formulas in Eqs. (2.1) and (2.2), we can obtain the corresponding values of asynchronous TOA and TDOA. Once the proposed mapping function is applied to these asynchronous TOA and TDOA values, their transformations become identical, thereby verifying the effectiveness of the proposed mapping function.

Real-Life data

The real-life data [92] was gathered from an office space measuring 5×3 square meters, from which most furniture was removed. In this setup, 12 fixed microphones recorded a chirp emitted from 65 different positions by a loudspeaker. The resulting real-life dataset comprises a 12×65 synchronous TOA matrix with an unknown constant time value, available for download on GitHub[1] [41,92], with the TDOA matrix calculated using Eq. (2.2). For further details on this dataset, readers are referred to references [41,92]. Additionally, both the start times of the microphones and the emission times of the sources fall within the range of $[-1,\ 1]s$ seconds. The real-life dataset comprises a total of 780 data points, corresponding to the 12 microphones and 65 sources used in the experiment. Fig. 3.1 shows the corresponding positions of 12 microphones.

3.3.2 Evaluations and Results

First, I present the values of the proposed mapping function applied to both TOA and TDOA measurements from simulation and real-life datasets. Subsequently, the validity of the proposed mapping function defined in Eq. (3.4) is evaluated by assessing the difference between the transformations of TOA and TDOA formulas:

$$\Delta f_{i,j} = f(\mathbf{t}_{i,j}) - f(\tau_{i,j}), \tag{3.15}$$

where i and j range from 1 to M and 1 to N, respectively. If $\Delta f_{i,j} = 0$ as indicated in Eq. (3.15), it confirms that the transformations of both TOA and TDOA formulas are equivalent, validating the proposed mapping function $f(\bullet)$.

In Fig. 3.2, both simulated and real datasets are analyzed to present the results. As shown in Figs. 3.2(a) and (c), the values of the proposed mapping function $f(\bullet)$ for TOA and TDOA measurements in the simulated data fall within the range of $[-0.05,\ 0.05]$ seconds, while in the real data, the range

[1] The real-life data can be accessed at https://github.com/swing-research/xtdoa/tree/master/matlab

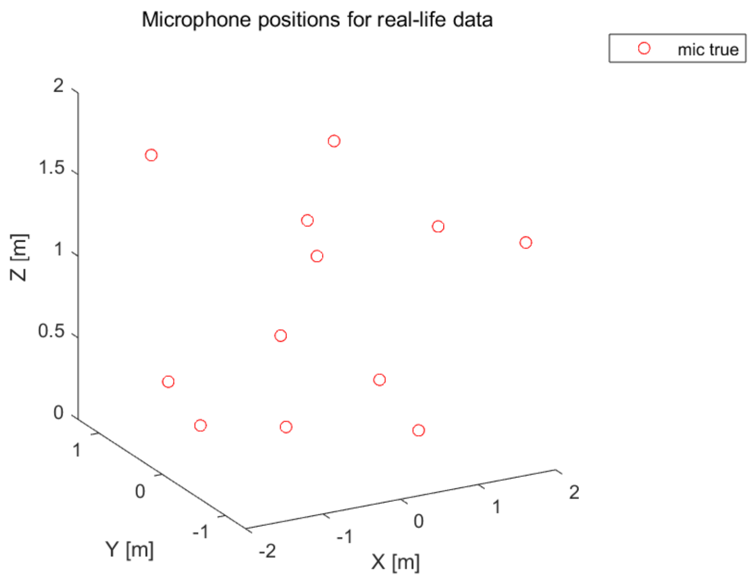
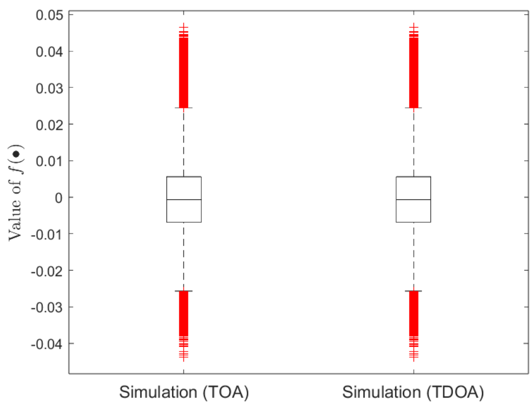


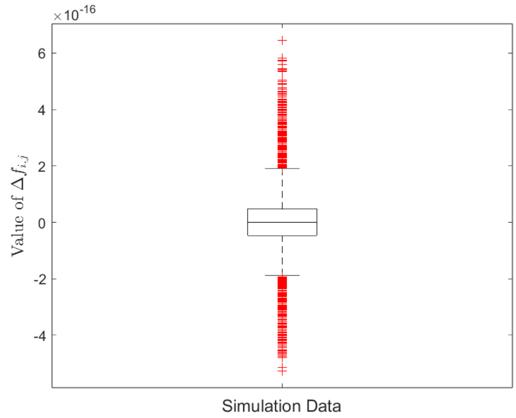
Figure 3.1: The illustration for 12 microphones' positions of real-life dataset [92].

is $[-0.02,\ 0.02]$ seconds. This variation is likely due to the different sizes of the rooms used in the simulation and real-life experiments. Additionally, Figs. 3.2(b) and (d) shows that $\Delta f_{i,j}$ has a magnitude of 10^{-16} seconds, influenced by MATLAB's computational precision. Hence, the condition $\Delta f_{i,j} = 0$ in Eq. (3.15) is satisfied, confirming that the transformations of the TOA formula in Eq. (2.1) and the TDOA formula in Eq. (2.2) are equivalent using the proposed mapping function $f(\bullet)$, thereby validating the statement in Eq. (3.4). Furthermore, Figs. 3.2(b) and (d) show that the discrepancies between the simulation and real-life dataset results are minimal, with $\Delta f_{i,j}$ being nearly zero. This minimal discrepancy is due to the application of the proposed mapping function, which transforms both asynchronous TOA and TDOA formulas. As a result, the transformed TOA and TDOA values are almost identical, as illustrated in Figs. 3.2(a) and (c).

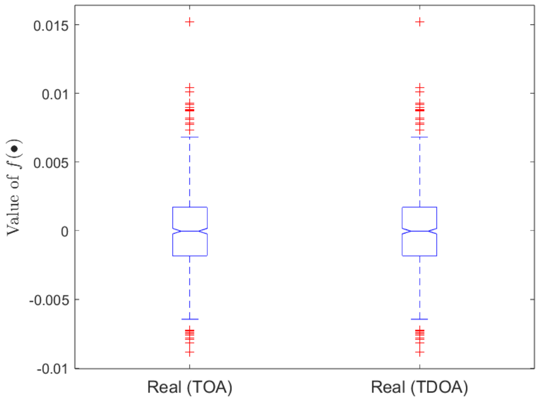
Given that real-life dataset contains noise, our proposed mapping function ensures that the transformation of both asynchronous TOA and TDOA measurements is identical, as indicated in Figs. 3.2(b) and (d). This can be explained as follows: when audio sources emit signals and microphones receive them, the noise present in the microphones is fixed. Consequently, the asynchronous TOA and TDOA measurements include this microphone noise. By applying our mapping function, the noise in the TOA and TDOA measurements is transformed in the same way, resulting in a $\Delta f_{i,j}$ that is nearly zero, this fits the theory results in Eq. (3.14). Additionally, the simulation dataset is generated from 20 microphones and 20 sources, whereas the real-life dataset comprises 12 microphones and



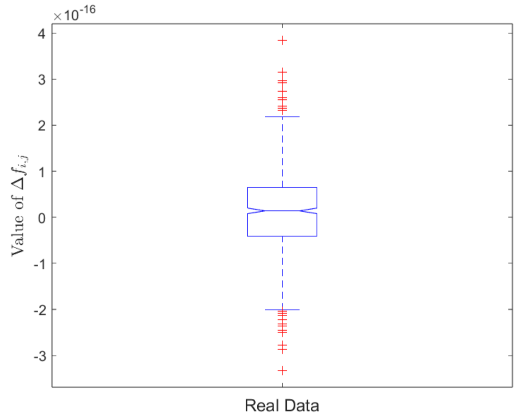
(a) Transformations of both TOA and TDOA measurements with proposed mapping function



(b) Assessment of $\Delta f_{i,j}$



(c) Transformations of both TOA and TDOA measurements with proposed mapping function



(d) Assessment of $\Delta f_{i,j}$

Figure 3.2: Validation of the proposed mapping function using both simulation and real-life datasets includes the following aspects. Left: Evaluation of the proposed mapping function $f(\bullet)$ for transformations of both TOA and TDOA measurements using Eqs. (3.1) and (3.2), respectively (up: simulation datasets; down: real-life dataset). Right: Assessment of $\Delta f_{i,j}$ using Eq. (3.15) (up: simulation datasets; down: real-life dataset)

65 sources. Despite these differences, the $\Delta f_{i,j}$ results from both datasets are nearly zero, further validating the effectiveness of the proposed mapping function.

TOA measurements depend on both microphone and source signals, whereas TDOA measurements rely solely on microphone signals. Figs. 3.2(b) and (d) show that the transformations of TOA and TDOA measurements are consistent, proving that using only microphone signals suffices for self-localization. This approach allows for the direct estimation of range measurements between microphones and sources once the UTIm in TOA/TDOA is identified, thus removing the necessity for source signals in the self-localization process. This discovery marks a significant advancement in the field of self-localization.

3.4 Summary

This chapter explores the feasibility of using microphone signals exclusively for self-localization, an aspect not previously investigated in existing literature. In situations where both the source signal emission time and the microphones' recording start times are unknown, a novel mapping function is introduced. This function transforms both TOA and TDOA formulas, demonstrating their equivalence for the first time. This shows that utilizing microphone signals alone is sufficient for self-localization. Moreover, once the UTIm in TOA/TDOA is estimated, direct range measurements between microphones and sources become feasible. This capability enhances the flexibility and efficiency of self-localization tasks. Therefore, the proposed mapping function represents a timely advancement in the field of self-localization.

Building on this work, the next chapter will address another key challenge in asynchronous self-localization and present the method I proposed to tackle it: estimating the start times of the microphones and the emission times of the sources.

Chapter 4

Low-Rank Matrices for Synchronization of Microphones and Audio Sources

4.1 Introduction

Building on my previous work in Chapter 3, which demonstrated the sufficiency of using microphone signals alone for localization in asynchronous environments, the relationship between asynchronous TOA and TDOA measurements is investigated. A novel mapping function that aligns the transformations of TOA and TDOA measurements is investigated, indicating that the low-rank structure information of UTIm exploited in TOA estimation could similarly benefit TDOA estimation (see Section 4.2). This insight suggests that once UTIm in TOA or TDOA is estimated, range measurements between microphones and sources can be obtained.

In this study, the CLRA method for UTIm estimation in TOA/TDOA is proposed. This method integrates linear constraints derived from LRP with three new variants of LRP designed to enhance low-rank structure exploitation between UTIm and TOA/TDOA. By leveraging these constraints to constrain UTIm, CLRA seeks a global optimization solution, thereby improving both convergence and recovery rates for UTIm as well estimation accuracy in noisy environments. Additionally, rigorous mathematical proofs for the proposed LRP variants are represented, solidifying the theoretical basis of proposed approach. In summary, this chapter not only introduces a novel method for UTIm estimation but also provides robust evidence of its effectiveness. This advancement promises to enhance the accuracy of UTIm estimation in TOA/TDOA, marking a significant contribution to signal processing and localization research.

4.2 Proposed New Low-Rank Properties

This section introduces three proposed variants of LRP aimed at estimating UTIm. LRP leverages the low-rank structure between UTIm and TOA/TDOA, and these variants enhance this structure, with the additional structured low-rank information from proposed variants of LRP, we can establish more linear constraints between UTIm and TOA/TDOA. These constraints confine the solutions of UTIm to a more precise solution space, thereby increasing the likelihood of achieving global optimal solutions for UTIm, thereby improving both recovery and convergence rates for UTIm estimation and reducing estimation errors in noisy environments.

Before introducing the three proposed variants of LRP, it is essential to extend the mapping function introduced in Chapter 3 to LRP, thereby leveraging the low-rank structure information between UTIm and TOA/TDOA. Referring to Eq. (3.4), where $\mathbf{t_p}_{i,j} = f(\mathbf{t}_{i,j}) = f(\tau_{i,j})$ represents TOA/TDOA measurements, $\delta_{\mathbf{p}_i}$ denotes the pseudo start time of the i^{th} microphone, and $\eta_{\mathbf{p}\,j}$ denotes the pseudo emission time of the j^{th} source. It is evident that LRP, as defined in Eq. (2.10), is applicable to asynchronous TOA/TDOA measurements. Henceforth in this PhD thesis, unless explicitly stated otherwise, $\mathbf{t_p}_{i,j}$ will denote asynchronous TOA/TDOA measurements, $\delta_{\mathbf{p}_i}$ will represent the start time of the i^{th} microphone, and $\eta_{\mathbf{p}\,j}$ will signify the emission time of the j^{th} source.

Introducing three new matrices

$$\mathbf{T}_{\mathbf{1}}^{*} = \begin{bmatrix} \mathbf{D} & \mathbf{U} \end{bmatrix} \in R^{(M-1)\times 2(N-1)}, \tag{4.1}$$

$$\mathbf{T}_{\mathbf{2}}^{*} = \begin{bmatrix} \mathbf{D}^{T} & \mathbf{U}^{T} \end{bmatrix} \in R^{(N-1)\times 2(M-1)}, \tag{4.2}$$

and

$$\mathbf{T}_{\mathbf{3}}^{*} = \begin{bmatrix} \mathbf{D} & \mathbf{U} \\ \mathbf{U} & \mathbf{D} \end{bmatrix} \in \mathbb{R}^{2(M-1)\times 2(N-1)}, \tag{4.3}$$

three variants of LRP are proposed under the condition that Eq. (2.9) holds.

LRPV1:

$$rank\,(\mathbf{T}_{\mathbf{1}}^{*}) \leq min\{M-1, N-1+3\}, \tag{4.4}$$

where matrix $\mathbf{T}_{\mathbf{1}}^{*}$ hold low-rank property when $M-1 > N-1+3$ (see proof in Section 4.3.1).

LRPV2:

$$rank\,(\mathbf{T}_{\mathbf{2}}^{*}) \leq min\{N-1, M-1+3\}, \tag{4.5}$$

where matrix $\mathbf{T_2^*}$ holds low-rank property when $N-1>M-1+3$ (see proof in Section 4.3.2).
LRPV3:

$$rank\left(\mathbf{T_3^*}\right) \leq min\{N-1+3, M-1+3\}, \tag{4.6}$$

where matrix $\mathbf{T_3^*}$ holds low-rank property when Eq. (2.9) holds (see proof in Section 4.3.3).
From Eqs. (4.1), (4.2), (4.3), (4.4), (4.5) and (4.6), it is evident that LRPV1, LRPV2, and LRPV3 consistently demonstrate low-rank structure information pertaining to known TOA/TDOA and UTIm, provided that $M-1>N-1+3$ holds for $\mathbf{T_1^*}$, and $N-1>M-1+3$ holds for $\mathbf{T_2^*}$.

4.3 Proof for Proposed Variants of LRP

The proof for the three variants of LRP relies on two main components:
1) The foundational LRP framework established in existing literature [42, 64, 65], encapsulated by Eq. (2.10) in Section 2.3, which asserts $rank(\mathbf{D}+\mathbf{U}) \leq 3$.
2) Principles from linear algebra [93–95]. This involves defining three matrices: $\mathbf{E} \in \mathbb{R}^{m\times n}$, $\Theta \in \mathbb{R}^{n\times h}$, and $\mathbf{O} \in \mathbb{R}^{m\times h}$, alongside two column vectors $\theta \in \mathbb{R}^n$ and $\mathbf{o} \in \mathbb{R}^m$. In accordance with established linear algebra theory [93–95], it follows:
Theorem 1: For a given linear system $\mathbf{E}\theta = \mathbf{o}$, where E is the coefficient matrix, and the augmented matrix $\begin{bmatrix}\mathbf{E} & \mathbf{o}\end{bmatrix} \in \mathbb{R}^{m\times(n+1)}$ with unknown θ, the following conditions are necessary and sufficient when $m \geq n$:

- If $rank(\mathbf{E}) = rank(\begin{bmatrix}\mathbf{E} & \mathbf{o}\end{bmatrix}) = n$, then $\mathbf{E}\theta = \mathbf{o}$ has a unique solution, and vice versa.
- If $rank(\mathbf{E}) = rank(\begin{bmatrix}\mathbf{E} & \mathbf{o}\end{bmatrix}) < n$, then $\mathbf{E}\theta = \mathbf{o}$ has multiple solutions, and vice versa.

Applying *Theorem 1*, a single linear system can be extended to multiple systems by substituting the column vectors θ and $\mathbf{o}$ with matrices Θ and $\mathbf{O}$, respectively. The conditions outlined in *Theorem 1* remain sufficient and necessary [95].

4.3.1 Proof for Proposed LRPV1

This part presents the proof for LRPV1 as defined in Eq. (4.4). The analysis proceeds by examining the conditions under which $rank\left(\mathbf{T_1^*}\right) < N-1+3$ or $rank\left(\mathbf{T_1^*}\right) = N-1+3$ in subsequent part.
The LRP framework established in existing literature [42, 64, 65] states that $rank(\mathbf{D}+\mathbf{U}) \leq 3$, indicating that three column vectors from matrix $D+U$ can represent the remaining column vectors [64].

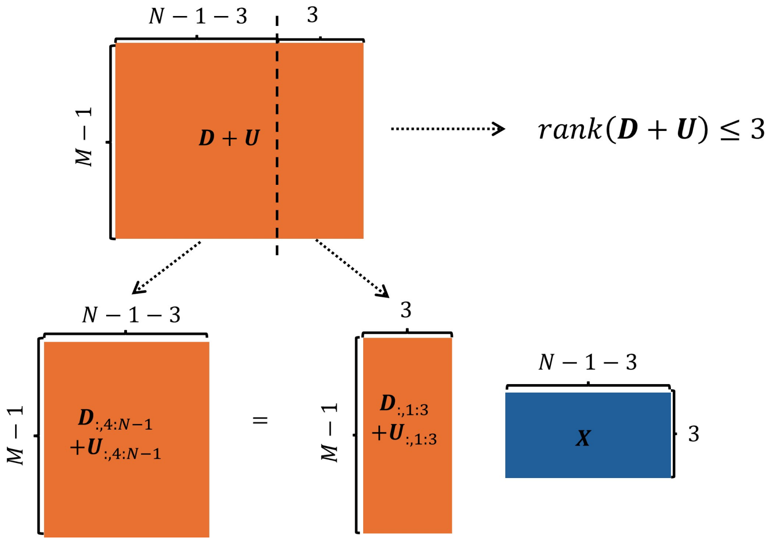


Figure 4.1: The relationship of LRP for the matrix $\mathbf{D}+\mathbf{U}$ [42, 64, 65], where M represents the number of microphones and N denotes the number of sources.

For analytical convenience, assuming the first three column vectors of $\mathbf{D}+\mathbf{U}$ are linearly independent, an unknown matrix is introduced:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{1,1} & \cdots & \mathbf{X}_{1,N-1-3} \\ \mathbf{X}_{2,1} & \cdots & \mathbf{X}_{2,N-1-3} \\ \mathbf{X}_{3,1} & \cdots & \mathbf{X}_{3,N-1-3} \end{bmatrix} \in \mathbb{R}^{3\times(N-1-3)},$$

that enables these vectors to express the others from $\mathbf{D}+\mathbf{U}$ [64, 65] (see Fig. 4.1). This relationship is captured by:

$$(\mathbf{D}_{:,1:3} + \mathbf{U}_{:,1:3})\mathbf{X} = \mathbf{D}_{:,3+1:N-1} + \mathbf{U}_{:,3+1:N-1}. \tag{4.7}$$

Upon inspection of Eq. (4.7), if follows

$$\mathbf{D}_{:,1:3}\mathbf{X} - \mathbf{D}_{:,3+1:N-1} + \mathbf{U}_{:,1:3}\mathbf{X} = \mathbf{U}_{:,3+1:N-1}, \tag{4.8}$$

therefore, this equation can be represented in matrix form as:

$$\begin{bmatrix} \mathbf{D}_{:,1:3} & \mathbf{D}_{:,4:N-1} & \mathbf{U}_{:,1:3} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ -\mathbf{I} \\ \mathbf{X} \end{bmatrix} = \begin{bmatrix} \mathbf{D} & \mathbf{U}_{:,1:3} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ -\mathbf{I} \\ \mathbf{X} \end{bmatrix} = U_{:,4:N-1}, \tag{4.9}$$

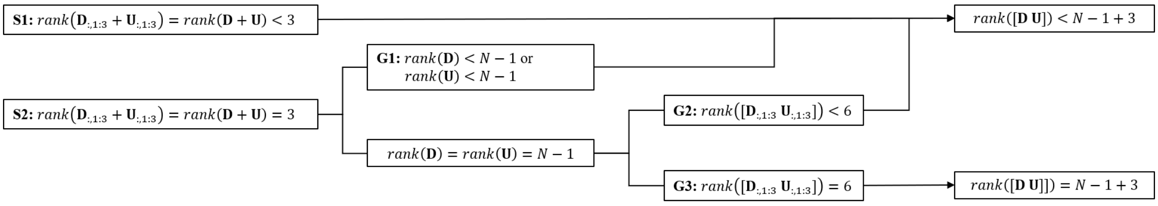


Figure 4.2: Steps to prove LRPV1, namely, $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) < N-1+3$ or $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) = N-1+3$ (N: number of sources; S1: Situation 1; S2: Situation 2; G1: Group 1; G2: Group 2; G3: Group 3).

where $\mathbf{I} \in \mathbb{R}^{(N-1-3)\times(N-1-3)}$ is the identity matrix.

Considering $\begin{bmatrix}\mathbf{X}\\ -\mathbf{I}\\ \mathbf{X}\end{bmatrix} \in \mathbb{R}^{(N-1+3)\times(N-1-3)}$, it is evident that this matrix has $N-1+3$ rows. Moreover, by examining the coefficient and augmented matrices in Eq. (4.9), denoted as $\begin{bmatrix}\mathbf{D} & \mathbf{U}_{:,1:3}\end{bmatrix}$ and $\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}$ respectively, and applying *Theorem 1* [93–95], it follows:

$$rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}_{:,1:3}\end{bmatrix}) = rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) \leq N-1+3. \tag{4.10}$$

Next, I analyze the conditions under which $\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}$ is a low-rank matrix. For a matrix to exhibit low-rank properties, the number of both its rows and columns must exceed its rank. Therefore, I consider two aspects for $\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix} \in \mathbb{R}^{(M-1)\times 2(N-1)}$:

1) The number of columns, $2(N-1)$. Given $N-1>3$, it follows that $2(N-1) > N-1+3$, indicating that the number of columns exceeds the rank of $\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}$, which is $N-1+3$.

2) The number of rows, $M-1$. Since the previous point establishes that the number of columns already exceeds the rank, for $\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}$ to be low-rank, $M-1$ must be greater than $N-1+3$. **This completes the proof for LRPV1**, demonstrating that $rank(\mathbf{T}_1^*) \leq N-1+3$ if $M-1 > N-1+3$.

Discussions for LRPV1

In this part, the discussions of the proposed LRPV1 are presented. From Eq. (4.10), it is evident that $N-1+3$ sets an upper bound on the rank of the matrix $\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}$ if $M-1 > N-1+3$. I will now analyze this upper bound, focusing on the conditions under which $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) < N-1+3$ or $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) = N-1+3$. The proof is divided into two scenarios: $rank(\mathbf{D}+\mathbf{U}) < 3$ and $rank(\mathbf{D}+\mathbf{U}) = 3$. The entire procedure is illustrated in Fig. 4.2.

Situation 1: If $rank(\mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3}) = rank(\mathbf{D}+\mathbf{U}) < 3$, then $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) < N-1+3$.

Proof: From Eq. (4.9), it can be observed that the sub-matrix $\mathbf{I}$ in $\begin{bmatrix} \mathbf{X} \\ -\mathbf{I} \\ \mathbf{X} \end{bmatrix}$ is an identity matrix. Given the uniqueness of the identity matrix, it is necessary to determine whether matrix $\mathbf{X}$ is unique. Based on the LRP [42, 64, 65], the relationship in Eq. (4.7) and *Theorem 1* [93–95], it can be confirmed that matrix $\mathbf{X}$ has multiple solutions if $rank(\mathbf{D}_{:,1:3} + \mathbf{U}_{:,1:3}) = rank(\mathbf{D} + \mathbf{U}) < 3$. Consequently, with Eq. (4.9), it can be concluded that $rank(\begin{bmatrix} \mathbf{D} & \mathbf{U} \end{bmatrix}) < N - 1 + 3$.

This completes the proof for *Situation 1*.

Situation 2: Once $rank(\mathbf{D}_{:,1:3} + \mathbf{U}_{:,1:3}) = rank(\mathbf{D} + \mathbf{U}) = 3$, there are three groups:

- ***Group 1:*** If $rank(\mathbf{D}) < N - 1$ or $rank(\mathbf{U}) < N - 1$, it leads to

$$rank(\begin{bmatrix} \mathbf{D} & \mathbf{U} \end{bmatrix}) < N - 1 + 3. \tag{4.11}$$

Proof: Using basic linear algebra principles [95], it is known that multiplying all elements of a certain column of a matrix by a constant and adding them to another column does not change the rank of the matrix. Hence, by multiplying all elements of the j^{th} column of matrix $\begin{bmatrix} \mathbf{D} & \mathbf{U} \end{bmatrix}$ by 1 and add them to the $\{j + N - 1\}^{th}$ column, where $j = 1, \quad \cdots, \quad N - 1$, it follows

$$rank(\begin{bmatrix} \mathbf{D} & \mathbf{U} \end{bmatrix}) = rank(\begin{bmatrix} \mathbf{D} & \mathbf{D} + \mathbf{U} \end{bmatrix}). \tag{4.12}$$

Since only three columns of matrix $D + U$ are independent, i.e., $rank(\mathbf{D} + \mathbf{U}) = rank(\mathbf{D}_{:,1:3} + \mathbf{U}_{:,1:3}) = 3$, it implies

$$rank(\begin{bmatrix} \mathbf{D} & \mathbf{D} + \mathbf{U} \end{bmatrix}) = rank(\begin{bmatrix} \mathbf{D} & \mathbf{D}_{:,1:3} + \mathbf{U}_{:,1:3} \end{bmatrix}). \tag{4.13}$$

For matrix $\begin{bmatrix} \mathbf{D} & \mathbf{D}_{:,1:3} + \mathbf{U}_{:,1:3} \end{bmatrix}$ in Eq. (4.13), it is evident that if $rank(\mathbf{D}) < N - 1$, it implies that one column vector of matrix $\mathbf{D}$ can be represented by the remaining column vectors, leading to

$$rank(\begin{bmatrix} \mathbf{D} & \mathbf{D}_{:,1:3} + \mathbf{U}_{:,1:3} \end{bmatrix}) < N - 1 + 3. \tag{4.14}$$

Combining Eq. (4.12), Eq. (4.13) and Eq. (4.14), it follows

$$rank(\begin{bmatrix} \mathbf{D} & \mathbf{U} \end{bmatrix}) < N - 1 + 3. \tag{4.15}$$

This completes the proof that $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) < N-1+3$ ***if*** $rank(\mathbf{D}) < N-1$.

Similarity, by multiplying all elements of the $\{j+N-1\}^{th}$ column of matrix $\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}$ by 1, then add them to the j^{th} column, where $j = 1, \quad \cdots, \quad N-1$, it follows

$$rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) = rank(\begin{bmatrix}\mathbf{D}+\mathbf{U} & \mathbf{U}\end{bmatrix}). \tag{4.16}$$

With Eq. (4.16), and following the same steps as Eqs. (4.13), (4.14) and (4.15), it is easy to prove that if $rank(\mathbf{U}) < N-1$, then

$$rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) < N-1+3. \tag{4.17}$$

This completes the proof that $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) < N-1+3$ ***if*** $rank(\mathbf{U}) < N-1$.

Combining Eqs. (4.15) and (4.17), it can be concluded that $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) < N-1+3$ if $rank(\mathbf{D}) < N-1$ or $rank(\mathbf{U}) < N-1$.

This completes the proof for Group 1.

- ***Group 2:*** If $rank(\mathbf{D}) = rank(\mathbf{U}) = N-1$ and $rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{U}_{:,1:3}\end{bmatrix}) < 6$, then

$$rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) < N-1+3. \tag{4.18}$$

Proof: To start, let's consider the precondition $rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{U}_{:,1:3}\end{bmatrix}) < 6$ in ***Group 2***. For any index $4 \leq n \leq N-1$, it follows that

$$rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{U}_{:,1:3}\end{bmatrix}) \leq rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{U}_{:,1:3}\end{bmatrix}) + 1. \tag{4.19}$$

Given $rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{U}_{:,1:3}\end{bmatrix}) < 6$, it follows $rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{U}_{:,1:3}\end{bmatrix}) + 1 < 6+1 = 7$. Thus, from Eq. (4.19), it is derived

$$rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{U}_{:,1:3}\end{bmatrix}) < 7. \tag{4.20}$$

Now, let's use the method of contradiction [96] to prove Eq. (4.18). Assume $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) <$

$N-1+3$ is false. Then, from Eq. (4.10), it is derived

$$rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) = N-1+3. \tag{4.21}$$

With Eq. (4.21), the following observation is made: ***Observation:*** For any $4 \leq n \leq N-1$, it holds that $rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{U}_{:,1:3}\end{bmatrix}) = 7$.

Proof: From Eq. (4.21), for any $4 \leq n \leq N-1$, it follows

$$rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{U}_{:,1:3} & \mathbf{U}_{:,n}\end{bmatrix}) = 4+3 = 7. \tag{4.22}$$

By performing the elementary operations on the matrix $\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{U}_{:,1:3} & \mathbf{U}_{:,n}\end{bmatrix}$ in Eq. (4.22), such as multiplying all elements of the j^{th} column by 1 and adding them to the $\{j+4\}^{th}$ column (for $j = 1, \quad \cdots, \quad 4$), it holds that

$$rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{U}_{:,1:3} & \mathbf{U}_{:,n}\end{bmatrix}) = rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3} & \mathbf{D}_{:,n}+\mathbf{U}_{:,n}\end{bmatrix}) = 7. \tag{4.23}$$

Since $\mathbf{D}_{:,n}+\mathbf{U}_{:,n}$ is dependent on $\mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3}$, i.e., $rank(\begin{bmatrix}\mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3} & \mathbf{D}_{:,n}+\mathbf{U}_{:,n}\end{bmatrix}) = rank(\mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3})$, from Eq. (4.23), it follows

$$rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3} & \mathbf{D}_{:,n}+\mathbf{U}_{:,n}\end{bmatrix}) = rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3}\end{bmatrix}) = 7. \tag{4.24}$$

Performing further elementary operations on the matrix $\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3}\end{bmatrix}$ in Eq. (4.24), such as multiplying all elements of the j^{th} column by -1 and adding them to the $\{j+4\}^{th}$ column (for $j = 1, \quad 2, \quad 3$), it holds that

$$rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3}\end{bmatrix}) = rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{U}_{:,1:3}\end{bmatrix}) = 7. \tag{4.25}$$

This completes the proof for Observation.

On one hand, Eq. (4.20) confirms that $rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{U}_{:,1:3}\end{bmatrix}) < 7$. On the other hand, the ***Observation*** shows that $rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{U}_{:,1:3}\end{bmatrix}) = 7$, this cause a contradiction. Since

Eq. (4.20) is correct, implying $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) = N-1+3$ is wrong and it holds that

$$rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) < N-1+3, \tag{4.26}$$

when $rank(\mathbf{D}) = rank(\mathbf{U}) = N-1$ and $rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{U}_{:,1:3}\end{bmatrix}) < 6$.

This completes the proof for Group 2.

- ***Group 3:*** Given $rank(\mathbf{D}) = rank(\mathbf{U}) = N-1$ and $rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{U}_{:,1:3}\end{bmatrix}) = 6$, it follows

$$rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) = N-1+3. \tag{4.27}$$

Proof: To prove Eq. (4.27), a contradiction method is employed [96]. Assume $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) = N-1+3$ in Eq. (4.27) is wrong, then Eq. (4.10) states $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) \leq N-1+3$, implying that

$$rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) < N-1+3. \tag{4.28}$$

With Eq. (4.28), it follows that $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) = N-1+2$ or $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) < N-1+2$. Both scenarios will be disproven in two steps.

Step 1: Suppose $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) = N-1+2$, the following observations are then made.

Observation 1: The rank of the matrix matrix $\begin{bmatrix}\mathbf{D} & \mathbf{U}_{:,1:3}\end{bmatrix}$ is a function of $N-1$, implying that $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}_{:,1:3}\end{bmatrix}) = N-1+2$.

Proof: Given $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) = N-1+2$ and $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}_{:,1:3}\end{bmatrix}) = rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix})$ in Eq. (4.10), it is clear that

$$rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}_{:,1:3}\end{bmatrix}) = N-1+2. \tag{4.29}$$

This completes the proof for Observation 1.

Observation 2: The rank of the matrix $\begin{bmatrix}\mathbf{D} & \mathbf{U}_{:,1:3}\end{bmatrix}$ is constant and does not vary with $N-1$.

Proof: If $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) = N-1+2$, then for any index $4 \leq n \leq N-1$, it follows:

$$rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{U}_{:,1:3} & \mathbf{U}_{:,n}\end{bmatrix}) = 4+2 = 6 \tag{4.30}$$

By performing elementary column operation on $\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{U}_{:,1:3} & \mathbf{U}_{:,n}\end{bmatrix}$ in Eq. (4.30), it holds that:

$$rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{U}_{:,1:3} & \mathbf{U}_{:,n}\end{bmatrix}) = rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3} & \mathbf{D}_{:,n}+\mathbf{U}_{:,n}\end{bmatrix}) = 6. \tag{4.31}$$

Since $\mathbf{D}_{:,n}+\mathbf{U}_{:,n}$ can be expressed using $\mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3}$, i.e., $rank(\begin{bmatrix}\mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3} & \mathbf{D}_{:,n}+\mathbf{U}_{:,n}\end{bmatrix})$ $= rank(\mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3})$, then from Eq. (4.31), it follows:

$$rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3} & \mathbf{D}_{:,n}+\mathbf{U}_{:,n}\end{bmatrix}) = rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3}\end{bmatrix}) = 6. \tag{4.32}$$

Applying further elementary column operations on $\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3}\end{bmatrix}$ in Eq. (4.32), it can be observed that:

$$rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3}\end{bmatrix}) = rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{U}_{:,1:3}\end{bmatrix}) = 6. \tag{4.33}$$

From the initial condition that $rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{U}_{:,1:3}\end{bmatrix}) = 6$ in ***Group 3*** and Eq. (4.33), it follows:

$$rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{U}_{:,1:3}\end{bmatrix}) = rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{U}_{:,1:3}\end{bmatrix}) = 6. \tag{4.34}$$

Eq. (4.34) implies that any $\mathbf{D}_{:,n}$ can be expressed by $\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{U}_{:,1:3}\end{bmatrix}$. Since $n \subset \{4, \ \cdots, \ N-1\}$, it holds that:

$$rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{U}_{:,1:3}\end{bmatrix}) = rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,4} & \cdots \mathbf{D}_{:,N-1} & \mathbf{U}_{:,1:3}\end{bmatrix}) = rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}_{:,1:3}\end{bmatrix}) = 6. \tag{4.35}$$

This completes the proof for Observation 2.

Observation 1 indicates that the rank of $\begin{bmatrix}\mathbf{D} & \mathbf{U}_{:,1:3}\end{bmatrix}$ varies with $N-1$, while ***Observation 2*** asserts that this rank is constant, leading to a contradiction. ***Therefore,*** $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) =$ $N-1+2$ ***is incorrect***.

Step 2: Suppose $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) < N-1+2$, the following observation is made.

Observation 3: $rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{U}_{:,1:3}\end{bmatrix}) < 6$.

Proof: If $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) < N-1+2$, for any $4 \leq n \leq N-1$, it follows:

$$rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{U}_{:,1:3} & \mathbf{U}_{:,n}\end{bmatrix}) < 4+2 = 6. \tag{4.36}$$

Performing elementary column operations on $\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{U}_{:,1:4} & \mathbf{U}_{:,n}\end{bmatrix}$ in Eq. (4.36), it holds that:

$$rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{U}_{:,1:3} & \mathbf{U}_{:,n}\end{bmatrix}) = rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3} & \mathbf{D}_{:,n}+\mathbf{U}_{:,n}\end{bmatrix}) < 6. \tag{4.37}$$

Since $\mathbf{D}_{:,n}+\mathbf{U}_{:,n}$ can be expressed using $\mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3}$, from Eq. (4.37), it follows:

$$rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3} & \mathbf{D}_{:,n}+\mathbf{U}_{:,n}\end{bmatrix}) = rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3}\end{bmatrix}) < 6. \tag{4.38}$$

Applying further elementary column operations on $\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3}\end{bmatrix}$ in Eq. (4.38), it follows:

$$rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3}\end{bmatrix}) = rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{U}_{:,1:3}\end{bmatrix}) < 6. \tag{4.39}$$

Given that

$$rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{U}_{:,1:3}\end{bmatrix}) \leq rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{D}_{:,n} & \mathbf{U}_{:,1:3}\end{bmatrix}), \tag{4.40}$$

then with Eq. (4.39), it holds that:

$$rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{U}_{:,1:3}\end{bmatrix}) < 6. \tag{4.41}$$

This completes the proof for Observation 3.

Observation 3 confirms the rank on $\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{U}_{:,1:3}\end{bmatrix} < 6$. on the other hand, the precondition in the ***Group 3*** indicates $\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{U}_{:,1:3}\end{bmatrix} = 6$, leading to a contradiction. ***Thus, it can be concluded that*** $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) < N-1+2$ ***is incorrect***.

In conclusion, ***Step 1*** indicates that $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) = N-1+2$ is incorrect and $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) < N-1+2$ is incorrect has been confirmed by ***Step 2***, implying that $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) \leq N-1+2$ is wrong, so that $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\end{bmatrix}) = N-1+3$ is correct.

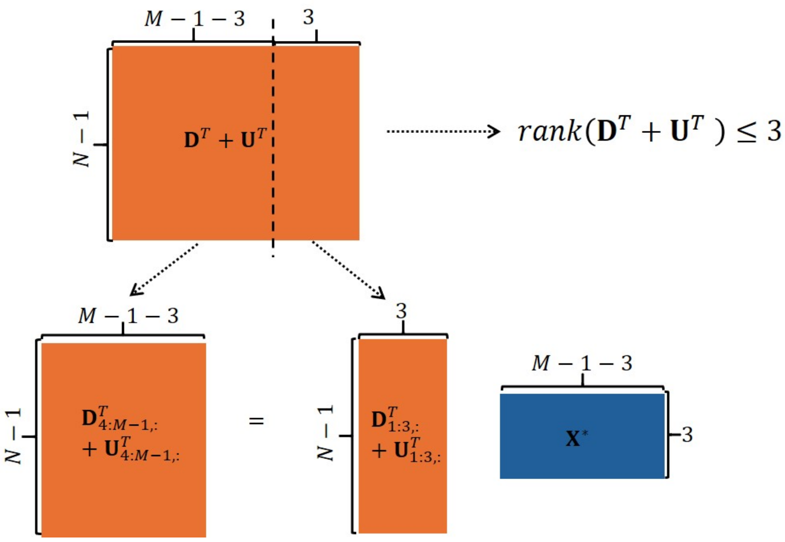


Figure 4.3: The relationship of LRP for the matrix $\mathbf{D}+\mathbf{U}$ [42,64,65], where M represents the number of microphones and N denotes the number of sources.

This completes the proof for Group 3.

4.3.2 Proof for Proposed LRPV2

This part presents the proof for the proposed LRPV2 in Eq.(4.5), subsequently, the discussions leading to $rank\,(\mathbf{T}_2^*) < M-1+3$ or $rank\,(\mathbf{T}_2^*) = M-1+3$ are presented.

From the state-of-the-art LRP methods [42,64,65], it is obvious that $rank(\mathbf{D}+\mathbf{U}) = rank(\mathbf{D}^T+\mathbf{U}^T) \leq 3$. This indicates that three row vectors from the matrix $\mathbf{D}+\mathbf{U}$ can represent the remaining row vectors [64]. For analytical convenience, the first three row vectors of $\mathbf{D}+\mathbf{U}$ are assumed to be independent. Hence, an unknown matrix is introduced

$$\mathbf{X}^* = \begin{bmatrix} \mathbf{X}^*{}_{1,1} & \cdots & \mathbf{X}^*{}_{1,M-1-3} \\ \mathbf{X}^*{}_{2,1} & \cdots & \mathbf{X}^*{}_{2,M-1-3} \\ \mathbf{X}^*{}_{3,1} & \cdots & \mathbf{X}^*{}_{3,M-1-3} \end{bmatrix} \in \mathbb{R}^{3\times(M-1-3)},$$

that allows the first three row vectors of $\mathbf{D}+\mathbf{U}$ to represent the remaining row vectors (see Fig. 4.3) [64,65], expressed as:

$$(\mathbf{D}^T_{1:3,:}+\mathbf{U}^T_{1:3,:})\mathbf{X}^* = \mathbf{D}^T_{3+1:M-1,:}+\mathbf{U}^T_{3+1:M-1,:}. \tag{4.42}$$

Upon inspecting Eq. (4.42), it implies

$$\mathbf{D}_{1:3,:}^{T}\mathbf{X}^{*}-\mathbf{D}_{3+1:M-1,:}^{T}+\mathbf{U}_{1:3,:}^{T}\mathbf{X}^{*}=\mathbf{U}_{3+1:M-1,:}^{T}. \tag{4.43}$$

Next, Eq. (4.43) can be expressed as the matrix multiplication form, leading to

$$\begin{bmatrix}\mathbf{D}_{1:3,:}^{T} & \mathbf{D}_{4:M-1,:}^{T} & \mathbf{U}_{1:3,:}^{T}\end{bmatrix}\begin{bmatrix}\mathbf{X}^{*}\\ -\mathbf{I}\\ \mathbf{X}^{*}\end{bmatrix}=\begin{bmatrix}\mathbf{D}^{T} & \mathbf{U}_{1:3,:}^{T}\end{bmatrix}\begin{bmatrix}\mathbf{X}^{*}\\ -\mathbf{I}\\ \mathbf{X}^{*}\end{bmatrix}=\mathbf{U}_{4:M-1,:}^{T}, \tag{4.44}$$

where $\mathbf{I}\in\mathbb{R}^{(M-1-3)\times(M-1-3)}$ is identity matrix.

Given that the matrix $\begin{bmatrix}\mathbf{X}^{*}\\ -\mathbf{I}\\ \mathbf{X}^{*}\end{bmatrix}\in\mathbb{R}^{(M-1+3)\times(M-1-3)}$, has $M-1+3$ rows, which is the number of rows in this matrix. Additionally, notice that the coefficient matrix and the augmented matrix in Eq. (4.44) are $\begin{bmatrix}\mathbf{D}^{T} & \mathbf{U}_{1:3,:}^{T}\end{bmatrix}$ and $\begin{bmatrix}\mathbf{D}^{T} & \mathbf{U}^{T}\end{bmatrix}$, respectively. Therefore, according to *Theorem 1* [93–95], it follows

$$rank(\begin{bmatrix}\mathbf{D}^{T} & \mathbf{U}_{1:3,:}^{T}\end{bmatrix})=rank(\begin{bmatrix}\mathbf{D}^{T} & \mathbf{U}^{T}\end{bmatrix})\leq M-1+3. \tag{4.45}$$

Next, the conditions that make the matrix $\begin{bmatrix}\mathbf{D}^{T} & \mathbf{U}^{T}\end{bmatrix}$ low-rank are examined. For matrix to have low rank property, both the number of rows and columns must exceed the corresponding rank. Therefore, two aspects are evaluated: the number of columns and rows of the matrix $\begin{bmatrix}\mathbf{D}^{T} & \mathbf{U}^{T}\end{bmatrix}\in\mathbb{R}^{(N-1)\times 2(M-1)}$.

1) First, let's consider the number of columns in the matrix $\begin{bmatrix}\mathbf{D}^{T} & \mathbf{U}^{T}\end{bmatrix}$, which is $2(M-1)$. Given that $M-1>3$, it implies that $2(M-1)>M-1+3$, meaning the number of columns in the matrix $\begin{bmatrix}\mathbf{D}^{T} & \mathbf{U}^{T}\end{bmatrix}$ exceeds its rank, $M-1+3$.

2) Next, the number of rows in the matrix $\begin{bmatrix}\mathbf{D}^{T} & \mathbf{U}^{T}\end{bmatrix}$ is considered., which is $N-1$. Since it has already been established that the number of columns in the matrix $\begin{bmatrix}\mathbf{D}^{T} & \mathbf{U}^{T}\end{bmatrix}$ is greater than its corresponding rank, if the number of rows in the matrix $\begin{bmatrix}\mathbf{D}^{T} & \mathbf{U}^{T}\end{bmatrix}$ is also greater than its rank, i.e., $N-1>M-1+3$, then it can be concluded that the matrix $\begin{bmatrix}\mathbf{D}^{T} & \mathbf{U}^{T}\end{bmatrix}$ is a low-rank matrix.

This completes the proof for LRPV2 that $rank(\mathbf{T}_{2}^{*})\leq M-1+3$ if $N-1>M-1+3$.

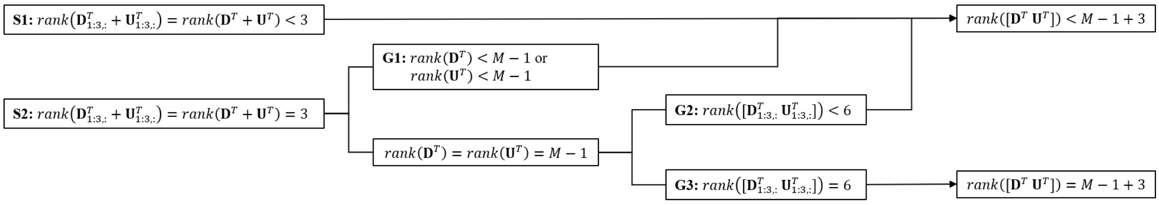


Figure 4.4: Steps to prove LRPV2, namely, $rank(\begin{bmatrix}\mathbf{D}^T & \mathbf{U}^T\end{bmatrix}) < M-1+3$ or $rank(\begin{bmatrix}\mathbf{D}^T & \mathbf{U}^T\end{bmatrix}) = M-1+3$ (M: number of microphones; S1: Situation 1; S2: Situation 2; G1: Group 1; G2: Group 2; G3: Group 3).

Discussions for LRPV2

From Eq. (4.45), it is evident that $M-1+3$ establishes an upper limit for the rank of the matrix $\begin{bmatrix}\mathbf{D}^T & \mathbf{U}^T\end{bmatrix}$ if $N-1 > M-1+3$. Let's explore this upper limit by analyzing the conditions under which $rank(\begin{bmatrix}\mathbf{D}^T & \mathbf{U}^T\end{bmatrix}) < M-1+3$ or $rank(\begin{bmatrix}\mathbf{D}^T & \mathbf{U}^T\end{bmatrix}) = M-1+3$. The proof can be divided into two situations: $rank(\mathbf{D}^T + \mathbf{U}^T) < 3$ and $rank(\mathbf{D}^T + \mathbf{U}^T) = 3$. The entire procedure is illustrated in Fig. 4.4.

Situation 1: If $rank(\mathbf{D}^T_{1:3,:} + \mathbf{U}^T_{1:3,:}) = rank(\mathbf{D}^T + \mathbf{U}^T) < 3$, then $rank(\begin{bmatrix}\mathbf{D}^T & \mathbf{U}^T\end{bmatrix}) < M-1+3$.

Situation 2: If $rank(\mathbf{D}^T_{1:3,:} + \mathbf{U}^T_{1:3,:}) = rank(\mathbf{D}^T + \mathbf{U}^T) = 3$, three groups are considered.

- ***Group 1:*** If $rank(\mathbf{D}^T) < M-1$ or $rank(\mathbf{U}^T) < M-1$, it follows that
$$rank(\begin{bmatrix}\mathbf{D}^T & \mathbf{U}^T\end{bmatrix}) < M-1+3. \tag{4.46}$$

- ***Group 2:*** If $rank(\mathbf{D}^T) = rank(\mathbf{U}^T) = M-1$ and $rank(\begin{bmatrix}\mathbf{D}^T_{1:3,:} & \mathbf{U}^T_{1:3,:}\end{bmatrix}) < 6$, it follows
$$rank(\begin{bmatrix}\mathbf{D}^T & \mathbf{U}^T\end{bmatrix}) < M-1+3. \tag{4.47}$$

- ***Group 3:*** If $rank(\mathbf{D}^T) = rank(\mathbf{U}^T) = M-1$ and $rank(\begin{bmatrix}\mathbf{D}^T_{1:3,:} & \mathbf{U}^T_{1:3,:}\end{bmatrix}) = 6$, it follows that
$$rank(\begin{bmatrix}\mathbf{D}^T & \mathbf{U}^T\end{bmatrix}) = M-1+3. \tag{4.48}$$

Proof: Using the same methodology as for LRPV1 in Section 4.3.1, the two situations above are easily proven.

This completes the proof for the conditions (see Fig. 4.4 for those conditions) that lead to $rank(\begin{bmatrix}\mathbf{D}^T & \mathbf{U}^T\end{bmatrix}) = M-1+3$ ***or*** $rank(\begin{bmatrix}\mathbf{D}^T & \mathbf{U}^T\end{bmatrix}) < M-1+3$.

4.3.3 Proof for Proposed LRPV3

This part provides the proof for the proposed LRPV3 in Eq. (4.6). Next, the discussions that result in either $rank\left(\mathbf{T}_3^*\right) < min(N-1+3, M-1+3)$ or $rank\left(\mathbf{T}_3^*\right) = min(N-1+3, M-1+3)$ are presented.

Two situations for LRPV3 are considered: $M \geq N$ and $M < N$.

If $M \geq N$, the proof for LRPV3 in Eq. (4.6) simplifies to:

$$rank\left(\begin{bmatrix} \mathbf{D} & \mathbf{U} \\ \mathbf{U} & \mathbf{D} \end{bmatrix}\right) \leq N-1+3. \tag{4.49}$$

From Eq. (4.7), it follows

$$\begin{cases} \mathbf{D}_{:,1:3}\mathbf{X} - \mathbf{D}_{:,3+1:N-1} + \mathbf{U}_{:,1:3}\mathbf{X} = \mathbf{U}_{:,3+1:N-1} \\ \mathbf{U}_{:,1:3}\mathbf{X} - \mathbf{U}_{:,3+1:N-1} + \mathbf{D}_{:,1:3}\mathbf{X} = \mathbf{D}_{:,3+1:N-1} \end{cases}. \tag{4.50}$$

Then Eq. (4.50) in matrix multiplication form can be written as:

$$\begin{cases} \begin{bmatrix} \mathbf{D}_{:,1:3} & \mathbf{D}_{:,3+1:N-1} & \mathbf{U}_{:,1:3} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ -\mathbf{I} \\ \mathbf{X} \end{bmatrix} = \mathbf{U}_{:,3+1:N-1} \\ \begin{bmatrix} \mathbf{U}_{:,1:3} & \mathbf{U}_{:,3+1:N-1} & \mathbf{D}_{:,1:3} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ -\mathbf{I} \\ \mathbf{X} \end{bmatrix} = \mathbf{D}_{:,3+1:N-1} \end{cases}, \tag{4.51}$$

where $\mathbf{I} \in \mathbb{R}^{(N-1-3)\times(N-1-3)}$ is the identity matrix. Thus, Eq. (4.51) can be written as

$$\begin{bmatrix} \mathbf{D} & \mathbf{U}_{:,1:3} \\ \mathbf{U} & \mathbf{D}_{:,1:3} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ -\mathbf{I} \\ \mathbf{X} \end{bmatrix} = \begin{bmatrix} \mathbf{U}_{:,3+1:N-1} \\ \mathbf{D}_{:,3+1:N-1} \end{bmatrix}. \tag{4.52}$$

Since matrix $\begin{bmatrix} \mathbf{X} \\ -\mathbf{I} \\ \mathbf{X} \end{bmatrix} \in \mathbb{R}^{(N-1+3)\times(N-1-3)}$, the number of rows in $\begin{bmatrix} \mathbf{X} \\ -\mathbf{I} \\ \mathbf{X} \end{bmatrix}$ is $N-1+3$.Additionally, the coefficient matrix and the augmented matrix in Eq. (4.52) are $\begin{bmatrix} \mathbf{D} & \mathbf{U}_{:,1:3} \\ \mathbf{U} & \mathbf{D}_{:,1:3} \end{bmatrix}$ and $\begin{bmatrix} \mathbf{D} & \mathbf{U} \\ \mathbf{U} & \mathbf{D} \end{bmatrix}$, respectively.

Based on *Theorem 1* [93–95], it follows:

$$rank(\begin{bmatrix} \mathbf{D} & \mathbf{U}_{:,1:3} \\ \mathbf{U} & \mathbf{D}_{:,1:3} \end{bmatrix}) = rank(\begin{bmatrix} \mathbf{D} & \mathbf{U} \\ \mathbf{U} & \mathbf{D} \end{bmatrix}) \leq N-1+3. \tag{4.53}$$

This completes the proof for LRPV3 when $M \geq N$.

If $M < N$, the proof for LRPV3 in Eq. (4.6) transforms to:

$$rank\left(\begin{bmatrix} \mathbf{D} & \mathbf{U} \\ \mathbf{U} & \mathbf{D} \end{bmatrix}\right) \leq M-1+3. \tag{4.54}$$

Upon inspection of Eq. (4.42), it holds that:

$$\begin{cases} \mathbf{D}^T_{1:3,:}\mathbf{X}^* - \mathbf{D}^T_{3+1:M-1,:} + \mathbf{U}^T_{1:3,:}\mathbf{X}^* = \mathbf{U}^T_{3+1:M-1,:} \\ \mathbf{U}^T_{1:3,:}\mathbf{X}^* - \mathbf{U}^T_{3+1:M-1,:} + \mathbf{D}^T_{1:3,:}\mathbf{X}^* = \mathbf{D}^T_{3+1:M-1,:} \end{cases}, \tag{4.55}$$

therefore, Eq. (4.55) is derived as matrix multiplication form:

$$\begin{cases} \begin{bmatrix} \mathbf{D}^T_{1:3,:} & \mathbf{D}^T_{3+1:M-1,:} & \mathbf{U}^T_{1:3,:} \end{bmatrix} \begin{bmatrix} \mathbf{X}^* \\ -\mathbf{I} \\ \mathbf{X}^* \end{bmatrix} = \mathbf{U}^T_{3+1:M-1,:} \\ \begin{bmatrix} \mathbf{U}^T_{1:3,:} & \mathbf{U}^T_{3+1:M-1,:} & \mathbf{D}^T_{1:3,:} \end{bmatrix} \begin{bmatrix} \mathbf{X}^* \\ -\mathbf{I} \\ \mathbf{X}^* \end{bmatrix} = \mathbf{D}^T_{3+1:M-1,:} \end{cases}, \tag{4.56}$$

where $\mathbf{I} \in \mathbb{R}^{(M-1-3)\times(M-1-3)}$ is the identity matrix.

Moreover, Eq. (4.56) is rewritten as

$$\begin{bmatrix} \mathbf{D}^T & \mathbf{U}^T_{1:3,:} \\ \mathbf{U}^T & \mathbf{D}^T_{1:3,:} \end{bmatrix} \begin{bmatrix} \mathbf{X}^* \\ -\mathbf{I} \\ \mathbf{X}^* \end{bmatrix} = \begin{bmatrix} \mathbf{U}^T_{3+1:M-1,:} \\ \mathbf{D}^T_{3+1:M-1,:} \end{bmatrix}. \tag{4.57}$$

Given that matrix $\begin{bmatrix} \mathbf{X}^* \\ -\mathbf{I} \\ \mathbf{X}^* \end{bmatrix} \in \mathbb{R}^{(M-1+3)\times(M-1-3)}$, it is evident that matrix $\begin{bmatrix} \mathbf{X}^* \\ -\mathbf{I} \\ \mathbf{X}^* \end{bmatrix}$ has $M-1+3$ rows. Additionally, it is observed that the coefficient matrix and the augmented matrix in Eq. (4.57) are $\begin{bmatrix} \mathbf{D}^T & \mathbf{U}^T_{1:3,:} \\ \mathbf{U}^T & \mathbf{D}^T_{1:3,:} \end{bmatrix}$ and $\begin{bmatrix} \mathbf{D}^T & \mathbf{U}^T \\ \mathbf{U}^T & \mathbf{D}^T \end{bmatrix}$, respectively. According to *Theorem 1* [93–95], it follows:

$$rank(\begin{bmatrix} \mathbf{D}^T & \mathbf{U}^T_{1:3,:} \\ \mathbf{U}^T & \mathbf{D}^T_{1:3,:} \end{bmatrix}) = rank(\begin{bmatrix} \mathbf{D}^T & \mathbf{U}^T \\ \mathbf{U}^T & \mathbf{D}^T \end{bmatrix}) \leq M-1+3. \tag{4.58}$$

Given that $\begin{bmatrix} \mathbf{D}^T & \mathbf{U}^T \\ \mathbf{U}^T & \mathbf{D}^T \end{bmatrix} = \begin{bmatrix} \mathbf{D} & \mathbf{U} \\ \mathbf{U} & \mathbf{D} \end{bmatrix}^T$, it follows:

$$rank(\begin{bmatrix} \mathbf{D} & \mathbf{U} \\ \mathbf{U} & \mathbf{D} \end{bmatrix}^T) = rank(\begin{bmatrix} \mathbf{D} & \mathbf{U} \\ \mathbf{U} & \mathbf{D} \end{bmatrix}) \leq M-1+3. \tag{4.59}$$

This completes the proof for LRPV3 when $M < N$.

Finally, combining Eq. (4.53) when $M \geq N$ with Eq. (4.59) when $M < N$, it holds that:

$$rank(\begin{bmatrix} \mathbf{D} & \mathbf{U} \\ \mathbf{U} & \mathbf{D} \end{bmatrix}) \leq min(N-1+3, M-1+3). \tag{4.60}$$

This completes entire proof for LRPV3.

Discussions for LRPV3

The discussions of LRPV3 in Eq. (4.53) are first presented when $M \geq N$, followed by the discussions of LRPV3 in Eq. (4.58) when $M < N$. The entire procedures are illustrated in Figs. 4.5 and 4.6 for $M \geq N$ and $M < N$, respectively.

(1) $M \geq N$: From Eq. (4.53), it is evident that $N-1+3$ serves as an upper bound for the rank of the matrix $\begin{bmatrix} \mathbf{D} & \mathbf{U} \\ \mathbf{U} & \mathbf{D} \end{bmatrix}$ under the condition $M \geq N$. The upper boundary is analyzed with two possible situations: either $rank(\mathbf{D}+\mathbf{U}) < 3$ or $rank(\mathbf{D}+\mathbf{U}) = 3$.

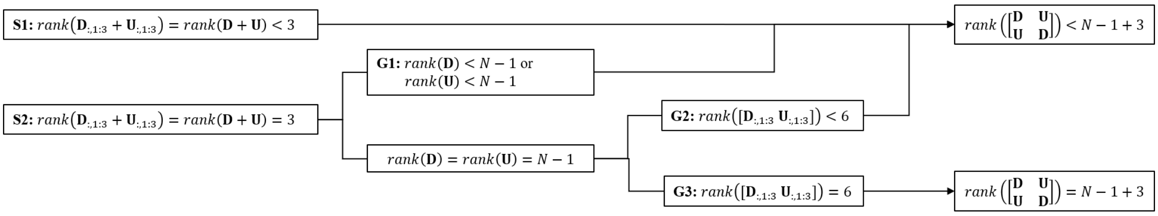


Figure 4.5: Steps to prove LRPV3 when $M \geq N$, namely, $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\\ \mathbf{U} & \mathbf{D}\end{bmatrix}) < N-1+3$ or $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\\ \mathbf{U} & \mathbf{D}\end{bmatrix}) = N-1+3$ (N: number of sources; S1: Situation 1; S2: Situation 2; G1: Group 1; G2: Group 2; G3: Group 3).

Situation 1: If $rank(\mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3}) = rank(\mathbf{D}+\mathbf{U}) < 3$, it follows that $rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\\ \mathbf{U} & \mathbf{D}\end{bmatrix}) < N-1+3$.

Situation 2: If $rank((\mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3}) = rank(\mathbf{D}+\mathbf{U}) = 3$, three groups emerge:

- ***Group 1:*** If $rank(\mathbf{D}) < N-1$ or $rank(\mathbf{U}) < N-1$, it results in

$$rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\\ \mathbf{U} & \mathbf{D}\end{bmatrix}) < N-1+3. \tag{4.61}$$

- ***Group 2:*** If $rank(\mathbf{D}) = rank(\mathbf{U}) = N-1$ and $rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{U}_{:,1:3}\end{bmatrix}) < 6$, it leads to

$$rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\\ \mathbf{U} & \mathbf{D}\end{bmatrix}) < N-1+3. \tag{4.62}$$

- ***Group 3:*** If $rank(\mathbf{D}) = rank(\mathbf{U}) = N-1$ and $rank(\begin{bmatrix}\mathbf{D}_{:,1:3} & \mathbf{U}_{:,1:3}\end{bmatrix}) = 6$, it results in

$$rank(\begin{bmatrix}\mathbf{D} & \mathbf{U}\\ \mathbf{U} & \mathbf{D}\end{bmatrix}) = N-1+3. \tag{4.63}$$

Proof: Initially, by performing elementary operations on the matrix $\begin{bmatrix}\mathbf{D} & \mathbf{U}\\ \mathbf{U} & \mathbf{D}\end{bmatrix}$–multiply all of elements of i^{th} column by 1 and adding them to the $\{i+M-1\}^{th}$ column—-it can be derived from Eq. (4.53):

$$rank(\begin{bmatrix} \mathbf{D} & \mathbf{U} \\ \mathbf{U} & \mathbf{D} \end{bmatrix}) = rank(\begin{bmatrix} \mathbf{D} & \mathbf{U} \\ \mathbf{D}+\mathbf{U} & \mathbf{U}+\mathbf{D} \end{bmatrix}) \leq N-1+3. \tag{4.64}$$

Subsequently, employing the proof procedure similar to LRPV1 in Section 4.3.1 for $\begin{bmatrix} \mathbf{D} & \mathbf{U} \\ \mathbf{D}+\mathbf{U} & \mathbf{U}+\mathbf{D} \end{bmatrix}$ in Eq. (4.64) verifies the aforementioned situations.

This completes the proof concerning conditions (see Fig. 4.5 for those conditions) that result in $rank(\begin{bmatrix} \mathbf{D} & \mathbf{U} \\ \mathbf{U} & \mathbf{D} \end{bmatrix}) < N-1+3$ ***or*** $rank(\begin{bmatrix} \mathbf{D} & \mathbf{U} \\ \mathbf{U} & \mathbf{D} \end{bmatrix}) = N-1+3$ ***when*** $M \geq N$***.***

(2) $M < N$: From Eq. (4.58), it is observed that $M-1+3$ sets an upper boundary for the rank of the matrix $\begin{bmatrix} \mathbf{D}^T & \mathbf{U}^T \\ \mathbf{U}^T & \mathbf{D}^T \end{bmatrix}$ under the condition $M < N$. This upper boundary is analyzed with two scenarios: either $rank(\mathbf{D}^T+\mathbf{U}^T) < 3$ or $rank(\mathbf{D}^T+\mathbf{U}^T) = 3$.

Situation 1: If $rank(\mathbf{D}^T_{1:3,:}+\mathbf{U}^T_{1:3,:}) = rank(\mathbf{D}^T+\mathbf{U}^T) < 3$, then $rank(\begin{bmatrix} \mathbf{D}^T & \mathbf{U}^T \\ \mathbf{U}^T & \mathbf{D}^T \end{bmatrix}) < M-1+3$.

Situation 2: If $rank(\mathbf{D}^T_{1:3,:}+\mathbf{U}^T_{1:3,:}) = rank(\mathbf{D}^T+\mathbf{U}^T) = 3$, three groups are identified:

- ***Group 1:*** If $rank(\mathbf{D}^T) < M-1$ or $rank(\mathbf{U}^T) < M-1$, it results in

$$rank(\begin{bmatrix} \mathbf{D}^T & \mathbf{U}^T \\ \mathbf{U}^T & \mathbf{D}^T \end{bmatrix}) < M-1+3. \tag{4.65}$$

- ***Group 2:*** If $rank(\mathbf{D}^T) = rank(\mathbf{U}^T) = M-1$ and $rank(\begin{bmatrix} \mathbf{D}^T_{1:3,:} & \mathbf{U}^T_{1:3,:} \end{bmatrix}) < 6$, it follows

$$rank(\begin{bmatrix} \mathbf{D}^T & \mathbf{U}^T \\ \mathbf{U}^T & \mathbf{D}^T \end{bmatrix}) < M-1+3. \tag{4.66}$$

- ***Group 3:*** If $rank(\mathbf{D}^T) = rank(\mathbf{U}^T) = M-1$ and $rank(\begin{bmatrix} \mathbf{D}^T_{1:3,:} & \mathbf{U}^T_{1:3,:} \end{bmatrix}) = 6$, it results in

$$rank(\begin{bmatrix} \mathbf{D}^T & \mathbf{U}^T \\ \mathbf{U}^T & \mathbf{D}^T \end{bmatrix}) = M-1+3. \tag{4.67}$$

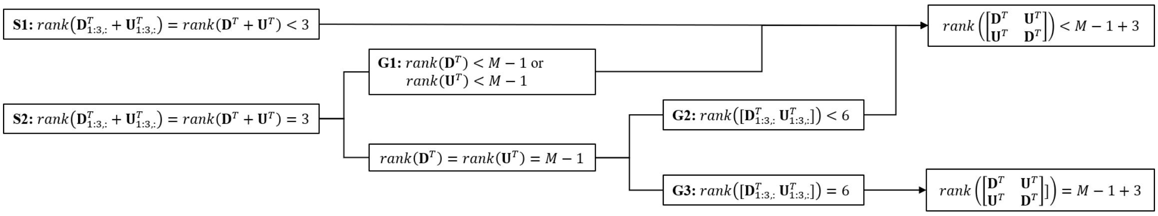


Figure 4.6: Steps to prove LRPV3 when $M < N$, namely, $rank(\begin{bmatrix} \mathbf{D}^T & \mathbf{U}^T \\ \mathbf{U}^T & \mathbf{D}^T \end{bmatrix}) < M - 1 + 3$ or $rank(\begin{bmatrix} \mathbf{D}^T & \mathbf{U}^T \\ \mathbf{U}^T & \mathbf{D}^T \end{bmatrix}) = M - 1 + 3$ (M: number of microphones; S1: Situation 1; S2: Situation 2; G1: Group 1; G2: Group 2; G3: Group 3).

Proof: Initially, by performing elementary operations on the matrix $\begin{bmatrix} \mathbf{D}^T & \mathbf{U}^T \\ \mathbf{U}^T & \mathbf{D}^T \end{bmatrix}$–multiplying all elements of the j^{th} row by 1 and adding them to the $\{j+N-1\}^{th}$ column–I derive from Eq. (4.58):

$$rank(\begin{bmatrix} \mathbf{D}^T & \mathbf{U}^T \\ \mathbf{U}^T & \mathbf{D}^T \end{bmatrix}) = rank(\begin{bmatrix} \mathbf{D}^T & \mathbf{U}^T \\ \mathbf{D}^T + \mathbf{U}^T & \mathbf{U}^T + \mathbf{D}^T \end{bmatrix}) \leq M - 1 + 3. \tag{4.68}$$

Subsequently, employing the proof procedure as LRPV1 in Section 4.3.1 for $\begin{bmatrix} \mathbf{D}^T & \mathbf{U}^T \\ \mathbf{D}^T + \mathbf{U}^T & \mathbf{U}^T + \mathbf{D}^T \end{bmatrix}$ in Eq. (4.68) verifies the aforementioned situations when $M < N$.

This completes the proof concerning $rank(\begin{bmatrix} \mathbf{D}^T & \mathbf{U}^T \\ \mathbf{U}^T & \mathbf{D}^T \end{bmatrix}) < M - 1 + 3$ ***or*** $rank(\begin{bmatrix} \mathbf{D}^T & \mathbf{U}^T \\ \mathbf{U}^T & \mathbf{D}^T \end{bmatrix}) = M - 1 + 3$ ***under different conditions (see Fig. 4.6 for those conditions) when*** $M < N$***.***

4.4 Proposed CLRA Algorithm

This section begins by outlining the four linear constraints, utilizing low-rank structural information as employed by LRP in Section 2.3 and the three LRP variants in Section 4.2. Using these constraints, I introduce a CLRA method, which is then applied to UTIm estimation.

Both LRP and the proposed three LRP variants exploit the low-rank structural information between UTIm and TOA/TDOA. This allows UTIm to leverage a more extensive pool of low-rank structural information compared to LRP alone, thereby facilitating the derivation of global solutions for UTIm. This process involves two main steps. First, based on the aforementioned four LRPs, corresponding linear constraints are formulated. Subsequently, these constraints are combined to formulate an objective function through the proposed CLRA method. The Gaussian Newton method [97] is then applied to solve this objective function. This approach helps mitigate the sub-optimal and locally minimal values of UTIm resulting from the randomness pertaining to the initialization of UTIm.

4.4.1 Linear constraint based on LRP

Upon inspecting Eqs. (2.8) and (2.10), matrices $\mathbf{D}$ and $\mathbf{U}$ can be expressed as

$$\begin{cases} \mathbf{D} = \begin{bmatrix} \mathbf{D}_{:,1:3} & \mathbf{D}_{:,4:N-1} \end{bmatrix} \\ \mathbf{U} = \begin{bmatrix} \mathbf{U}_{:,1:3} & \mathbf{U}_{:,4:N-1} \end{bmatrix} \end{cases} . \tag{4.69}$$

Consequently, it is established that

$$rank([\mathbf{D}_{:,1:3} + \mathbf{U}_{:,1:3} \;\; \mathbf{D}_{:,4:N-1} + \mathbf{U}_{:,4:N-1}]) = rank(\mathbf{D}_{:,1:3} + \mathbf{U}_{:,1:3}) \leq 3. \tag{4.70}$$

From Eq. (4.70), I infer the existence of a matrix $\mathbf{X}$ such that

$$(\mathbf{D}_{:,1:3} + \mathbf{U}_{:,1:3})\mathbf{X} = \mathbf{D}_{:,4:N-1} + \mathbf{U}_{:,4:N-1}, \tag{4.71}$$

where $\mathbf{X} \in R^{3\times(N-1-3)}$ is an unknown matrix to be estimated in Section 4.4.5.

4.4.2 Linear constraint based on LRPV1

Upon examining Eqs. (4.1) and (4.4), $\mathbf{T}^*_\mathbf{1}$ is decomposed as $\mathbf{T}^*_\mathbf{1} = \begin{bmatrix} \mathbf{T}^*_\mathbf{11} & \mathbf{T}^*_\mathbf{12} \end{bmatrix}$, where $\mathbf{T}^*_\mathbf{11} \in \mathbb{R}^{(M-1)\times(N-1+3)}$ and $\mathbf{T}^*_\mathbf{12} \in \mathbb{R}^{(M-1)\times(N-1-3)}$. This yields

$$rank(\begin{bmatrix} \mathbf{T}^*_\mathbf{11} & \mathbf{T}^*_\mathbf{12} \end{bmatrix}) = rank(\mathbf{T}^*_\mathbf{11}) \leq N-1+3. \tag{4.72}$$

From Eq. (4.72), I infer the existence of a matrix $\mathbf{Z}$ such that

$$\mathbf{T}^*_\mathbf{11}\mathbf{Z} = \mathbf{T}^*_\mathbf{12}, \tag{4.73}$$

where $\mathbf{Z} \in R^{(N-1+3)\times(N-1-3)}$ is an unknown matrix to be estimated in Section 4.4.5.

4.4.3 Linear constraint based on LRPV2

Upon inspecting Eqs. (4.2) and (4.5), I decompose $\mathbf{T}_2^*$ as $\mathbf{T}_2^* = \begin{bmatrix} \mathbf{T}_{21}^* & \mathbf{T}_{22}^* \end{bmatrix}$ where $\mathbf{T}_{21}^* \in R^{(N-1)\times(M-1+3)}$ and $\mathbf{T}_{22}^* \in R^{(N-1)\times(M-1-3)}$. This yields

$$rank(\begin{bmatrix} \mathbf{T}_{21}^* & \mathbf{T}_{22}^* \end{bmatrix}) = rank(\mathbf{T}_{21}^*) \leq M-1+3. \tag{4.74}$$

Upon examining Eq. (4.74), I infer the existence of a matrix $\mathbf{W}$ such that

$$\mathbf{T}_{21}^* \mathbf{W} = \mathbf{T}_{22}^*, \tag{4.75}$$

where $\mathbf{W} \in R^{(M-1+3)\times(M-1-3)}$ is an unknown matrix to be estimated in Section 4.4.5.

4.4.4 Linear constraint based on LRPV3

Upon reviewing Eqs. (4.3) and (4.6), I initially define M_N as $min(N-1+3, M-1+3)$ and decompose $\mathbf{T}_3^*$ as $\mathbf{T}_3^* = \begin{bmatrix} \mathbf{T}_{31}^* & \mathbf{T}_{32}^* \end{bmatrix}$ where $\mathbf{T}_{31}^* \in R^{2(M-1)\times M_N}$ and $\mathbf{T}_{32}^* \in R^{2(M-1)\times(2(N-1)-M_N)}$. This gives us

$$rank(\begin{bmatrix} \mathbf{T}_{31}^* & \mathbf{T}_{32}^* \end{bmatrix}) = rank(\mathbf{T}_{31}^*) \leq M_N. \tag{4.76}$$

From Eq. (4.76), I infer the existence of a matrix such that

$$\mathbf{T}_{31}^* \mathbf{Y} = \mathbf{T}_{32}^*, \tag{4.77}$$

where $\mathbf{Y} \in R^{M_N\times(2(N-1)-M_N)}$ is an unknown matrix to be estimated in Section 4.4.5.

4.4.5 Algorithm

The STLS [64, 65, 86] relies solely on LRP for UTIm estimation. This approach often traps the UTIm solution in local minima, thereby limiting both recovery accuracy and convergence rates, especially in noisy environments. To achieve a globally optimal solution for UTIm estimation, three variants of LRP are introduced, namely LRPV1, LRPV2 and LRPV3. These three variants leverage additional low-rank structure information compared to the LRP alone, leading to improved estimation robustness.

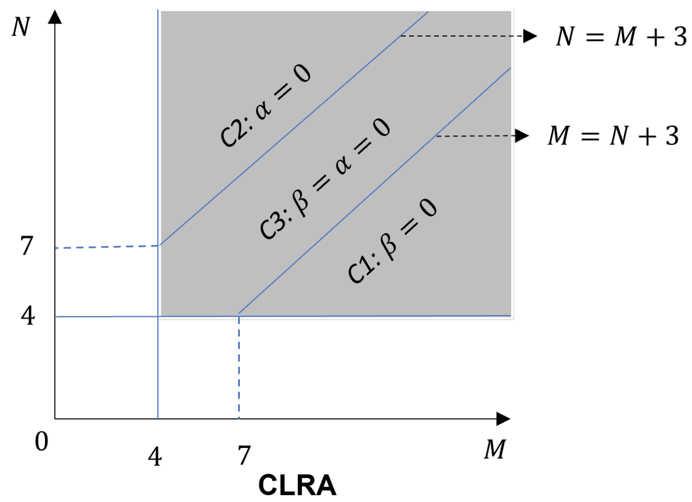


Figure 4.7: Illustration of the combinations of four low-rank properties in the proposed CLRA method with varying numbers of microphones M and sources N (α for LRPV1, β for LRPV2, γ for LRPV3; $\alpha = \beta = \gamma = 0$ in the shadow area for STLS [64, 65, 86]; C1: $\alpha \neq 0$ and $\gamma \neq 0$, C2: $\beta \neq 0$ and $\gamma \neq 0$, C3: $\gamma \neq 0$).

The objective function is formulated as follows:

$$f(\delta_{\mathbf{p}}, \eta_{\mathbf{p}}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}) = \|\mathbf{U}\|_F^2 + \lambda^2 \|(\mathbf{D}_{:,1:3} + \mathbf{U}_{:,1:3})\mathbf{X} - (\mathbf{D}_{:,4:N-1} + \mathbf{U}_{:,4:N-1})\|_F^2 + \alpha^2 \|\mathbf{T}_{\mathbf{11}}^* \mathbf{Z} - \mathbf{T}_{\mathbf{12}}^*\|_F^2 + \beta^2 \|\mathbf{T}_{\mathbf{21}}^* \mathbf{W} - \mathbf{T}_{\mathbf{22}}^*\|_F^2 + \gamma^2 \|\mathbf{T}_{\mathbf{31}}^* \mathbf{Y} - \mathbf{T}_{\mathbf{32}}^*\|_F^2, \tag{4.78}$$

where $\| \bullet \|_F$ denotes the Frobenius norm, $\|\mathbf{U}\|_F^2$ serves as a regularization term [64, 65, 86], and λ, α, β and γ are penalty parameters associated with the respective low-rank properties (LRP, LRPV1, LRPV2, and LRPV3). Notably, when $\gamma = \alpha = \beta = 0$, the proposed CLRA method reduces to STLS, which employs only LRP for UTIm estimation [64, 65, 86].

Additionally, for any given number of microphones M and sources N, I categorize into three cases based on their difference: 1) Case 1 (C1): $M - N > 3$; 2) Case 2 (C2): $N - M > 3$; 3) Case 3 (C3): $|M - N| \leq 3$. From Eqs. (4.4) and (4.5), specifically $M - 1 > N - 1 + 3$ and $N - 1 > M - 1 + 3$, it's evident that LRPV1 and LRPV2 exhibit low-rank properties exclusively in C1 and C2, respectively. Thus, the combination of the four low-rank properties is summarized as follows:

1) $\beta = 0$ in C1;

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

3) $\alpha = 0$ and $\beta = 0$ in C3. Fig. 4.7 illustrates the combination of the four low-rank properties for CLRA.

Subsequently, a comprehensive solution for minimizing the objective function in Eq. (4.78) is outlined. I utilize column-wise matrix vectorization, $v(\cdot)$ (i.e., $v(\mathbf{X}) = [\mathbf{X}_{1,1}, \cdots, \mathbf{X}_{3,1}, \cdots, \mathbf{X}_{1,(N-1-3)}, \cdots, \mathbf{X}_{3,(N-1-3)}]^T$), to define

$$\mathbf{p} = \begin{bmatrix} \delta_{\mathbf{p}}{}^T & \eta_{\mathbf{p}}{}^T & v(\mathbf{X})^T & v(\mathbf{Y})^T & v(\mathbf{Z})^T & v(\mathbf{W})^T \end{bmatrix}^T,$$

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

$$\begin{cases} \mathbf{f_A} = v(\mathbf{U}) \\ \mathbf{f_B} = v((\mathbf{D}_{:,1:3} + \mathbf{U}_{:,1:3})\mathbf{X} - (\mathbf{D}_{:,4:N-1} + \mathbf{U}_{:,4:N-1})) \\ \mathbf{f_C} = v(\mathbf{T}^*_{\mathbf{31}}\mathbf{Y} - \mathbf{T}^*_{\mathbf{32}}) \\ \mathbf{f_D} = v(\mathbf{T}^*_{\mathbf{11}}\mathbf{Z} - \mathbf{T}^*_{\mathbf{12}}) \\ \mathbf{f_E} = v(\mathbf{T}^*_{\mathbf{21}}\mathbf{W} - \mathbf{T}^*_{\mathbf{22}}) \end{cases}. \tag{4.79}$$

The objective function Eq. (4.78) can thus be rewritten as $f(\mathbf{p}) = \|\mathbf{q}\|_2^2$, where the dimension pertaining to 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.

Finally, to solve the nonlinear least squares problem, the Gauss-Newton algorithm [97] is employed. This involves computing the Jacobian matrix

$$\mathbf{J} = \partial \mathbf{q} / \partial \mathbf{p} \in R^{Q \times P}, \tag{4.80}$$

which represents the derivative of vector $\mathbf{q}$ with respect to vector $\mathbf{p}$ (see details in Section 4.4.6). The iterative update of $\mathbf{p}$ proceeds as

$$\mathbf{p}^{m+1} = \mathbf{p}^m - (\mathbf{J}^{mT}\mathbf{J}^m)^{-1}\mathbf{J}^{mT}\mathbf{q}^m, \tag{4.81}$$

where m denotes m^{th} iteration. The overall flowchart of the proposed CLRA method is depicted in Fig. 4.8.

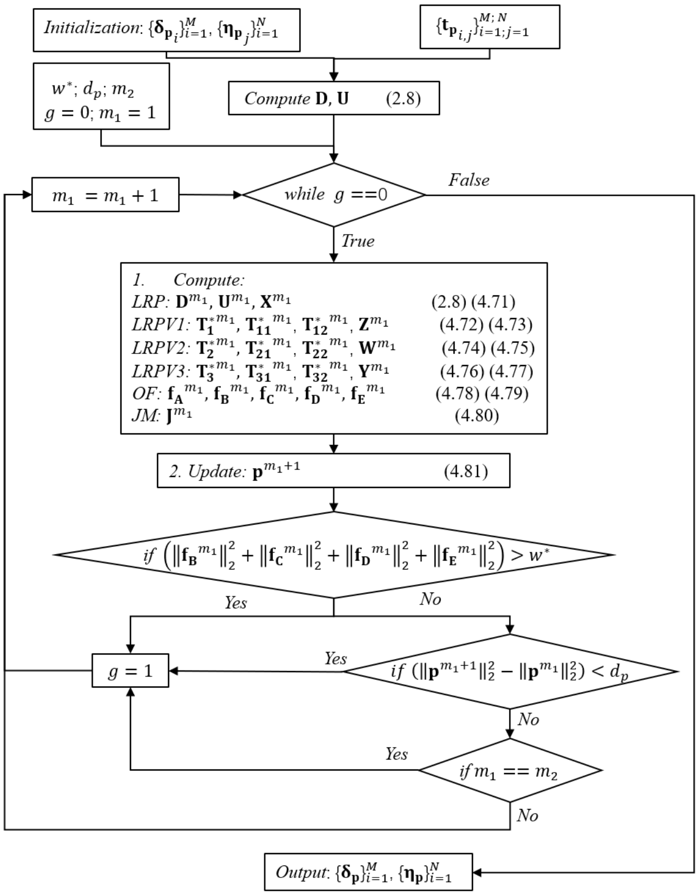


Figure 4.8: The algorithm flowchart pertaining to proposed CLRA method (w^*: threshold for divergence; d_p: stopping threshold for iterations; m_2: maximal number of iterations; m_1: m_1^{th} iteration; OF: Objective function; JM: Jacobian matrix; $\|\bullet\|_2$: L2 norm).

4.4.6 Form of the Jacobian matrix for CLRA Method

In this part, the form of the Jacobian matrix pertaining to the proposed CLRA method in Eq. (4.80) is detailed, thereby, the Jacobian matrix $\mathbf{J} = \frac{\partial \mathbf{q}}{\partial \mathbf{p}}$ can be derived as follows:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{f_A}}{\partial \delta_\mathbf{p}} & \frac{\partial \mathbf{f_A}}{\partial \eta_\mathbf{p}} & \cdots & \frac{\partial \mathbf{f_A}}{\partial \mathbf{w}} \\ \lambda \frac{\partial \mathbf{f_B}}{\partial \delta_\mathbf{p}} & \lambda \frac{\partial \mathbf{f_B}}{\partial \eta_\mathbf{p}} & \cdots & \lambda \frac{\partial \mathbf{f_B}}{\partial \mathbf{w}} \\ \vdots & \vdots & \ddots & \vdots \\ \beta \frac{\partial \mathbf{f_E}}{\partial \delta_\mathbf{p}} & \beta \frac{\partial \mathbf{f_E}}{\partial \eta_\mathbf{p}} & \cdots & \beta \frac{\partial \mathbf{f_E}}{\partial \mathbf{w}} \end{bmatrix}. \tag{4.82}$$

Subsequently, the computation of block matrices in (4.82) is derived as:

$$\begin{cases} \frac{\partial \mathbf{f_A}}{\partial \delta_\mathbf{p}} = \begin{bmatrix} v(\frac{\partial \mathbf{U}}{\partial \delta_{\mathbf{p}_1}}) & \cdots & v(\frac{\partial \mathbf{U}}{\partial \delta_{\mathbf{p}_M}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_A}}{\partial \eta_\mathbf{p}} = \begin{bmatrix} v(\frac{\partial \mathbf{U}}{\partial \eta_{\mathbf{p}_2}}) & \cdots & v(\frac{\partial \mathbf{U}}{\partial \eta_{\mathbf{p}_N}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_A}}{\partial \mathbf{x}} = \begin{bmatrix} v(\frac{\partial \mathbf{U}}{\partial \mathbf{X}_{1,1}}) & \cdots & v(\frac{\partial \mathbf{U}}{\partial \mathbf{X}_{3,(N-1-3)}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_A}}{\partial \mathbf{y}} = \begin{bmatrix} v(\frac{\partial \mathbf{U}}{\partial \mathbf{Y}_{1,1}}) & \cdots & v(\frac{\partial \mathbf{U}}{\partial \mathbf{Y}_{M_N,(2(N-1)-M_N)}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_A}}{\partial \mathbf{z}} = \begin{bmatrix} v(\frac{\partial \mathbf{U}}{\partial \mathbf{Z}_{1,1}}) & \cdots & v(\frac{\partial \mathbf{U}}{\partial \mathbf{Z}_{(N-1+3),(N-1-3)}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_A}}{\partial \mathbf{w}} = \begin{bmatrix} v(\frac{\partial \mathbf{U}}{\partial \mathbf{W}_{1,1}}) & \cdots & v(\frac{\partial \mathbf{U}}{\partial \mathbf{W}_{(M-1+3),(M-1-3)}}) \end{bmatrix} \end{cases}, \tag{4.83}$$

where

$$\frac{\partial \mathbf{U}_{i-1,j-1}}{\partial \delta_{\mathbf{p}_k}} = \begin{cases} -2(\mathbf{t}_{\mathbf{p}_{1,j}} - \mathbf{t}_{\mathbf{p}_{1,1}}) + 2\eta_{\mathbf{p}_j}, & k = 1 \\ (2(\mathbf{t}_{\mathbf{p}_{i,j}} - \mathbf{t}_{\mathbf{p}_{i,1}}) - 2\eta_{\mathbf{p}_j}) \bullet \uparrow_{i,k}, & k = 2, \cdots, M \end{cases}$$

for $\uparrow_{i,k} = \begin{cases} 0, & i \neq k \\ 1, & i = k \end{cases}$ and $i = 2, \cdots, M$ and $j = 2, \cdots, N$;

$$\frac{\partial \mathbf{U}_{i-1,j-1}}{\partial \eta_{\mathbf{p}_k}} = (-2(\mathbf{t}_{\mathbf{p}_{i,j}} - \mathbf{t}_{\mathbf{p}_{1,j}}) + 2(\delta_{\mathbf{p}_1} - \delta_{\mathbf{p}_i})) \bullet \uparrow_{j,k}$$

for $k = 2, \cdots, N$;

$$\frac{\partial \mathbf{U}_{i-1,j-1}}{\partial \mathbf{X}_{k,l}} = 0$$

for $l = 1, \cdots, N-1-3$ and $k = 1, \cdots, 3$;

$$\frac{\partial \mathbf{U}_{i-1,j-1}}{\partial \mathbf{Y}_{k,l}} = 0$$

for $l = 1, \cdots, 2(N-1) - M_N$ and $k = 1, \cdots, M_N$;

$$\frac{\partial \mathbf{U}_{i-1,j-1}}{\partial \mathbf{Z}_{k,l}} = 0$$

for $l = 1, \cdots, N-1-3$ and $k = 1, \cdots, N-1+3$;

$$\frac{\partial \mathbf{U}_{i-1,j-1}}{\partial \mathbf{W}_{k,l}} = 0$$

for $l=1,\cdots,M-1-3$ and $k=1,\cdots,\ M-1+3$.

By denoting $\mathbf{V}=(\mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3})\mathbf{X}-\mathbf{D}_{:,4:N-1}-\mathbf{U}_{:,4:N-1}$, it follows:

$$\begin{cases} \frac{\partial \mathbf{f_B}}{\partial \delta_\mathbf{p}} = \begin{bmatrix} v(\frac{\partial \mathbf{V}}{\partial \delta_{\mathbf{p}_1}}) & \cdots & v(\frac{\partial \mathbf{V}}{\partial \delta_{\mathbf{p}_M}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_B}}{\partial \eta_\mathbf{p}} = \begin{bmatrix} v(\frac{\partial \mathbf{V}}{\partial \eta_{\mathbf{p}_2}}) & \cdots & v(\frac{\partial \mathbf{V}}{\partial \eta_{\mathbf{p}_N}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_B}}{\partial \mathbf{x}} = \begin{bmatrix} v(\frac{\partial \mathbf{V}}{\partial \mathbf{X}_{1,1}}) & \cdots & v(\frac{\partial \mathbf{V}}{\partial \mathbf{X}_{3,(N-1-3)}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_B}}{\partial \mathbf{y}} = \begin{bmatrix} v(\frac{\partial \mathbf{V}}{\partial \mathbf{Y}_{1,1}}) & \cdots & v(\frac{\partial \mathbf{V}}{\partial \mathbf{Y}_{M_N,2(N-1)-M_N}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_B}}{\partial \mathbf{z}} = \begin{bmatrix} v(\frac{\partial \mathbf{V}}{\partial \mathbf{Z}_{1,1}}) & \cdots & v(\frac{\partial \mathbf{V}}{\partial \mathbf{Z}_{N-1+3,N-1-3}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_B}}{\partial \mathbf{w}} = \begin{bmatrix} v(\frac{\partial \mathbf{V}}{\partial \mathbf{W}_{1,1}}) & \cdots & v(\frac{\partial \mathbf{V}}{\partial \mathbf{W}_{M-1+3,M-1-3}}) \end{bmatrix} \end{cases}, \tag{4.84}$$

where

$$\frac{\partial \mathbf{V}_{i-1,j-1}}{\partial \delta_{\mathbf{p}_1}} = \frac{\partial (\mathbf{U}_{:,1:3}\mathbf{X}-\mathbf{U}_{:,4:N-1})_{i-1,j-1}}{\partial \delta_{\mathbf{p}_1}} = \frac{\partial \sum_{u=1}^{3}(\mathbf{U}_{i-1,u}\mathbf{X}_{u,j-1})-\mathbf{U}_{i,j-1+3}}{\partial \delta_{\mathbf{p}_1}}$$
$$=\sum_{u=1}^{3}(2(\eta_{\mathbf{p}_u}-\mathbf{t}_{\mathbf{p}_{i,u+1}}+\mathbf{t}_{\mathbf{p}_{1,1}})\mathbf{X}_{u,j-1})+2(\eta_{\mathbf{p}_{j+3}}-\mathbf{t}_{\mathbf{p}_{i,j+3}}+\mathbf{t}_{\mathbf{p}_{1,1}})$$

for $j=2,\cdots,N-3$ and $i=2,\cdots,M$;

$$\frac{\partial \mathbf{V}_{i-1,j-1}}{\partial \delta_{\mathbf{p}_k}}$$
$$=\frac{\partial (\mathbf{U}_{:,1:3}\mathbf{X}-\mathbf{U}_{:,4:N-1})_{i-1,j-1}}{\partial \delta_{\mathbf{p}_k}} = \uparrow_{i,k} \bullet \{\sum_{u=1}^{3}(2(\mathbf{t}_{\mathbf{p}_{i,u+1}}-\mathbf{t}_{\mathbf{p}_{i,1}}-\eta_{\mathbf{p}_{u+1}})\mathbf{X}_{u,j-1})-2(\mathbf{t}_{\mathbf{p}_{i,j+3}}-\mathbf{t}_{\mathbf{p}_{i,1}}-\eta_{\mathbf{p}_{j+3}})\}$$

for $k=2,\cdots,M$;

$$\frac{\partial V_{i-1,j-1}}{\partial \eta_{\mathbf{p}_k}} = \frac{\partial (\mathbf{U}_{:,1:3}\mathbf{X}-\mathbf{U}_{:,4:N-1})_{i-1,j-1}}{\partial \eta_{\mathbf{p}_k}} = \begin{cases} (-2(\mathbf{t}_{\mathbf{p}_{i,k}}-\mathbf{t}_{\mathbf{p}_{1,k}})+2(\delta_{\mathbf{p}_1}-\delta_{\mathbf{p}_i}))\mathbf{X}_{k-1,j-1} & k=2,\cdots,4 \\ (-2(\mathbf{t}_{\mathbf{p}_{i,j+3}}-\mathbf{t}_{\mathbf{p}_{1,j+3}})+2(\delta_{\mathbf{p}_1}-\delta_{\mathbf{p}_i}))\bullet \uparrow_{k,j+3} & k=5,\cdots,N \end{cases};$$

$$\frac{\partial \mathbf{V}_{i-1,j-1}}{\partial \mathbf{X}_{k-1,l-1}} = \frac{\partial ((\mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3})\mathbf{X})_{i-1,j-1}}{\partial \mathbf{X}_{k-1,l-1}} = \frac{\partial \sum_{u=1}^{3}(\mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3})_{i-1,u}\mathbf{X}_{u,j-1}}{\partial \mathbf{X}_{k-1,l-1}}$$
$$=\uparrow_{l-1,j-1}\bullet(\mathbf{D}_{:,1:3}+\mathbf{U}_{:,1:3})_{i-1,k-1}$$

for $l = 2, \cdots, N-3$ and $k = 2, \cdots, 4$;

$$\frac{\partial \mathbf{V}_{i-1,j-1}}{\partial \mathbf{Y}_{k,l}} = 0$$

for $l = 1, \cdots, 2(N-1) - M_N$ and $k = 1, \cdots, M_N$;

$$\frac{\partial \mathbf{V}_{i-1,j-1}}{\partial \mathbf{Z}_{k,l}} = 0$$

for $l = 1, \cdots, N-1-3$ and $k = 1, \cdots, N-1+3$;

$$\frac{\partial \mathbf{V}_{i-1,j-1}}{\partial \mathbf{W}_{k,l}} = 0$$

for $l = 1, \cdots, M-1-3$ and $k = 1, \cdots, M-1+3$.

Given that

$$\mathbf{T}_3^* = \begin{bmatrix} \mathbf{D} & \mathbf{U} \\ \mathbf{U} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{31}^* & \mathbf{T}_{32}^* \end{bmatrix}$$

and

$$\begin{cases} \frac{\partial \mathbf{D}_{i-1,j-1}}{\partial \delta_{\mathbf{p}_k}} = 0 \\ \frac{\partial \mathbf{D}_{i-1,j-1}}{\partial \eta_{\mathbf{p}_l}} = 0 \end{cases},$$

and $\frac{\partial \mathbf{U}_{i-1,j-1}}{\partial \delta_{\mathbf{p}_k}}$ and $\frac{\partial \mathbf{U}_{i-1,j-1}}{\partial \eta_{\mathbf{p}_l}}$ in Eq. (4.83) ($i = 2, \cdots, M$, $j = 2, \cdots, N$, $k = 1, , \cdots, M$ and $l = 2, \cdots, N$), denote $\mathbf{V1} = \mathbf{T}_{31}^* \mathbf{Y} - \mathbf{T}_{32}^*$, it holds that:

$$\begin{cases} \frac{\partial \mathbf{f_C}}{\partial \delta_{\mathbf{p}}} = \begin{bmatrix} v(\frac{\partial \mathbf{V1}}{\partial \delta_{\mathbf{p}_1}}) & \cdots & v(\frac{\partial \mathbf{V1}}{\partial \delta_{\mathbf{p}_M}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_C}}{\partial \eta_{\mathbf{p}}} = \begin{bmatrix} v(\frac{\partial \mathbf{V1}}{\partial \eta_{\mathbf{p}_2}}) & \cdots & v(\frac{\partial \mathbf{V1}}{\partial \eta_{\mathbf{p}_N}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_C}}{\partial \mathbf{x}} = \begin{bmatrix} v(\frac{\partial \mathbf{V1}}{\partial \mathbf{X}_{1,1}}) & \cdots & v(\frac{\partial \mathbf{V1}}{\partial \mathbf{X}_{3,(N-1-3)}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_C}}{\partial \mathbf{y}} = \begin{bmatrix} v(\frac{\partial \mathbf{V1}}{\partial \mathbf{Y}_{1,1}}) & \cdots & v(\frac{\partial \mathbf{V1}}{\partial \mathbf{Y}_{M_N,2(N-1)-M_N}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_C}}{\partial \mathbf{z}} = \begin{bmatrix} v(\frac{\partial \mathbf{V1}}{\partial \mathbf{Z}_{1,1}}) & \cdots & v(\frac{\partial \mathbf{V1}}{\partial \mathbf{Z}_{N-1+3,N-1-3}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_C}}{\partial \mathbf{w}} = \begin{bmatrix} v(\frac{\partial \mathbf{V1}}{\partial \mathbf{W}_{1,1}}) & \cdots & v(\frac{\partial \mathbf{V1}}{\partial \mathbf{W}_{M-1+3,M-1-3}}) \end{bmatrix} \end{cases}, \tag{4.85}$$

where

$$\frac{\partial \mathbf{V1}_{i-1,j-1}}{\partial \delta_{\mathbf{p}_k}} = \frac{[\mathbf{T}_{31}^* \mathbf{Y} - \mathbf{T}_{32}^*]_{i-1,j-1}}{\partial \delta_{\mathbf{p}_k}} = \sum_{u=1}^{M_N} [\frac{\partial \mathbf{T}_{31}^*}{\partial \delta_{\mathbf{p}_k}}]_{i-1,u} \mathbf{Y}_{u,j-1} - [\frac{\partial \mathbf{T}_{32}^*}{\partial \delta_{\mathbf{p}_k}}]_{i-1,j-1}$$

for $k = 1, \cdots, M$, $i = 2, \cdots, 2M-1$ and $j = 2, \cdots, 2N-1-M_N$;

$$\frac{\partial \mathbf{V1}_{i-1,j-1}}{\partial \eta_{\mathbf{p}_k}} = \frac{\partial [\mathbf{T}_{31}^* \mathbf{Y} - \mathbf{T}_{32}^*]_{i-1,j-1}}{\partial \eta_{\mathbf{p}_k}} = \sum_{u=1}^{M_N} [\frac{\partial \mathbf{T}_{31}^*}{\partial \eta_{\mathbf{p}_k}}]_{i,u} \mathbf{Y}_{u,j} - [\frac{\partial \mathbf{T}_{32}^*}{\partial \eta_{\mathbf{p}_k}}]_{i,j}$$

for $k = 2, \cdots, N$;

$$\frac{\partial \mathbf{V1}_{i-1,j-1}}{\partial \mathbf{X}_{k,l}} = \frac{\partial [\mathbf{T}_{31}^* \mathbf{Y} - \mathbf{T}_{32}^*]_{i-1,j-1}}{\partial \mathbf{X}_{k,l}} = 0$$

for $l = 1, \cdots, N-1-3$, $k = 1, \cdots, 3$ and $l = 1, \cdots, N-1-3$;

$$\frac{\partial \mathbf{V1}_{i-1,j-1}}{\partial \mathbf{Y}_{k-1,l-1}} = \frac{\partial (\mathbf{T}_{31}^* \mathbf{Y} - \mathbf{T}_{32}^*)_{i-1,j-1}}{\partial \mathbf{Y}_{k-1,l-1}} = \uparrow_{l-1,j-1} \bullet (\mathbf{T}_{31}^*)_{i-1,k-1}$$

for $l = 2, \cdots, 2N-1-M_N$ and $k = 2, \cdots, M_N+1$;

$$\frac{\partial \mathbf{V1}_{i-1,j-1}}{\partial \mathbf{Z}_{k,l}} = 0$$

for $l = 1, \cdots, N-1-3$ and $k = 1, \cdots, N-1+3$;

$$\frac{\partial \mathbf{V1}_{i-1,j-1}}{\partial \mathbf{W}_{k,l}} = 0,$$

for $l = 1, \cdots, M-1-3$ and $k = 1, \cdots, M-1+3$.

Given that $\mathbf{T}_1^* = \begin{bmatrix} \mathbf{D} & \mathbf{U} \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{11}^* & \mathbf{T}_{12}^* \end{bmatrix}$, denote $\mathbf{V2} = \mathbf{T}_{11}^* \mathbf{Z} - \mathbf{T}_{12}^*$, it is derived:

$$\begin{cases} \frac{\partial \mathbf{f_D}}{\partial \delta_\mathbf{p}} = \begin{bmatrix} v(\frac{\partial \mathbf{V2}}{\partial \delta_{\mathbf{p}_1}}) & \cdots & v(\frac{\partial \mathbf{V2}}{\partial \delta_{\mathbf{p}_M}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_D}}{\partial \eta_\mathbf{p}} = \begin{bmatrix} v(\frac{\partial \mathbf{V2}}{\partial \eta_{\mathbf{p}_2}}) & \cdots & v(\frac{\partial \mathbf{V2}}{\partial \eta_{\mathbf{p}_N}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_D}}{\partial \mathbf{x}} = \begin{bmatrix} v(\frac{\partial \mathbf{V2}}{\partial \mathbf{X}_{1,1}}) & \cdots & v(\frac{\partial \mathbf{V2}}{\partial \mathbf{X}_{3,(N-1-3)}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_D}}{\partial \mathbf{y}} = \begin{bmatrix} v(\frac{\partial \mathbf{V2}}{\partial \mathbf{Y}_{1,1}}) & \cdots & v(\frac{\partial \mathbf{V2}}{\partial \mathbf{Y}_{M_N,2(N-1)-M_N}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_D}}{\partial \mathbf{z}} = \begin{bmatrix} v(\frac{\partial \mathbf{V2}}{\partial \mathbf{Z}_{1,1}}) & \cdots & v(\frac{\partial \mathbf{V2}}{\partial \mathbf{Z}_{N-1+3,N-1-3}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_D}}{\partial \mathbf{w}} = \begin{bmatrix} v(\frac{\partial \mathbf{V2}}{\partial \mathbf{W}_{1,1}}) & \cdots & v(\frac{\partial \mathbf{V2}}{\partial \mathbf{W}_{M-1+3,M-1-3}}) \end{bmatrix} \end{cases}, \tag{4.86}$$

where

$$\frac{\partial \mathbf{V2}_{i-1,j-1}}{\partial \delta_{\mathbf{p}_k}} = \frac{[\mathbf{T}_{11}^* \mathbf{Z} - \mathbf{T}_{12}^*]_{i-1,j-1}}{\partial \delta_{\mathbf{p}_k}} = \sum_{u=1}^{N-1+3} [\frac{\partial T_{11}^*}{\partial \delta_{\mathbf{p}_k}}]_{i-1,u} \mathbf{Z}_{u,j-1} - [\frac{\partial \mathbf{T_{12}}^*}{\partial \delta_{\mathbf{p}_k}}]_{i-1,j-1}$$

for $k=1,\cdots,M$, $i=2,\cdots,M$ and $j=2,\cdots,N-3$;

$$\frac{\partial \mathbf{V2}_{i-1,j-1}}{\partial \eta_{\mathbf{p}_k}} = \frac{\partial [\mathbf{T}_{\mathbf{11}}^{*}\mathbf{Z}-\mathbf{T}_{\mathbf{12}}^{*}]_{i-1,j-1}}{\partial \eta_{\mathbf{p}_k}} = \sum_{u=1}^{N-1+3} [\frac{\partial \mathbf{T}_{\mathbf{11}}^{*}}{\partial \eta_{\mathbf{p}_k}}]_{i-1,u}\mathbf{Z}_{u,j-1} - [\frac{\partial \mathbf{T_{12}}^{*}}{\partial \eta_{\mathbf{p}_k}}]_{i-1,j-1}$$

for $k=2,\cdots,N$;

$$\frac{\partial \mathbf{V2}_{i-1,j-1}}{\partial \mathbf{X}_{k,l}} = \frac{\partial [\mathbf{T}_{\mathbf{11}}^{*}\mathbf{Z}-\mathbf{T}_{\mathbf{12}}^{*}]_{i-1,j-1}}{\partial \mathbf{X}_{k,l}} = 0$$

for $l=1,\cdots,N-1-3$ and $k=1,\cdots,3$;

$$\frac{\partial \mathbf{V2}_{i-1,j-1}}{\partial \mathbf{Y}_{k,l}} = \frac{\partial [\mathbf{T}_{\mathbf{11}}^{*}\mathbf{Z}-\mathbf{T}_{\mathbf{12}}^{*}]_{i-1,j-1}}{\partial \mathbf{Y}_{k,l}} = 0$$

for $l=1,\cdots,2(N-1)-M_N$ and $k=1,\cdots,M_N$;

$$\frac{\partial \mathbf{V2}_{i-1,j-1}}{\partial \mathbf{Z}_{k-1,l-1}} = \frac{\partial (\mathbf{T}_{\mathbf{11}}^{*}\mathbf{Z}-\mathbf{T}_{\mathbf{12}}^{*})_{i-1,j-1}}{\partial \mathbf{Z}_{k-1,l-1}} = \uparrow_{l-1,j-1} \bullet (\mathbf{T}_{\mathbf{11}}^{*})_{i-1,k-1}$$

for $l=2,\cdots,N-3$ and $k=2,\cdots,N+3$;

$$\frac{\partial \mathbf{V2}_{i,j}}{\partial \mathbf{W}_{k,l}} = \frac{\partial (\mathbf{T}_{\mathbf{21}}^{*}\mathbf{Z}-\mathbf{T}_{\mathbf{22}}^{*})_{i,j}}{\partial \mathbf{W}_{k,l}} = 0$$

for $l=1,\cdots,M-1-3$ and $k=1,\cdots,M-1+3$.
Given that $T_2^{*} = \begin{bmatrix}\mathbf{D}^T & \mathbf{U}^T\end{bmatrix} = \begin{bmatrix}\mathbf{T}_{\mathbf{21}}^{*} & \mathbf{T}_{\mathbf{22}}^{*}\end{bmatrix}$, denote $\mathbf{V3} = \mathbf{T}_{\mathbf{21}}^{*}\mathbf{W} - \mathbf{T}_{\mathbf{22}}^{*}$, it follows:

$$\begin{cases} \frac{\partial \mathbf{f_E}}{\partial \delta_{\mathbf{p}}} = \begin{bmatrix} v(\frac{\partial \mathbf{V3}}{\partial \delta_{\mathbf{p}_1}}) & \cdots & v(\frac{\partial \mathbf{V3}}{\partial \delta_{\mathbf{p}_M}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_E}}{\partial \eta_{\mathbf{p}}} = \begin{bmatrix} v(\frac{\partial \mathbf{V3}}{\partial \eta_{\mathbf{p}_2}}) & \cdots & v(\frac{\partial \mathbf{V3}}{\partial \eta_{\mathbf{p}_N}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_E}}{\partial \mathbf{x}} = \begin{bmatrix} v(\frac{\partial \mathbf{V3}}{\partial \mathbf{X}_{1,1}}) & \cdots & v(\frac{\partial \mathbf{V3}}{\partial \mathbf{X}_{3,(N-1-3)}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_E}}{\partial \mathbf{y}} = \begin{bmatrix} v(\frac{\partial \mathbf{V3}}{\partial \mathbf{Y}_{1,1}}) & \cdots & v(\frac{\partial \mathbf{V3}}{\partial \mathbf{Y}_{M_N,2(N-1)-M_N}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_E}}{\partial \mathbf{z}} = \begin{bmatrix} v(\frac{\partial \mathbf{V3}}{\partial \mathbf{Z}_{1,1}}) & \cdots & v(\frac{\partial \mathbf{V3}}{\partial \mathbf{Z}_{N-1+3,N-1-3}}) \end{bmatrix} \\ \frac{\partial \mathbf{f_E}}{\partial \mathbf{w}} = \begin{bmatrix} v(\frac{\partial \mathbf{V3}}{\partial \mathbf{W}_{1,1}}) & \cdots & v(\frac{\partial \mathbf{V3}}{\partial \mathbf{W}_{M-1+3,M-1-3}}) \end{bmatrix} \end{cases}, \tag{4.87}$$

where

$$\frac{\partial \mathbf{V3}_{i-1,j-1}}{\partial \delta_{\mathbf{p}_k}} = \frac{[\mathbf{T}_{\mathbf{21}}^{*}\mathbf{W}-\mathbf{T}_{\mathbf{22}}^{*}]_{i-1,j-1}}{\partial \delta_{\mathbf{p}_k}} = \sum_{u=1}^{M-1+3} [\frac{\partial \mathbf{T}_{\mathbf{21}}^{*}}{\partial \delta_{\mathbf{p}_k}}]_{i-1,u}\mathbf{W}_{u,j-1} - [\frac{\partial \mathbf{T}_{\mathbf{22}}^{*}}{\partial \delta_{\mathbf{p}_k}}]_{i-1,j-1}$$

for $k = 1, \cdots, M$, $i = 2, \cdots, N$ and $j = 2, \cdots, M-3$;

$$\frac{\partial \mathbf{V3}_{i-1,j-1}}{\partial \eta_{\mathbf{p}_k}} = \frac{\partial [\mathbf{T}^*_{21}\mathbf{W} - \mathbf{T}^*_{22}]_{i-1,j-1}}{\partial \eta_{\mathbf{p}_k}} = \sum_{u=1}^{M-1+3} [\frac{\partial \mathbf{T}^*_{21}}{\partial \eta_{\mathbf{p}_k}}]_{i-1,u}\mathbf{W}_{u,j-1} - [\frac{\partial \mathbf{T}^*_{22}}{\partial \eta_{\mathbf{p}_k}}]_{i-1,j-1}$$

for $k = 2, \cdots, N$;

$$\frac{\partial \mathbf{V3}_{i-1,j-1}}{\partial \mathbf{X}_{k,l}} = \frac{\partial [\mathbf{T}^*_{21}\mathbf{W} - \mathbf{T}^*_{22}]_{i-1,j-1}}{\partial \mathbf{X}_{k,l}} = 0$$

for $l = 1, \cdots, N-1-3$ and $k = 1, \cdots, 3$;

$$\frac{\partial \mathbf{V3}_{i-1,j-1}}{\partial \mathbf{Y}_{k,l}} = \frac{\partial [\mathbf{T}^*_{21}\mathbf{W} - \mathbf{T}^*_{22}]_{i-1,j-1}}{\partial \mathbf{Y}_{k,l}} = 0$$

for $l = 1, \cdots, 2(N-1) - M_N$ and $k = 1, \cdots, M_N$;

$$\frac{\partial \mathbf{V3}_{i-1,j-1}}{\partial \mathbf{Z}_{k,l}} = \frac{\partial (\mathbf{T}^*_{21}\mathbf{W} - \mathbf{T}^*_{22})_{i-1,j-1}}{\partial \mathbf{Z}_{k,l}} = 0$$

for $l = 1, \cdots, N-1-3$ and $k = 1, \cdots, N-1+3$;

$$\frac{\partial \mathbf{V3}_{i-1,j-1}}{\partial \mathbf{W}_{k-1,l-1}} = \frac{\partial (\mathbf{T}^*_{21}\mathbf{W} - \mathbf{T}^*_{22})_{i-1,j-1}}{\partial \mathbf{W}_{k-1,l-1}} = \uparrow_{l-1,j-1} \bullet (\mathbf{T}^*_{21})_{i-1,k-1} \tag{4.88}$$

for $l = 2, \cdots, M-3$ and $k = 2, \cdots, M+3$.

4.5 Experimental Results

The simulation settings are detailed in Section 4.5.1. The impact of parameters from the proposed three variants of LRP on the CLRA method is discussed in Section 4.5.2. Sections 4.5.3, 4.5.4, and 4.5.5 then present the performance comparison of the CLRA method with both STLS and auxiliary function-based algorithms [59]. Robustness analysis is conducted in Section 4.5.6, examining both STLS and CLRA under noise conditions in both simulated and real datasets. Finally, Section 4.5.7 discusses the limitations observed in the proposed variants of LRP and the CLRA method.

4.5.1 Simulation Setup

The simulation data regarding the locations of microphones and sources, as well as their respective start and emission times, are generated randomly using MATLAB R2019a on a computer equipped with a 3.7-GHz CPU, six cores, and 16.0G RAM. Specifically, microphone and source locations are

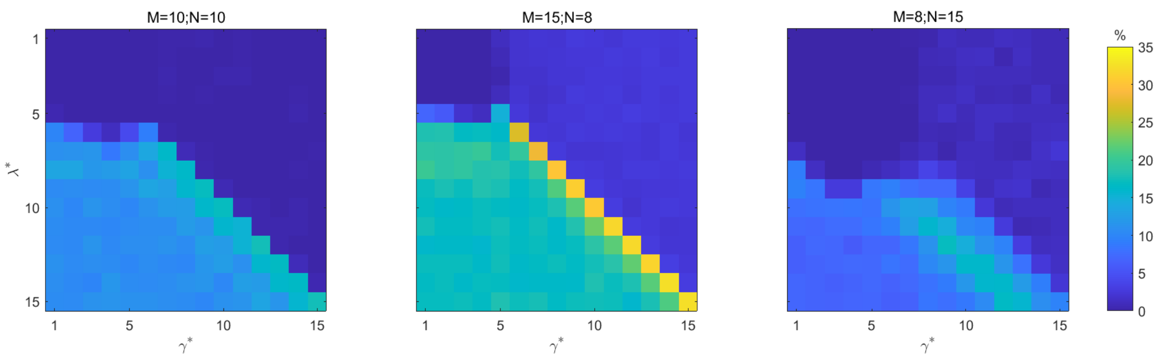


Figure 4.9: The impact of parameters (λ^* for LRP and γ^* for LRPV3) on proposed CLAR1 pertaining to recovery rate under three cases.

uniformly distributed within a $10m \times 10m \times 3m$ room suitable for real-world applications [41, 64]. The start time (δ) and emission time (η) are uniformly generated from the range $\begin{bmatrix}-1 & 1\end{bmatrix}$ s [41, 64]. Additionally, the speed of sound c is set to $340m/s$. For the parameters of the proposed CLRA method (refer to Fig. 4.8), I set $w^* = 10^{30}$, $d_p = 10^{-9}$ and $m_2 = 100$. These parameters dictate that the algorithm terminates if: the objective function value exceeds 10^{30}, the difference in values for variable p in Eq. (4.81) between two consecutive iterations falls below 10^{-9}, or the number of iterations exceeds 100.

To evaluate the robustness of the CLRA method against initialization changes, multiple initializations are employed for each configuration. Typically, a single initialization suffices for microphone start times and source emission times in a specific configuration. However, in cases where random initialization fails to achieve the globally optimal solution for UTIm, multiple initializations are necessary. By counting the number of globally optimal UTIm solutions achieved across multiple random initializations using the CLRA method and comparing these with other state-of-the-art methods, the superior recovery rate performance of the proposed CLRA method can be demonstrated. The recovery rate $Rr(M,N)$ is defined as:

$$Rr(M,N) = \frac{\sum_{i=1}^{Nc(M,N)} Ne_i(M,N)}{I_n(M,N)Nc(M,N)}, \tag{4.89}$$

where $Ne_i(M,N)$ is the count of globally optimal solutions achieved by the algorithms from initializations in the i^{th} configuration, $Nc(M,N)$ is the total number of configurations, and $I_n(M,N)$ is the total number of initializations per configuration. Successful recovery is determined when the errors

(er) between the estimated and ground truth values of UTIm is less than $10^{-4}s$, computed as:

$$er = \frac{\sum_{i=1}^{M} \|\delta_{\mathbf{p}_i} - \delta_{\mathbf{ep}_i}\|}{M} + \frac{\sum_{j=1}^{N} \|\eta_{\mathbf{p}j} - \eta_{\mathbf{ep}j}\|}{N}, \tag{4.90}$$

where $\delta_{\mathbf{p}_i}$ and $\eta_{\mathbf{p}j}$ denote the ground truth values, and $\delta_{\mathbf{ep}_i}$ and $\eta_{\mathbf{ep}j}$ represent the estimated values of UTIm.

Furthermore, the convergence rate $Cr(M,N)$ is defined as

$$\begin{aligned} Cr(M,N) &= \frac{Ce(M,N)}{Nc(M,N)}, \\ Ce(M,N) &= \sum_{i=1}^{Nc(M,N)} sr_i, \\ sr_i &= \begin{cases} 1 & Ne_i(M,N) \neq 0 \\ 0 & Ne_i(M,N) = 0 \end{cases}, \end{aligned} \tag{4.91}$$

where $Cr(M,N)$ indicates the ratio of configurations with successful recovery to the total number of configurations $Nc(M,N)$, and $Ce(M,N)$ counts the total number of successful recovery instances across all configurations (Requirement: er in Eq. (4.90) between the estimated and ground truth values of UTIm is less than $10^{-4}s$). The high convergence rate highlights the CLRA method's robustness against changes in microphone or source configurations, especially when the number of microphones and sources is not sufficient for UTIm.

4.5.2 Parameters Analysis

This subsection examines the parameters of the three proposed variants of LRP within the CLRA framework. I categorize them as follows: CLRA1 when $\alpha = \beta = 0$ (combining LRP with proposed LRPV3), CLRA2 when $\alpha = \gamma = 0$ (combining LRP with proposed LRPV2), and CLRA3 when $\beta = \gamma = 0$ (combining LRP with proposed LRPV1). I set the number of configurations $N_c(M,N) = 10$ configurations $I_n(M,N) = 100$ initializations per configuration, resulting in 1000 implementations for fixed M and N, revealing the impact of parameters on CLRA1, CLRA2, and CLRA3.

Additionally, as illustrated in Fig. 4.7, parameters β and α are zero in C1 and C2, respectively; In C3, both the α and β are zero. Nonetheless, even when both α and β are zero, CLRA behaves differently depending on the number of microphones M and sources N, due to the consistent low-rank property of proposed LRPV3 across C1, C2, and C3. Therefore, I analyze the parameters (λ and γ of CLRA1 (defined in Eq. (4.78)) across scenarios C1, C2, and C3, with configurations set as $M = 15$,

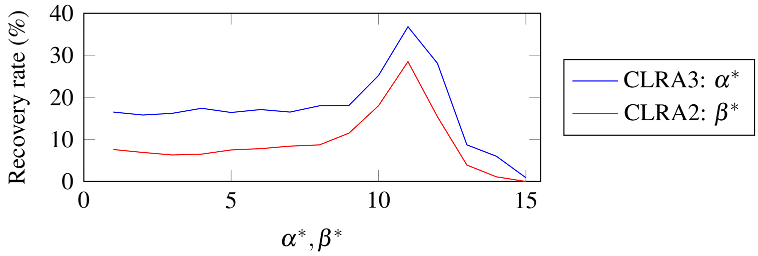


Figure 4.10: The impact of β^* on recovery rate using CLRA2 when $M = 8$, $N = 15$ and α^* on recovery rate using proposed CLRA3 when $M = 15$, $N = 8$.

$N = 8$ for C1, $M = 8$, $N = 15$ for C2 and $M = 10$, $N = 10$ for C3. Following the analysis of these two parameters, I proceed to examine the parameter β for CLRA2 with $M = 15$ and $N = 8$, while keeping $M = 8$ and $N = 15$ constant for the parameter α of CLRA3.

1) Analysis of Parameters λ and γ for CLRA1: I set parameters $\lambda = 10^{\lambda^*}$ and $\gamma = 10^{\gamma^*}$ for CLRA1 in Eq. (4.78), varying λ^* and γ^* from 1 to 15. Fig. 4.9 demonstrates the influence of λ^* and γ^* on CLRA1's recovery rate across different numbers of microphones and sources. In the left and middle sub-figures of Fig. 4.9, both λ^* and γ^* exhibit similar effects on CLRA1's recovery rate: when λ^* and γ^* are small (less than 5), the recovery rate is nearly 0%. As λ^* surpasses 5 and $\gamma^* \leq \lambda^*$, the recovery rate exceeds 0%. Notably, the peaks in recovery rate occur when $\lambda^* = \gamma^*$. From the right sub-figure in Fig. 4.9, it can be observed that when $\lambda^* < 8$, CLRA1's recovery rate is nearly 0%. However, for $\lambda^* > 8$ and $\gamma^* \leq \lambda^* + 2$, the recovery rate surpasses 0%. Moreover, the recovery rate peaks when $\lambda^* = \gamma^* + 3$. Therefore, unless specified otherwise, I use $\lambda^* = 10$ and $\gamma^* = 10$ for both C1 and C3, and $\lambda^* = 12$ and $\gamma^* = 9$ for C2.

2) Analysis of Parameter β for CLRA2: I set parameter $\lambda = 10^{10}$ and analyze the impact of β on CLRA2. Denoting $\beta = 10^{\beta^*}$, β^* is varied from 1 to 15. Fig. 4.10 depicts CLRA2's recovery rate as β^* varies. The plot reveals that β^* significantly affects CLRA2's performance: with small β^* (less than 8), the recovery rate stabilizes around 8%. As β^* reaches 11, CLRA2 achieves a recovery rate of approximately 28%, which then diminishes with further increases in β^*. Thus, $\beta^* = 11$ for CLRA2 is adopted unless specified otherwise.

3) Analysis of Parameter α for CLRA3: Setting $\lambda = 10^{10}$, I analyze the effect of α on CLRA3. Denote $\alpha = 10^{\alpha^*}$ and vary α^* from 1 to 15. Fig. 4.10 indicates that α^* also significantly impacts CLRA3: for $\alpha^* < 9$, CLRA3's recovery rate stabilizes around 17%. However, as α^* continues to increase, the recovery rate rises, reaching about 37% when $\alpha^* = 11$. Hence, $\alpha^* = 11$ is selected for CLRA3 unless specified otherwise.

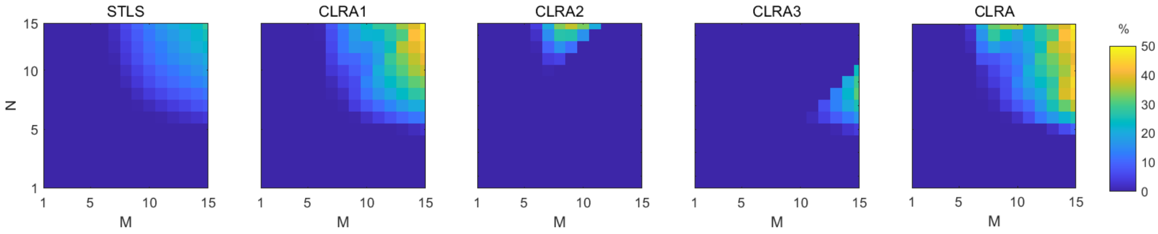
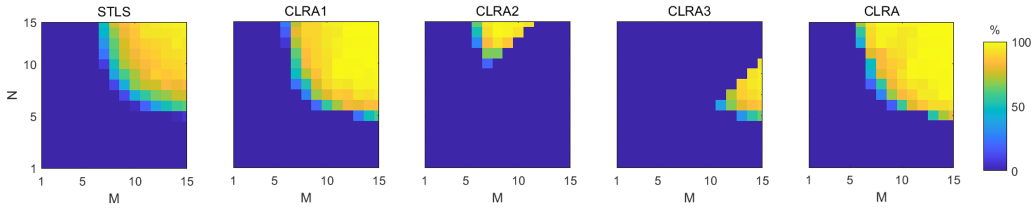
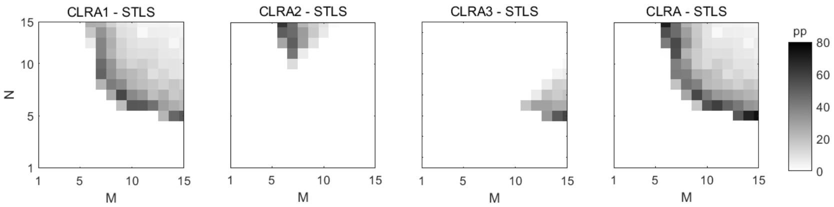


Figure 4.11: Performance comparison of STLS and the proposed CLRA method in terms of recovery rate (CLAR1: $\alpha = \beta = 0$; CLAR2: $\alpha = \gamma = 0$; CLAR3:$\beta = \gamma = 0$).

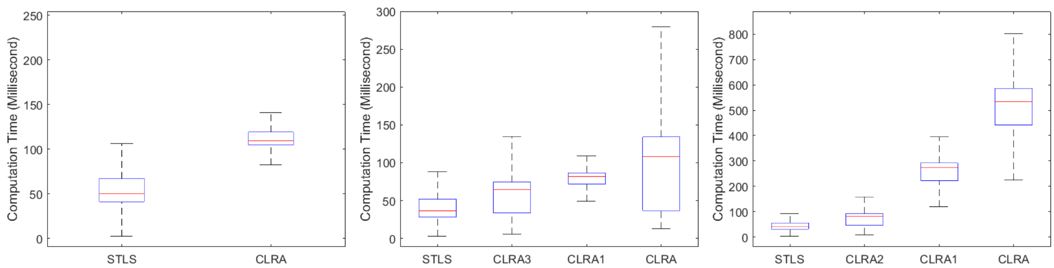


(a) Convergency rate.



(b) Percentage point.

Figure 4.12: The performance comparison for STLS and proposed CLRA in terms of convergency rate (CLAR1: $\alpha = \beta = 0$; CLAR2: $\alpha = \gamma = 0$; CLAR3: $\beta = \gamma = 0$; pp in the colorbar of figure (b) denotes percentage point for the difference of convergency rate between proposed CLRA methods and STLS).



(a) $M = 10$; $N = 10$ (b) $M = 15$; $N = 8$ (c) $M = 8$; $N = 15$

Figure 4.13: Comparison of the running time between the proposed CLRA methods and STLS.

4.5.3 Comparison of the Performance for Low-Rank Properties

This subsection demonstrates the performance of the proposed CLRA methods compared to the STLS method [64].

Upon inspection of the linear constraints in Eq. (4.71) formulated by LRP, the number of equations ($(M-1)(N-1-3)$) should be larger than or equal to the number of unknowns ($3(N-1-3)+M+N-1$), it implies $(M-5)(N-5) \geq 8$, so that the minimal number of microphones and sources for UTIm is $6/7/8$ and $13/9/8$, respectively, and vice versa. 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.

For the STLS method, the parameter λ in Eq. (4.78) is set to 10^{10} ($\gamma = \alpha = \beta = 0$). The parameters for the proposed CLAR methods, which combine one low-rank property with LRP (i.e., CLRA1, CLRA2, and CLRA3), are detailed in Section 4.5.2. For the proposed CLAR method utilizing all four low-rank properties, I categorize them into three cases: 1) C3: The parameters λ and γ in Eq. (4.78) are both set to 10^{10} and 10^{10}, and $\alpha = \beta = 0$. 2) C1: The parameters are set to $\lambda = 10^{10}$, $\gamma = 10^{10}$, $\alpha = 10^{11}$ and $\beta = 0$. 3) C2: The parameters are set to $\lambda = 10^{12}$, $\gamma = 10^{9}$, $\beta = 10^{13}$ and $\alpha = 0$. Furthermore, both M and N are varied from 1 to 15. For each fixed pair of M and N, I randomly configure 200 scenarios ($N_c(M,N) = 200$) and perform 100 initializations ($I_n(M,N) = 100$) for each configuration.

Comparison of Recovery Rate

Fig. 4.11 presents the recovery rate comparison between the proposed CLRA methods and STLS. Firstly, examining the STLS results, it is clear that when $M \leq 6$ or $N \leq 6$, the recovery rate is approximately 0%. However, when both $M \geq 7$ and $N \geq 7$, the recovery rate for STLS ranges from 0% to 28%

Secondly, when comparing the proposed CLRA1 to STLS, it can be observed that with $M \leq 6$ or $N \leq 5$, CLRA1's recovery rate is about 0%, matching STLS. But when $M = 7$ and $N \geq 10$, CLRA1's recovery rate varies from 0% to 8%, outperforming STLS. Notably, when $M \geq 14$ and $N = 6$, CLRA1 achieves about a 1% recovery rate, while STLS remains at 0%. Moreover, for $M > 12$ and $N > 8$, CLRA1's recovery rate exceeds STLS by about 10% to 20%. Therefore, when $M \geq 7$ and $N \geq 6$, CLRA1 performs better than STLS, validating LRPV3 and CLRA1.

Thirdly, comparing CLRA2 to STLS, it can be observed that when $M = 6$ and $N > 13$, CLRA2 achieves a recovery rate of 2% to 4%, compared to STLS's 0%. When $M \geq 7$ and $N > M+3$, CLRA2 significantly outperforms STLS, particularly with $N = 15$ and $7 \leq M \leq 10$, achieving 18% to 27% higher recovery rates. Hence, with $M \geq 6$ and $N > M + 3$, CLRA2 surpasses STLS, confirming the efficacy of LRPV2 and CLRA2.

Fourthly, evaluating CLRA3 against STLS, it can be observed that when $N = 6$ and $M > 13$, CLRA3 achieves a 2% to 4% recovery rate, better than STLS's 0%. When $N \geq 7$ and $M > N + 3$, CLRA3 again outperforms STLS, especially with $M = 15$ and $7 \leq N \leq 10$, achieving 12% to 19% higher recovery rates. Thus, when $N \geq 6$ and $M > N + 3$, CLRA3 outshines STLS, validating LRPV1 and CLRA3.

Finally, comparing the comprehensive CLRA method (LRP combined with LRPV1, LRPV2, and LRPV3) to STLS, it it obvious a significant improvement in recovery rates when $M > 5$ and $N > 5$. Additionally, combining LRP with all three LRP variants yields better recovery rates than combining LRP with only one LRP variant, further validating the effectiveness of LRPV1, LRPV2, and LRPV3.

Comparison of Convergency Rate

Fig. 4.12(a) shows the convergence rate of the proposed CLRA methods compared to STLS, while Fig. 4.12(b) highlights the corresponding percentage points for better illustration. Firstly, analyzing the STLS results in Fig. 4.12(a), it is clear that when $M < 7$ or $N < 6$, the convergence rate is around 0%. As both M and N increase, the convergence rate for STLS improves. Notably, for $M \geq 13$ and $N \geq 12$, the convergence rate exceeds 90%.

Secondly, examining Fig. 4.12(b), I compare the percentage points between CLRA1 and STLS. When $M < 6$ or $N < 6$, the percentage points between CLRA1 and STLS are approximately 0% since both have a 0% convergency rate. For $M > 10$ and $N > 9$, CLRA1's convergence rate is 2% to 18% higher than STLS. Additionally, when $6 \leq M \leq 9$ or $5 \leq N \leq 8$, CLRA1's convergence rate is 6% to 58% higher than STLS, validating the proposed LRPV3 and CLRA1.

Next, I compare CLRA2 with STLS. When $M = 5$, both STLS and CLRA2 have a 0% convergence rate, resulting in 0% percentage points between them. When $M = 6$ and $N > 12$, CLRA2's convergence rate is 28% to 63% higher than STLS. For $M \geq 7$ and $N > M + 3$, CLRA2 consistently outperforms STLS, particularly when $M = 7$ and $N > 10$, with CLRA2's convergence rate being 11% to 50% higher. Therefore, when $M \geq 6$ and $N > M + 3$, CLRA2 shows superior performance, validating LRPV2 and CLRA2.

I then compare CLRA3 with STLS. When $N = 6$ and $M > 12$, CLRA3's convergence rate is 32% to 58% higher than STLS. For $N \geq 7$ and $M > N + 3$, CLRA3 outperforms STLS, CLRA3

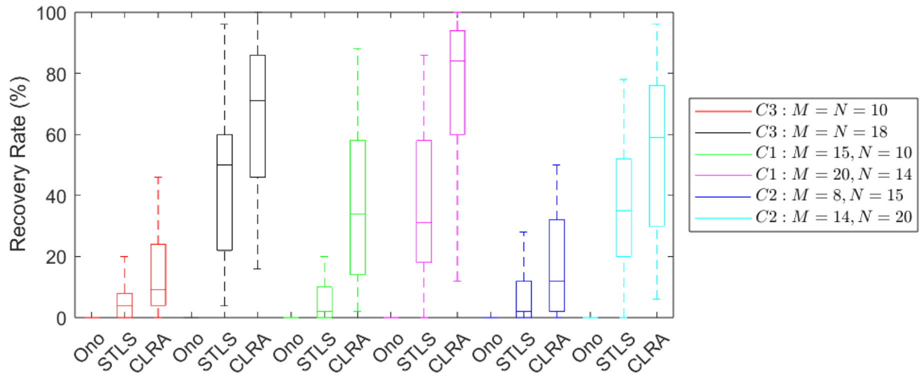


Figure 4.14: Comparison of recovery rates across different configurations using Ono, STLS, and CLRA Methods.

outperforms STLS, particularly when $N = 7$ and $M \geq 11$, with a 32% to 58% higher convergence rate. Additionally, when $N \geq 8$ and $M > N+3$, CLRA3's convergence rate is 14% to 19% higher. Thus, for $N \geq 6$ and $M > N+3$, CLRA3 surpasses STLS, validating LRPV1 and CLRA3.

Finally, comparing the comprehensive CLRA method (LRP combined with LRPV1, LRPV2, and LRPV3) to STLS, it is observed that CLRA achieves a much higher convergence rate for $M > 5$ and $N > 5$. Combining all three rank properties with LRP results in better performance than using just one, further validating LRPV1, LRPV2, and LRPV3.

4.5.4 Computational Complexity Analysis

In this subsection, I analyze the computational complexity of the proposed CLRA method compared to STLS, which uses only LRP [64].

Both STLS and the proposed CLRA method are based on the Gauss-Newton method, where the most computationally intensive part is updating the variables (see Eq. (4.81) and Fig. 4.8). There are three main operations in Eq. (4.81): 1) calculating the multiplication of two Jacobian matrices, $\mathbf{J}^{(m)^T}\mathbf{J}^{(m)}$ (see Section 4.4.6 for the Jacobian matrix form); 2) calculating the inverse of $\mathbf{J}^{(m)^T}\mathbf{J}^{(m)}$, i.e., $(\mathbf{J}^{(m)^T}\mathbf{J}^{(m)})^{-1}$; 3) calculating the multiplication of $(\mathbf{J}^{(m)^T}\mathbf{J}^{(m)})^{-1}$ and $\mathbf{J}^{(m)^T}$.

When only LRP is used for UTIm, three sub-variables in variable $\mathbf{p}$ need to be estimated: $\delta_{\mathbf{p}}$, $\eta_{\mathbf{p}}$ and $\mathbf{X}$. Therefore, the size of the corresponding Jacobian matrix is $(M-1)(2(N-1)-3)$ by $M-3^2+(3+1)(N-1)$. Let $J_{S_{1,r}}$ and $J_{S_{1,c}}$ be $(M-1)(2(N-1)-3)$ and $M-3^2+(3+1)(N-1)$, respectively, resulting in the computational complexity: 1) $O(J_{S_{1,c}}^2 J_{S_{1,r}})$ for the multiplication of two Jacobian matrices; 2) $O(J_{S_{1,c}}^3)$ for the inverse of $\mathbf{J}^{(m)^T}\mathbf{J}^{(m)}$; 3) $O(J_{S_{1,c}}^2 J_{S_{1,r}})$ for the multiplication of

$\left(\mathbf{J}^{(m)^T}\mathbf{J}^{(m)}\right)^{-1}$ and $\mathbf{J}^{(m)^T}$. Thus, the overall computational complexity for using only LRP is

$$O(min(J_{S_{1,c}}^2 J_{S_{1,r}}, J_{S_{1,c}}^3)). \tag{4.92}$$

For the proposed CLRA method, computational complexity is analyzed in three cases based on the number of microphones and sources: C1, C2 and C3. In C3, using only LRP and LRPV3, four variables need to be estimated: $\delta_\mathbf{p}$, $\eta_\mathbf{p}$, $\mathbf{X}$ and $\mathbf{Y}$. The size of the Jacobian matrix is $(M-1)(6(N-1)-3-M_N)$ by $M-3^2-M_N^2+(3+1+2M_N)(N-1)$. Let $J_{S_{1,4,r}}$ and $J_{S_{1,4,c}}$ be $(M-1)(6(N-1)-3-M_N)$ and $M-3^2-M_N^2+(3+1+2M_N)(N-1)$, respectively. The computational complexity using both LRP and LRPV3 is

$$O(min(J_{S_{1,4,c}}^2 J_{S_{1,4,r}}, J_{S_{1,4,c}}^3)). \tag{4.93}$$

In C1, using LRP, LRPV1, and LRPV3, five variables need to be estimated: $\delta_\mathbf{p}$, $\eta_\mathbf{p}$, $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$. The Jacobian matrix size is $(M-1)(7(N-1)-23-M_N)$ by $M-23^2-M_N^2+(3+N+2M_N)(N-1)$. Let $J_{S_{1,2,4,r}}$ and $J_{S_{1,2,4,c}}$ be $(M-1)(7(N-1)-23-M_N)$ and $M-23^2-M_N^2+(3+N+2M_N)(N-1)$, respectively. The computational complexity for using LRP, LRPV1, and LRPV3 is

$$O(min(J_{S_{1,2,4,c}}^2 J_{S_{1,2,4,r}}, J_{S_{1,2,4,c}}^3)). \tag{4.94}$$

In C2, using LRP, LRPV2, and LRPV3, five variables need to be estimated: $\delta_\mathbf{p}$, $\eta_\mathbf{p}$, $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{W}$. The Jacobian matrix size is $(M-1)(7(N-1)-3-M_N)-(N-1)3$ by $M-3^2-M_N^2+(3+1+2M_N)(N-1)+(M-1-3)(M-1+3)$. Let $J_{S_{1,3,4,r}}$ and $J_{S_{1,3,4,c}}$ be $(M-1)(7(N-1)-3-M_N)-(N-1)3$ and $M-3^2-M_N^2+(3+1+2M_N)(N-1)+(M-1-3)(M-1+3)$, respectively. The computational complexity for using LRP, LRPV2, and LRPV3 is

$$O(min(J_{S_{1,3,4,c}}^2 J_{S_{1,3,4,r}}, J_{S_{1,3,4,c}}^3)). \tag{4.95}$$

With those analysis, it can be concluded that the differences in computational complexities from Eqs. (4.92) to (4.95) are:

$$\begin{cases} J_{S_{1,4,r}} - J_{S_{1,r}} = 2(M-1)(2(N-1)-M_N) \\ J_{S_{1,2,4,r}} - J_{S_{1,4,r}} = (M-1)(N-1-3) \\ J_{S_{1,3,4,r}} - J_{S_{1,4,r}} = (N-1)(M-1-3) \end{cases}, \tag{4.96}$$

and

$$\begin{cases} J_{S_{1,4,c}} - J_{S_{1,c}} = M_N(2(N-1) - M_N) \\ J_{S_{1,2,4,c}} - J_{S_{1,4,c}} = (N-1+3)(N-1-3) \\ J_{S_{1,3,4,c}} - J_{S_{1,4,c}} = (M-1+3)(M-1-3) \end{cases}, \tag{4.97}$$

From Eqs. (4.92) to (4.97), it is clear that the computational complexity increases with additional rank properties (LRPV1, LRPV2, or LRPV3) because the size of the Jacobian matrix grows.
From Eqs. (4.92) to (4.97), we see that the computational complexity increases with additional rank properties (LRPV1, LRPV2, or LRPV3) because the size of the Jacobian matrix grows. Given M microphones and N sources, the number of configurations $N_c(M,N)$ is set to 50, and for each configuration, the number of initializations $I_n(M,N)$ is set to 100, totaling 5000 implementations. Fig. 4.13 shows the running time for these 5000 implementations across three cases. Fig. 4.13(a) for C3 with $M = 10$ and $N = 10$, using only LRP and LRPV3. Here, STLS's running time ranges from 3 to 110 milliseconds, and the proposed CLRA method with LRP and LRPV3 takes up to 150 milliseconds. Fig. 4.13(b) for C1 with $M = 15$ and $N = 8$, using LRP, LRPV1, and LRPV3. STLS remains fast, with running times from 3 to 100 milliseconds. CLRA1 and CLRA3 take more time due to the larger Jacobian matrix sizes, with CLRA1 having a larger median value than CLRA3 because of the larger matrix size for LRPV3. The overall running time for CLRA is under 300 milliseconds.
Figure 4.13 illustrates the running time for the 5000 implementations across three cases: C3, C1, and C2. In Fig. 4.13(a), with $M = 10$ and $N = 10$, only LRP and LRPV3 are used in the proposed CLRA method. STLS's running time with only LRP ranges from 3 to 110 milliseconds, indicating high speed. When both LRP and LRPV3 are utilized, the Jacobian matrix size in CLRA increases compared to STLS, leading to longer update times for UTIm. However, it remains efficient, with running times under 150 milliseconds per implementation.
In Fig. 4.13(b), with $M = 15$ and $N = 8$, the proposed CLRA method employs LRP, LRPV1, and LRPV3. The running time of STLS, using only LRP, is again very fast, ranging from 3 to 100 milliseconds for the 5000 implementations. With both LRP and LRPV1 or both LRP and LRPV3, the Jacobian matrix size increases, causing CLRA1 and CLRA3 to take more time than STLS for updating the UTIm. The median running time for CLRA1 is higher than that for CLRA3, as the LRPV3 matrix is larger than the LRPV1 matrix (refer to Eqs. (4.1) and (4.3)). Consequently, CLRA takes longer than both CLRA1 and CLRA3 due to its larger Jacobian matrix. Despite this, the running time remains under 300 milliseconds per implementation.
In Fig. 4.13(c), with $M = 8$ and $N = 15$, LRP, LRPV2, and LRPV3 are used in the proposed CLRA

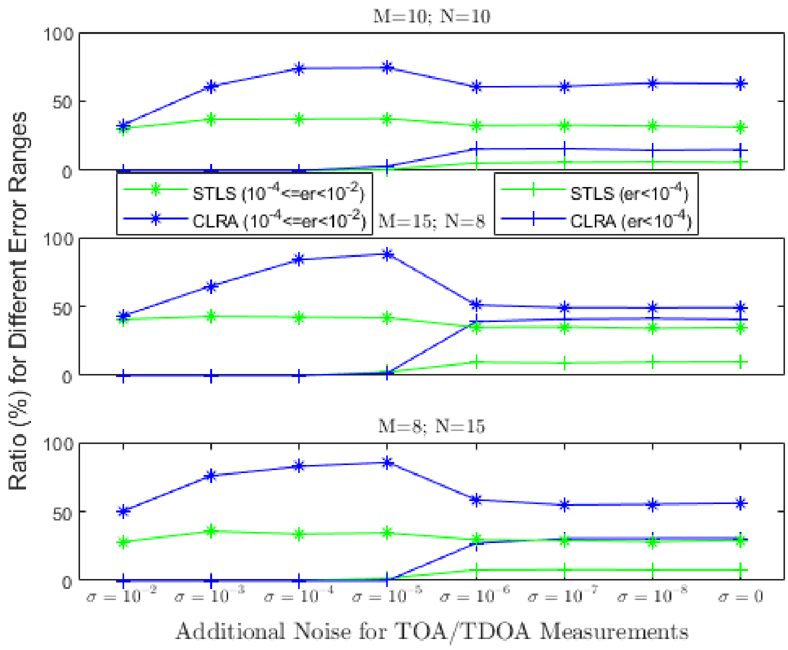


Figure 4.15: Comparison of recovery error ranges for UTIm achieved by STLS and the proposed CLRA, categorized as $er < 10^{-4}s$ and $10^{-4}s \leq er < 10^{-2}s$.

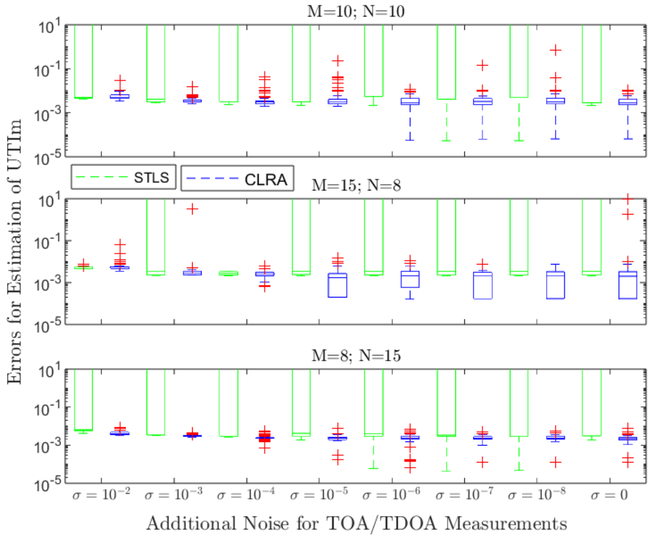
method. The running time pattern in Fig. 4.13(c) is similar to that in Fig. 4.13(b), with all methods taking less than 900 milliseconds per implementation.

4.5.5 Comparison of the Performance with Other Methods

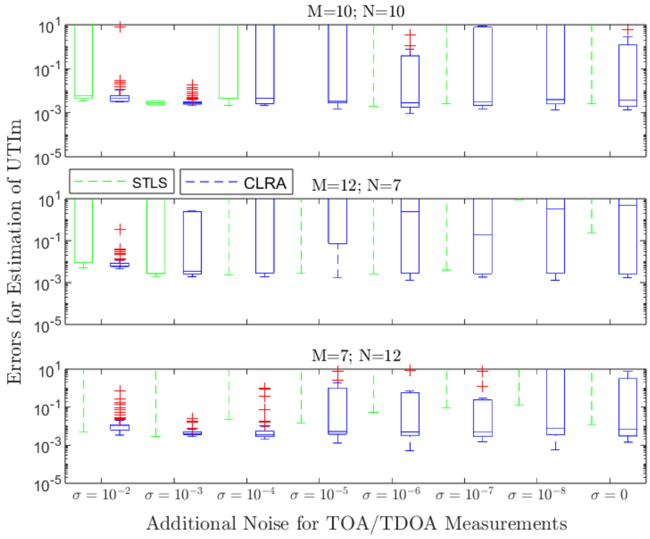
In this subsection, I compare the proposed CLRA method with the auxiliary function-based algorithm Ono [59].

Experiments are conducted for three cases: C1, C2, and C3. Given M microphones and N sources, the number of configurations $N_c(M,N)$ is set to 50, with 50 initializations $I_n(M,N)$ for each configuration, resulting in 2500 implementations for each M and N. The maximum iteration number for each implementation of Ono [59] is set to 2×10^5.

Fig. 4.14 displays the recovery rate within 50 initializations for each configuration. As shown, in all three cases (C3: $M = 10$, $N = 10$ and $M = 18$, $N = 18$; C1: $M = 15$, $N = 8$ and $M = 20$, $N = 14$; C2: $M = 8$, $N = 15$ and $M = 14$, $N = 20$), the recovery rate achieved by the Ono algorithm is consistently zero. In contrast, as both M and N increase, the recovery rates for both STLS and CLRA methods improve. Additionally, the recovery rate for STLS is superior to that of the Ono algorithm. Furthermore, for fixed values of M and N, CLRA demonstrates a significantly higher recovery rate than STLS in terms of minimum, maximum, and median values, reaffirming the effectiveness of the



(a) Realistic Simulation



(b) Office Real Data

Figure 4.16: Estimation errors for UTIm achieved by STLS and proposed CLRA under varying levels of additional noise in TOA/TDOA measurements.

proposed CLRA method.

4.5.6 Robust Analysis

In this subsection, Gaussian noise will be introduced to the TOA/TDOA measurements, with a mean of zero and standard deviations $\sigma = \{10^{-2}, 10^{-3}, \cdots, 10^{-8}\}$, to demonstrate the robustness of the proposed CLRA method.

First, the experimental results using TOA/TDOA measurements as described in Section 4.5.1 are presented. Both the number of configurations $N_c(M,N)$ and the number of initializations $I_n(M,N)$ are set to 50, resulting in 2500 implementations for each M and N. I show the ratio of two different estimation error ranges achieved by STLS and the proposed CLRA: $er < 10^{-4}s$ and $10^{-4}s \leq er < 10^{-2}s$. An error of $er < 10^{-4}s$ corresponds to a distance error of 0.034 m, which is acceptable for localization tasks, while $er < 10^{-2}s$ introduces about 3.4 m of error, significantly impacting localization. These two error ranges illustrate the performance impact of STLS and CLRA under different noise intensities σ.

As shown in Fig. 4.15, when the noise intensity $\sigma > 10^{-6}$, the ratio for $er < 10^{-4}$ achieved by both STLS and CLRA is 0. However, when $\sigma \leq 10^{-6}$, CLRA outperforms STLS significantly, for example, with $M = 15$ and $N = 8$, CLRA achieves a ratio of about 40% for $er < 10^{-4}$, compared to just 6% for STLS. Additionally, the ratio for $10^{-4} \leq er < 10^{-2}$ achieved by CLRA is consistently higher than STLS, indicating that STLS often results in higher errors ($er > 10^{-2}$) compared to CLRA. This verifies the robustness of the proposed CLRA method.

Besides TOA/TDOA simulation data, I also analyze robustness using two other data types:

1. Realistic Simulation: Microphone and source locations are randomly generated within a $5m \times 5m \times 3m$ room, with a 1s chirp signal. The simulation audio signals[1] are generated with a 48k Hz sampling rate and a sound speed of 340 m/s. TOA/TDOA measurements are obtained using the GCC-PHAT method [18]. The mean TOA/TDOA measurement errors are approximately $5 \times 10^{-6}s$ based on the 48k Hz sampling rate.

2. Real Data: Real data collected in a $5m \times 5m \times 3m$ office with 12 microphones at a 96kHz sampling rate. A chirp signal was played from various positions, producing a 12×23 TOA/TDOA matrix[2] [92]. Due to microphone sampling rates and environmental noise, the mean TOA/TDOA measurement errors are around $1 \times 10^{-4}s$, making real applications more challenging than simulations.

For both data types, the number of initializations $I_n(M,N)$ is set to 100. Fig. 4.16 shows the estimation errors er for these data types. As seen in Figs. 4.16(a) and (b), with varying noise intensities σ values achieved by both STLS and CLRA exceed $10^{-4}s$, translating to distance errors over $0.034m$. This is because the noise in TOA/TDOA measurements from realistic simulation and real data is greater than $10^{-6}s$. However, generally, the proposed CLRA method results in lower estimation errors than STLS across different noise levels in both realistic simulation and real data.

[1] https://www.audiolabs-erlangen.de/fau/professor/habets/software/signal-generator

[2] https://github.com/swing-research/xtdoa/tree/master/matlab

These findings indicate that the proposed CLRA has greater potential for real-world applications compared to state-of-the-art methods.

4.5.7 Limitations

The proposed CLRA method incorporates three additional variants of LRP compared to the STLS method, which utilizes only LRP. By integrating these three additional linear constraints formulated by proposed LRP variants, the CLRA method can explore more globally optimal solutions with different initializations, outperforming STLS in both simulation and real data scenarios. However, there are some limitations to the proposed CLRA method:
1) The proposed LRPV1 is constrained by the number of microphones and sources, functioning only when $M-1 > N-1+3$ and $N-1 > 3$.
2) The proposed LRPV2 is similarly limited by the number of microphones and sources, operating only when $N-1 > M-1+3$ and $N-1 > 3$.
3) The CLRA method is unable to denoise TOA/TDOA measurements when environmental noise is present.

4.6 Summary

In this chapter, the primary objective is to synchronize microphones and sources by estimating the UTIm of TOA/TDOA. By constructing matrices $\mathbf{D}$ and $\mathbf{U}$ of Eq. (2.8) in various combinations, three new LRP variants were introduced to exploit the additional low-rank structure information between UTIm and TOA/TDOA. A proof for these three LRP variants was also provided. Utilizing the low-rank structure information from these LRP variants, combined with LRP to constrain the UTIm, I developed the CLRA method for estimating UTIm. Experimental results demonstrated that the proposed CLRA method outperforms state-of-the-art techniques in recovery and convergence rates as well as estimation accuracy, validating the effectiveness of the low-rank structure information exploited by the three LRP variants, enhancing the accuracy of range measurements for self-localization.
Once the microphones and sources are synchronized and the range measurements between them are obtained, the next chapter will address a key challenge in localizing both: relaxing the minimal configuration requirements for the number of microphones and sources necessary for self-localization.

Chapter 5

Numerical Solutions for Relaxing the Minimal Configurations of Joint Microphones and Sources Localization

5.1 Introduction

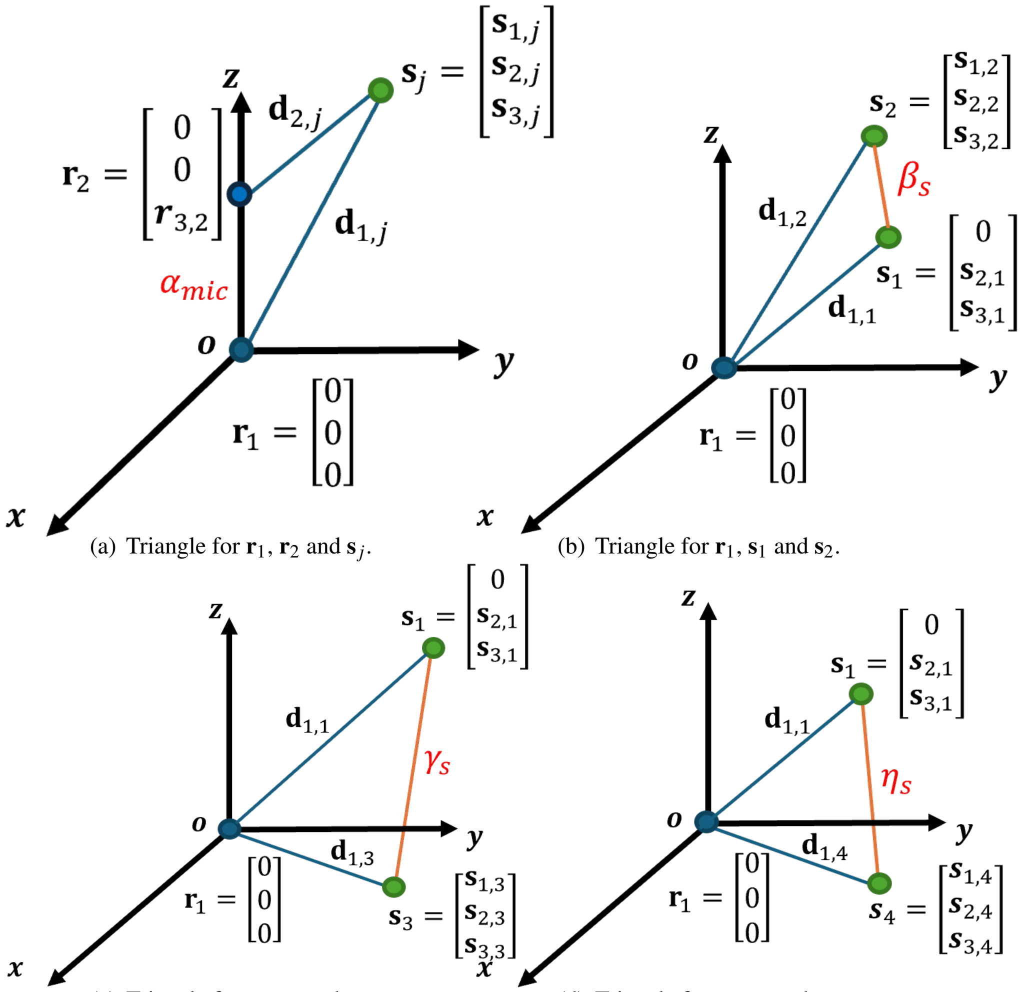
Building in Chapter 4, I proposed a CLAR method comprising three additional variants of the LRP to estimate UTIm in TOA/TDOA, achieving globally optimal solutions for acquiring range measurements between microphones and sources. Alternatively, in scenarios where centralized control of both microphones and sources is feasible, synchronization can be implemented to facilitate the acquisition of range measurements between them. Therefore, this chapter will study localizing both microphones and sources using range measurements between them.

Despite the advancements made in localizing both microphones and sources, state-of-the-arts [33, 82–85] usually necessitate a minimum number of microphones and sources—four/five/six microphones and six/five/four sources. This requirement can limit the efficiency of self-localization when the needed number of microphones and audio sources is not available. Therefore, a critical and substantial question remains: Can we relax the minimal configurations established by current state-of-the-art methods for JMSL? Specifically, consider the challenging scenario where the number of microphones and/or sources is fewer than the minimal configurations proposed in existing literature. Presently, there are neither iterative nor closed-form methods available in the literature that can handle localization under such constraints, severely limiting the efficiency of JMSL. However, if an alternative approach that circumvents the established principles in JMSL regarding the required number of equations and unknown locations is discovered, it might be feasible to further reduce the number of microphones and sources for JMSL applications. Answering this question holds profound implications for the field, not only because it remains largely unexplored but also because it could enhance the flexibility and efficiency of JMSL configurations by enabling the use of fewer devices/microphones and sources.

In this chapter, I address the problem of JMSL in 3D space using LRP and synchronized TOA or range measurements. I introduce a novel numerical method aimed at relaxing the minimal configurations defined by state-of-the-art approaches over the past decades, thereby facilitating the localization of both microphones and sources and enhancing the flexibility of JMSL configurations. By formulating the JMSL problem in terms of triangles and applying the laws of cosine, the localization problem is transformed into estimating four unknown pairs of distances: one pair for microphones and three pairs for sources. This approach reduces the problem to determining four unknown variables that represent the locations of all microphones and sources. Using triangle inequalities, the lower and upper bounds for these four unknown distance pairs are established based on known synchronized TOA or range measurements between microphones and sources. Finally, a numerical optimization method is employed to find optimal solutions within these boundaries, iteratively refining the candidates with a small step size. With proposed numerical method, I demonstrate that the minimal configurations required for JMSL can be achieved with just four microphones and four sources. This relaxation of the minimal configurations defined by state-of-the-art methods represents a significant contribution to the field, suggesting that four microphones and four sources are sufficient for effective JMSL. Therefore, this study marks a substantial advancement and represents a groundbreaking advancement in the realm of JMSL, potentially revolutionizing JSSL techniques and expanding their applicability in more challenging environments.

5.2 Proposed Numerical Method Based On Triangles

In this section, leveraging fundamental properties of triangles such as the laws of cosine and triangle inequality, I introduce a novel numerical method aimed at reducing the required number of microphones and sources for JMSL. This approach not only relaxes the minimal configurations traditionally required for JMSL but also enhances the adaptability of JMSL configurations, thereby facilitating localization tasks in more challenging environments. Next, in Section 5.2.1, I employ the construction of multiple triangles to transform the localization problems of both microphones and



(a) Triangle for $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{s}_j$.

(b) Triangle for $\mathbf{r}_1$, $\mathbf{s}_1$ and $\mathbf{s}_2$.

(c) Triangle for $\mathbf{r}_1$, $\mathbf{s}_1$ and $\mathbf{s}_3$.

(d) Triangle for $\mathbf{r}_1$, $\mathbf{s}_1$ and $\mathbf{s}_4$.

Figure 5.1: Triangles for transforming JMSL to four unknown pairs of distance measurements.

sources, focusing on determining solutions for four unknown distances: one pair for microphones and three pairs for sources. Section 5.2.2 illustrates how these four unknown distances directly influence the estimated locations of microphones and sources. In Sections 5.2.3 and 5.2.4, I detail the process of obtaining numerical solutions for these four unknown distances. Section 5.2.3 establishes lower and upper boundaries for the four unknown distances using triangle inequalities. Section 5.2.4 outlines the method used to iteratively search for optimal solutions within these boundaries. Finally, in Section 5.2.5, I extend the proposed numerical solutions to various scenarios of JMSL, demonstrating the flexibility and robustness of proposed approach in handling different configurations and environmental conditions.

5.2.1 Laws of cosine for transformations of JMSL

Upon examining Eq. (2.14), it becomes evident that

$$\begin{cases}(\mathbf{r}_i - \mathbf{r}_1)^T = \mathbf{U}^*_{\mathbf{p}_{i-1,:}}\mathbf{C}^{-1} \\ -2(\mathbf{s}_j - \mathbf{s}_1) = \mathbf{C}\mathbf{V}^*_{\mathbf{p}_{:,j-1}}\end{cases}, \tag{5.1}$$

where $i = 2, \quad \cdots, \quad M$ and $j = 2, \quad \cdots, \quad N$. Given that $\mathbf{r}_1 = \begin{bmatrix}0, & 0, & 0\end{bmatrix}^T$, $\mathbf{r}_2 = \begin{bmatrix}0, & 0, & \mathbf{r}_{3,2}\end{bmatrix}^T$ and $\mathbf{s}_1 = \begin{bmatrix}0, & \mathbf{s}_{2,1}, & \mathbf{s}_{3,1}\end{bmatrix}^T$, respectively, where $\mathbf{r}_{3,2} > 0$ and $\mathbf{s}_{2,1} > 0$, it is apparent that once solutions for $\mathbf{s}_{2,1}$ and $\mathbf{s}_{3,1}$ and the nine unknown variables in matrix $\mathbf{C}$ are found, localization of all microphones and sources can be achieved. Next, the solutions of these eleven variables are expressed as four unknown pairs of distance measurements related to one pair of microphones and three pairs of sources. Defining α_{mic} as the distance between the 1^{st} and 2^{nd} microphones, β_s as the distance between the 1^{st} and 2^{nd} sources, γ_s as the distance between the 1^{st} and 3^{rd} sources, η_s as the distance between the 1^{st} and 4^{th} sources, we derive the following equation using the law of cosines from Fig. 5.1(a):

$$\alpha_{mic}^2 = \mathbf{d}_{1,j}^2 + \mathbf{d}_{2,j}^2 - 2(\mathbf{r}_1 - \mathbf{s}_j)^T(\mathbf{r}_2 - \mathbf{s}_j), \tag{5.2}$$

where $(\mathbf{r}_1 - \mathbf{s}_j)^T(\mathbf{r}_2 - \mathbf{s}_j) = \|\mathbf{r}_1 - \mathbf{s}_j\|_2\|\mathbf{r}_2 - \mathbf{s}_j\|_2 cos(\theta_{\mathbf{r}_1 - \mathbf{s}_j, \mathbf{r}_2 - \mathbf{s}_j})$ and $\theta_{\mathbf{r}_1 - \mathbf{s}_j, \mathbf{r}_2 - \mathbf{s}_j}$ is the angle between vectors $\mathbf{r}_1 - \mathbf{s}_j$ and $\mathbf{r}_2 - \mathbf{s}_j$.

Upon inspection of Eq. (5.2), with $\mathbf{r}_1 = \begin{bmatrix}0, & 0, & 0\end{bmatrix}^T$, $\mathbf{r}_2 = \begin{bmatrix}0, & 0, & \mathbf{r}_{3,2}\end{bmatrix}^T$ and $(\mathbf{r}_1 - \mathbf{s}_j)^T(\mathbf{r}_1 - \mathbf{s}_j) = \mathbf{d}_{1,j}^2$, where $\mathbf{r}_{3,2} = \alpha_{mic}$ and j ranges from 1 to N, it implies

$$\mathbf{s}_{3,j} = \frac{\alpha_{mic}^2 + \mathbf{d}_{1,j}^2 - \mathbf{d}_{2,j}^2}{2\alpha_{mic}}, \tag{5.3}$$

where $j = 1, \quad \cdots, \quad N$.

Upon inspecting Eq. (5.3) and $(\mathbf{r}_1 - \mathbf{s}_1)^T(\mathbf{r}_1 - \mathbf{s}_1) = \mathbf{d}_{1,1}^2$, since $\mathbf{s}_{1,1} = 0$ and $\mathbf{s}_{2,1} > 0$, it follows

$$\mathbf{s}_{2,1} = \sqrt{\mathbf{d}_{1,1}^2 - \mathbf{s}_{3,1}^2}. \tag{5.4}$$

With Eqs. (5.3) and (5.4), it becomes evident that the two unknown variables $\mathbf{s}_{2,1}$ and $\mathbf{s}_{3,1}$ can be represented by the unknown distance α_{mic} between 1^{st} and 2^{nd} microphones. Therefore, the subsequent task is to obtain solutions for the nine variables in matrix $\mathbf{C}$. From Eq. (5.1), $-2(\mathbf{s}_j - \mathbf{s}_1) =$

$\mathbf{C}\mathbf{V}^*_{\mathbf{p}_{:,j}}$ can be rewritten as

$$\mathbf{C} = -2\begin{bmatrix}\mathbf{s}_2-\mathbf{s}_1 & \mathbf{s}_3-\mathbf{s}_1 & \mathbf{s}_4-\mathbf{s}_1\end{bmatrix}\mathbf{V}^{*-1}_{\mathbf{p}_{:,1:3}}, \tag{5.5}$$

thus the unknown three variables in the third row of matrix $\mathbf{C}$ are

$$\mathbf{C}_{3,:} = -2\begin{bmatrix}\mathbf{s}_{3,2}-\mathbf{s}_{3,1} & \mathbf{s}_{3,3}-\mathbf{s}_{3,1} & \mathbf{s}_{3,4}-\mathbf{s}_{3,1}\end{bmatrix}\mathbf{V}^{*-1}_{\mathbf{p}_{:,1:3}}. \tag{5.6}$$

By applying Eq. (5.3) to Eq. (5.6), it holds that

$$\begin{aligned}\mathbf{C}_{3,:} &= -\frac{\begin{bmatrix}\mathbf{d}^2_{1,2}-\mathbf{d}^2_{2,2}-\mathbf{d}^2_{1,1}+\mathbf{d}^2_{2,1}\\ \mathbf{d}^2_{1,3}-\mathbf{d}^2_{2,3}-\mathbf{d}^2_{1,1}+\mathbf{d}^2_{2,1}\\ \mathbf{d}^2_{1,4}-\mathbf{d}^2_{2,4}-\mathbf{d}^2_{1,1}+\mathbf{d}^2_{2,1}\end{bmatrix}^T\mathbf{V}^{*-1}_{\mathbf{p}_{:,1:3}}}{\alpha_{mic}}\\ &= \frac{\mathbf{D}_{1,1:3}\mathbf{V}^{*-1}_{\mathbf{p}_{:,1:3}}}{\alpha_{mic}} = \frac{\mathbf{U}^*_{\mathbf{p}_{1,:}}}{\alpha_{mic}}.\end{aligned} \tag{5.7}$$

Upon inspection of Eq. (5.7), since matrix $\mathbf{U}^*_{\mathbf{p}}$ is known, it is evident that the three variables in the third row of matrix $\mathbf{C}$ are the functions of unknown α_{mic}. Therefore, by inspecting Eqs. (5.3), (5.4), and (5.7), it can be observed that the five unknown variables $\mathbf{s}_{2,1}$, $\mathbf{s}_{3,1}$ and $\mathbf{C}_{3,:}$ are the functions of α_{mic} only. Next task is to find solutions for the remaining six unknown variables in matrix $\mathbf{C}$.
From Figs. 5.1(b), (c), and (d), applying the laws of cosine to the corresponding three triangles yields:

$$\begin{cases}\beta_s^2 = \mathbf{d}^2_{1,1}+\mathbf{d}^2_{1,2}-2\mathbf{s}_1^T\mathbf{s}_2\\ \gamma_s^2 = \mathbf{d}^2_{1,1}+\mathbf{d}^2_{1,3}-2\mathbf{s}_1^T\mathbf{s}_3\\ \eta_s^2 = \mathbf{d}^2_{1,1}+\mathbf{d}^2_{1,4}-2\mathbf{s}_1^T\mathbf{s}_4\end{cases}. \tag{5.8}$$

Since $\mathbf{s}_j^T\mathbf{s}_j = \mathbf{d}^2_{1,j}$, Eq. (5.8) can be rewritten as:

$$\begin{cases}\beta_s^2 = \mathbf{d}^2_{1,2}-\mathbf{d}^2_{1,1}-2\mathbf{s}_1^T(\mathbf{s}_2-\mathbf{s}_1)\\ \gamma_s^2 = \mathbf{d}^2_{1,3}-\mathbf{d}^2_{1,1}-2\mathbf{s}_1^T(\mathbf{s}_3-\mathbf{s}_1)\\ \eta_s^2 = \mathbf{d}^2_{1,4}-\mathbf{d}^2_{1,1}-2\mathbf{s}_1^T(\mathbf{s}_4-\mathbf{s}_1)\end{cases}. \tag{5.9}$$

Upon inspection of Eq. (5.9), by applying $-2(\mathbf{s}_j-\mathbf{s}_1) = \mathbf{C}\mathbf{V}^*_{\mathbf{p}_{:,j-1}}$ and $\mathbf{s}_1 = \begin{bmatrix}0, & \mathbf{s}_{2,1}, & \mathbf{s}_{3,1}\end{bmatrix}^T$ to Eq.

(5.9), it follows:

$$\begin{cases} \beta_s^2 = \mathbf{d}_{1,2}^2 - \mathbf{d}_{1,1}^2 + \mathbf{s}_{2,1}\mathbf{C}_{2,:}\mathbf{V}^*_{\mathbf{p}_{:,1}} + \mathbf{s}_{3,1}\mathbf{C}_{3,:}\mathbf{V}^*_{\mathbf{p}_{:,1}} \\ \gamma_s^2 = \mathbf{d}_{1,3}^2 - \mathbf{d}_{1,1}^2 + \mathbf{s}_{2,1}\mathbf{C}_{2,:}\mathbf{V}^*_{\mathbf{p}_{:,2}} + \mathbf{s}_{3,1}\mathbf{C}_{3,:}\mathbf{V}^*_{\mathbf{p}_{:,2}} \\ \eta_s^2 = \mathbf{d}_{1,4}^2 - \mathbf{d}_{1,1}^2 + \mathbf{s}_{2,1}\mathbf{C}_{2,:}\mathbf{V}^*_{\mathbf{p}_{:,3}} + \mathbf{s}_{3,1}\mathbf{C}_{3,:}\mathbf{V}^*_{\mathbf{p}_{:,3}} \end{cases}. \tag{5.10}$$

From Eq. (5.10), the three variables in the second row of matrix C can be expressed as:

$$\mathbf{C}_{2,:} = \frac{\begin{bmatrix} \beta_s^2 - \mathbf{d}_{1,2}^2 + \mathbf{d}_{1,1}^2 \\ \gamma_s^2 - \mathbf{d}_{1,3}^2 + \mathbf{d}_{1,1}^2 \\ \eta_s^2 - \mathbf{d}_{1,4}^2 + \mathbf{d}_{1,1}^2 \end{bmatrix}^T \mathbf{V}^{*-1}_{\mathbf{p}_{:,1:3}} - \mathbf{s}_{3,1}\mathbf{C}_{3,:}}{\mathbf{s}_{2,1}}. \tag{5.11}$$

Upon inspection of Eq. (5.11), since $\mathbf{s}_{2,1}$, $\mathbf{s}_{3,1}$ and $\mathbf{C}_{3,:}$ are the functions of α_{mic}, it follows that the solutions of three variables $\mathbf{C}_{2,:}$ depend on four unknown distances α_{mic}, β_s, γ_s and η_s. Therefore, from the solutions of $\mathbf{C}_{2,:}$ in Eq. (5.11), it is evident that when $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{s}_1$ are co-linear, the stability of $\mathbf{C}_{2,:}$ solutions is compromised in the presence of estimation errors for α_{mic}, β_s, γ_s and η_s. However, as previously stated, no three or four positions among microphones and sources should lie on a single line or plane. Therefore, this special case for the solution of $\mathbf{C}_{2,:}$ can be excluded. Finally, once the remaining three variables $\mathbf{C}_{1,:}$ are expressed as the functions of four unknown distances α_{mic}, β_s, γ_s and η_s, all microphone and source locations can be determined based on the four unknown distance measurements.

Considering $\mathbf{s}_j^T\mathbf{s}_j = \mathbf{d}_{1,j}^2$ (see Fig. 5.1(a)), it follows:

$$\begin{cases} \mathbf{d}_{1,2}^2 = \mathbf{s}_{1,2}^2 + \mathbf{s}_{2,2}^2 + \mathbf{s}_{3,2}^2 \\ \mathbf{d}_{1,3}^2 = \mathbf{s}_{1,3}^2 + \mathbf{s}_{2,3}^2 + \mathbf{s}_{3,3}^2 \\ \mathbf{d}_{1,4}^2 = \mathbf{s}_{1,4}^2 + \mathbf{s}_{2,4}^2 + \mathbf{s}_{3,4}^2 \end{cases}. \tag{5.12}$$

To express the three variables in $\mathbf{C}_{1,:}$ as functions of α_{mic}, β_s, γ_s, η_s, Eq. (5.12) is written as:

$$\begin{cases} (\mathbf{s}_{1,2} - \mathbf{s}_{1,1} + \mathbf{s}_{1,1})^2 + (\mathbf{s}_{2,2} - \mathbf{s}_{2,1} + \mathbf{s}_{2,1})^2 = \mathbf{d}_{1,2}^2 - \mathbf{s}_{3,2}^2 \\ (\mathbf{s}_{1,3} - \mathbf{s}_{1,1} + \mathbf{s}_{1,1})^2 + (\mathbf{s}_{2,3} - \mathbf{s}_{2,1} + \mathbf{s}_{2,1})^2 = \mathbf{d}_{1,3}^2 - \mathbf{s}_{3,3}^2 \\ (\mathbf{s}_{1,4} - \mathbf{s}_{1,1} + \mathbf{s}_{1,1})^2 + (\mathbf{s}_{2,4} - \mathbf{s}_{2,1} + \mathbf{s}_{2,1})^2 = \mathbf{d}_{1,4}^2 - \mathbf{s}_{3,4}^2 \end{cases}, \tag{5.13}$$

then applying $-2(\mathbf{s}_j - \mathbf{s}_1) = \mathbf{C}\mathbf{V}^*_{\mathbf{p}_{j-1}}$ and $\mathbf{s}_{1,1} = 0$ to Eq. (5.13), where j ranges from 2 to N, results in:

$$\begin{cases} (\mathbf{C}_{1,:}\mathbf{V}^*_{\mathbf{p}:,1})^2 = 4(\mathbf{d}^2_{1,2} - \mathbf{s}^2_{3,2}) - (\mathbf{C}_{2,:}\mathbf{V}^*_{\mathbf{p}:,1} - 2\mathbf{s}_{2,1})^2 \\ (\mathbf{C}_{1,:}\mathbf{V}^*_{\mathbf{p}:,2})^2 = 4(\mathbf{d}^2_{1,3} - \mathbf{s}^2_{3,3}) - (\mathbf{C}_{2,:}\mathbf{V}^*_{\mathbf{p}:,2} - 2\mathbf{s}_{2,1})^2 \\ (\mathbf{C}_{1,:}\mathbf{V}^*_{\mathbf{p}:,3})^2 = 4(\mathbf{d}^2_{1,4} - \mathbf{s}^2_{3,4}) - (\mathbf{C}_{2,:}\mathbf{V}^*_{\mathbf{p}:,3} - 2\mathbf{s}_{2,1})^2 \end{cases}. \tag{5.14}$$

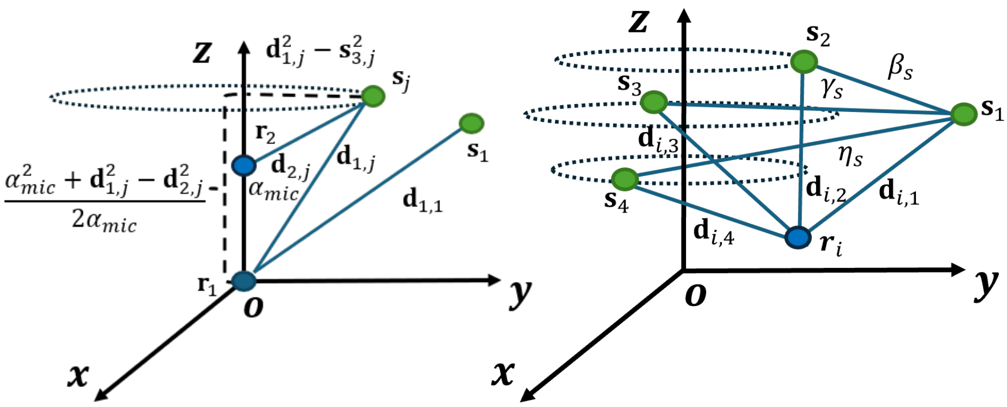
Upon inspection of Eq. (5.14), $C_{1,:}$ can be expressed as

$$\mathbf{C}_{1,:} = \begin{bmatrix} \pm\sqrt{4(\mathbf{d}^2_{1,2} - \mathbf{s}^2_{3,2}) - (\mathbf{C}_{2,:}\mathbf{V}^*_{\mathbf{p}:,1} - 2\mathbf{s}_{2,1})^2} \\ \pm\sqrt{4(\mathbf{d}^2_{1,3} - \mathbf{s}^2_{3,3}) - (\mathbf{C}_{2,:}\mathbf{V}^*_{\mathbf{p}:,2} - 2\mathbf{s}_{2,1})^2} \\ \pm\sqrt{4(\mathbf{d}^2_{1,4} - \mathbf{s}^2_{3,4}) - (\mathbf{C}_{2,:}\mathbf{V}^*_{\mathbf{p}:,3} - 2\mathbf{s}_{2,1})^2} \end{bmatrix}^T \mathbf{V}^{*-1}_{\mathbf{p}:,1:3}. \tag{5.15}$$

From Eq. (5.15), $\mathbf{C}_{1,:}$ is clearly a function of α_{mic}, β_s, γ_s and η_s, considering $\mathbf{s}_{3,j}$, $\mathbf{s}_{2,1}$ as functions of α_{mic} (see Eqs. (5.3) and (5.4)) and the three variables of $\mathbf{C}_{2,:}$ as functions of α_{mic}, β_s, γ_s and η_s (see Eq. (5.11)). Additionally, Eq. (5.15) indicates eight ambiguities in the solutions of $\mathbf{C}_{1,:}$, posing challenges in obtaining a definitive solution. Fortunately, due to the reflection invariance concerning the geometry of microphones and sources, the eight ambiguities in Eq. (5.15) can be reduced to four distinct ambiguities: $\{+,+,+\}$, $\{+,+,-\}$, $\{+,-,+\}$ and $\{+,-,-\}$.
Considering Eqs. (5.3), (5.4), (5.7), it is evident that the five variables $\mathbf{s}_{2,1}$, $\mathbf{s}_{3,1}$ and $\mathbf{C}_{3,:}$, are the function of unknown variable α_{mic}. Moreover, upon inspection of Eqs. (5.11) and (5.15), the six variables in $\mathbf{C}_{1,:}$ and $\mathbf{C}_{2,:}$ are the functions of α_{mic}, β_s, γ_s and η_s. Thus, the eleven variables above depend solely on α_{mic}, β_s, γ_s and η_s, indicating that the locations of all microphones and sources can be determined using only these four unknown variables. Furthermore, while there are four ambiguities in $\mathbf{C}_{1,:}$, these ambiguities can be resolved by comparing the distance errors between the ground truth range measurements $\mathbf{d}_{i,j}$ and the four distance sets $\mathbf{d}^{(se)}_{i,j} = \|\mathbf{r}^{(se)}_i - \mathbf{s}^{(se)}_j\|$ (where $se = 1, \cdots, 4$, $i = 1, \cdots, M$, $j = 1, \cdots, N$), , leading to $min\{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(\mathbf{d}_{i,j} - \mathbf{d}^{(se)}_{i,j})^2}\}$, where $se = 1, \cdots, 4$. . In the subsequent subsection, the impact of the four unknowns α_{mic}, β_s, γ_s and η_s on the locations of microphones of sources is illustrated.

5.2.2 Effects of four unknowns on locations of microphones and sources

Upon examining Eq. (5.3) and Fig. 5.2(a), it is clear that the third coordinate of all sources, $s_{3,j}$, can be determined by the unknown variable α_{mic}. Consequently, the possible solutions for the first and second coordinates of all sources lie on different circles, each of which can be defined by α_{mic} since $\mathbf{s}^2_{1,j} + \mathbf{s}^2_{2,j} = \mathbf{d}^2_{1,j} - \mathbf{s}^2_{3,j}$. Furthermore, the first source's coordinates, $\mathbf{s}_{1,1}$ and $\mathbf{s}_{2,1}$, being zero and



(a) Effect of α_{mic} where $j \geq 2$.

(b) Effects of β_s, γ_s and η_s.

Figure 5.2: Effects of four unknowns on locations of microphones and sources.

greater than zero respectively, indicate that the location of the first source s_1 can be determined solely by α_{mic} (refer to Eqs. (5.3) and (5.4)).

Moreover, as established by Eqs. (5.7), (5.11) and (5.15), the nine variables in matrix C can be dominated by four unknown variables α_{mic}, β_s, γ_s and η_s. This allows for determining the locations of all microphones and sources. Specifically, from Fig. 5.2(b), it is evident that once β_s, γ_s and η_s are known, three sources $\mathbf{s}_2$, $\mathbf{s}_3$ and $\mathbf{s}_4$ can be localized because the distance $\mathbf{d}_{i,j}$ between microphone $\mathbf{r}_i$ and source $\mathbf{s}_j$ is known. Thus, with the position of $\mathbf{s}_1$, $\mathbf{s}_2$, $\mathbf{s}_3$ and $\mathbf{s}_4$, the solutions of eleven variables $\mathbf{s}_{2,1}$ $\mathbf{s}_{3,1}$, $\mathbf{C}$ can be obtained, resulting in the solutions for localizing all microphones and sources (see Eq. (5.1)). Hence, estimating the four unknowns α_{mic}, β_s, γ_s and η_s is crucial for JMSL. Unfortunately, there is no prior information available regarding these unknowns, making their estimation challenging. Particularly, when there are only four microphones and four sources, the number of valid equations from $\mathbf{d}_{i,j} = \|\mathbf{r}_i - \mathbf{s}_j\|$ is only two (as analyzed in Section 5.2.4), making it impossible to solve for the four unknowns. This leads to the conclusion that the minimal configuration requires at least six/five/four microphones and four/five/six sources, respectively. However, if an alternative method can be found to bypass the principle of the number of valid equations (synchronized TOA/range/distance measurement) and unknown locations, numerical solutions for the four unknowns may still be possible even with only four microphones and four sources for JMSL. Therefore, the next subsection will derive the boundaries of these four unknowns.

5.2.3 Triangle inequality for lower and upper boundaries of four unknowns

In this subsection, before determining the solutions of four unknown variables α_{mic}, β_s, γ_s and η_s, I will establish both the lower and upper boundaries for these variables using the triangle inequality. The triangle inequality states that in any triangle, the length of one side must be greater than or equal to the absolute difference between the lengths of the other two sides, and less than or equal to their sum. Specifically, for a triangle with side length a_1, a_2 and a_3, respectively, it holds that $|a_2 - a_3| \leq a_1 \leq |a_2 + a_3|$, where $|\bullet|$ denotes the absolute value. Therefore, from Figs. 5.2(a) and (b), it can be derived that the lower and upper boundaries for the four unknown variables α_{mic}, β_s, γ_s and η_s:

$$\begin{cases} max\{|\mathbf{d}_{1,j} - \mathbf{d}_{2,j}|\} \leq \alpha_{mic} \leq min\{|\mathbf{d}_{1,j} + \mathbf{d}_{2,j}|\} \\ max\{|\mathbf{d}_{i,1} - \mathbf{d}_{i,2}\} \leq \beta_s \leq min\{|\mathbf{d}_{i,1} + \mathbf{d}_{i,2}|\} \\ max\{|\mathbf{d}_{i,1} - \mathbf{d}_{i,3}\} \leq \gamma_s \leq min\{|\mathbf{d}_{i,1} + \mathbf{d}_{i,3}|\} \\ max\{|\mathbf{d}_{i,1} - \mathbf{d}_{i,4}\} \leq \eta_s \leq min\{|\mathbf{d}_{i,1} + \mathbf{d}_{i,4}|\} \end{cases}, \tag{5.16}$$

where i ranges from 1 to M and j ranges from 1 to N. In the following subsection, I will explore an alternative numerical method to obtain numerical solutions for these four unknowns within the boundaries defined in Eq. (5.16), even when the number of both microphones and sources is limited to four.

5.2.4 Numerical solutions for JMSL

In this subsection, I derive numerical solutions for the four unknown variables α_{mic}, β_s, γ_s and η_s using the boundaries established in Eq. (5.16). Before estimating the numerical solutions, I will first demonstrate the number of valid equations needed to clarify the solutions provided by the proposed numerical method.

Given that the distance $\mathbf{d}_{i,j}$ between i^{th} microphone and j^{th} source is formulated as $||\mathbf{r}_i - \mathbf{s}_j||_2$: $\mathbf{d}_{i,j} = ||\mathbf{r}_i - \mathbf{s}_j||_2$ ($i = 1, \ \cdots, \ M$ and $j = 1, \ \cdots, \ N$) can be categorized into three cases

$$\begin{cases} \mathbf{d}_{1,j}^2 = \mathbf{r}_1^T\mathbf{r}_1 + \mathbf{s}_j^T\mathbf{s}_j - 2\mathbf{r}_1^T\mathbf{s}_j \quad j = 1, \cdots, N \\ \mathbf{d}_{i,1}^2 = \mathbf{r}_i^T\mathbf{r}_i + \mathbf{s}_1^T\mathbf{s}_1 - 2\mathbf{r}_i^T\mathbf{s}_1 \quad i = 1, \cdots, M \\ \mathbf{d}_{i,j}^2 = \mathbf{r}_i^T\mathbf{r}_i + \mathbf{s}_j^T\mathbf{s}_j - 2\mathbf{r}_i^T\mathbf{s}_j \quad i = 2, \cdots, M; j = 2, \cdots, N \end{cases}. \tag{5.17}$$

From Eq. (5.17), since $\mathbf{r}_1 = \begin{bmatrix} 0, & 0, & 0 \end{bmatrix}^T$, it follows that $\mathbf{d}_{1,j}^2 = \mathbf{s}_j^T\mathbf{s}_j$ for $j = 1, \cdots, N$. Therefore, with $\mathbf{d}_{1,1}^2 = \mathbf{s}_1^T\mathbf{s}_1$, it can be derived that $\mathbf{d}_{i,1}^2 = \mathbf{r}_i^T\mathbf{r}_i + \mathbf{d}_{1,1}^2 - 2\mathbf{r}_i^T\mathbf{s}_1$ for $i = 1,\cdots,M$. Similarity, with $\mathbf{d}_{1,j}^2 = \mathbf{s}_j^T\mathbf{s}_j$ and $-2\mathbf{r}_i^T(\mathbf{s}_j - \mathbf{s}_1) = \mathbf{U}^*_{\mathbf{p}_{i-1,:}}\mathbf{V}^*_{\mathbf{p}_{:,j-1}} = \mathbf{d}_{i,j}^2 - \mathbf{d}_{i,1}^2 - \mathbf{d}_{1,j}^2 + \mathbf{d}_{1,1}^2$ (see Eqs. (2.11), (2.14) and (5.1)), it holds that $\mathbf{d}_{i,j}^2 = \mathbf{r}_i^T\mathbf{r}_i + \mathbf{s}_j^T\mathbf{s}_j - 2\mathbf{r}_i^T\mathbf{s}_j$ ($i = 2,\cdots,M$; $j = 2,\cdots,N$) in Eq. (5.17) as

$$\mathbf{d}_{i,1}^2 - \mathbf{d}_{1,1}^2 = \mathbf{r}_i^T\mathbf{r}_i - 2\mathbf{r}_i^T\mathbf{s}_1, \tag{5.18}$$

which matches the variant of $\mathbf{d}_{i,1}^2 = \mathbf{r}_i^T\mathbf{r}_i + \mathbf{s}_1^T\mathbf{s}_1 - 2\mathbf{r}_i^T\mathbf{s}_1$ in Eq. (5.17) for $i = 2,\cdots,M$. Therefore, Eq. (5.17) can be written as:

$$\begin{cases} \mathbf{d}_{1,j}^2 = \mathbf{s}_j^T\mathbf{s}_j \quad j = 1,\cdots,N \\ \mathbf{d}_{i,1}^2 - \mathbf{d}_{1,1}^2 = \mathbf{r}_i^T\mathbf{r}_i - 2\mathbf{r}_i^T\mathbf{s}_1 \quad i = 1,\cdots,M \end{cases}. \tag{5.19}$$

Next, I will show the valid equations in Eq. (5.19) for obtaining the numerical solutions for the four unknowns. Since $\mathbf{r}_1 = \begin{bmatrix} 0, & 0, & 0 \end{bmatrix}^T$, the equation $\mathbf{d}_{i,1}^2 - \mathbf{d}_{1,1}^2 = \mathbf{r}_i^T\mathbf{r}_i - 2\mathbf{r}_i^T\mathbf{s}_1$ is invalid when $i = 1$. Additionally, when $i = 2$, $\mathbf{d}_{i,1}^2 - \mathbf{d}_{1,1}^2 = \mathbf{r}_i^T\mathbf{r}_i - 2\mathbf{r}_i^T\mathbf{s}_1$ is also invalid as it is the same as the equation $\alpha_{mic}^2 = \mathbf{d}_{1,j}^2 + \mathbf{d}_{2,j}^2 - 2(\mathbf{r}_1 - \mathbf{s}_j)^T(\mathbf{r}_2 - \mathbf{s}_j)$ in Eq. (5.2) for $j = 1$. Therefore, there are just $M-2$ valid equations for $\mathbf{d}_{i,1}^2 - \mathbf{d}_{1,1}^2 = \mathbf{r}_i^T\mathbf{r}_i - 2\mathbf{r}_i^T\mathbf{s}_1$. Next, I shall present the number of valid equations for $\mathbf{d}_{1,j}^2 = \mathbf{s}_j^T\mathbf{s}_j$ in Eq. (5.19), where $j = 1,\cdots,N$. When $j = 1$, $\mathbf{d}_{1,1}^2 = \mathbf{s}_1^T\mathbf{s}_1$ is invalid since as it is already used to formulate variables $\mathbf{s}_{2,1}$ and $\mathbf{s}_{3,1}$ (see Eqs. (5.3) and (5.4)). When $j = 2$, 3 and 4, $\mathbf{d}_{1,j}^2 = \mathbf{s}_j^T\mathbf{s}_j$ are also invalid as they are used to formulate the three variables $\mathbf{C}_{1,:}$ (see Eqs. (5.12), (5.13), (5.14) and (5.15)). Thus, there are only $N-4$ valid equations for $\mathbf{d}_{1,j}^2 = \mathbf{s}_j^T\mathbf{s}_j$, where $j = 1,\cdots,N$. From this analysis, it is clear that the number of valid equations for solving the four unknowns is only $M+N-6$ with synchronized TOA measurements.

Given that the number of valid equations is $M+N-6$ in Eq. (5.19), , it is evident that obtaining either closed-form solutions or iterative solutions for the four unknowns using optimization methods is impossible when 1) $M = 4$ and $N = 4$; 2) $M = 4$ and $N = 5$ and 3) $M = 5$ and $N = 4$. Fortunately, using the boundaries derived in Eq. (5.16), obtaining numerical solutions for the four unknowns becomes possible. In more detail, first, the boundaries of the four unknowns can be divided into several candidate values with a given small step. Second, with these candidate values, several sets of locations for the microphones and sources are obtained. Finally, the numerical solutions for JMSL can be found by selecting the sets of microphone and source locations that minimize the error with the valid equations in Eq. (5.19):

$$Er = \sqrt{\sum_{i=3}^{M}(\mathbf{d}_{i,1}^2 - \mathbf{d}_{1,1}^2 - \mathbf{r}_i^T\mathbf{r}_i + 2\mathbf{r}_i^T\mathbf{s}_1)^2} + \sqrt{\sum_{j=5}^{N}(\mathbf{d}_{1,j}^2 - \mathbf{s}_j^T\mathbf{s}_j)^2}. \tag{5.20}$$

Algorithm: **Numerical solutions for JMSL**

Input: 1. M and N;
2. Range measurements between microphones and sources.

Output: Locations for M microphones and N sources.

Step 1: Obtain the lower and upper boundaries for α_{mic}, β_s, γ_s and η_s with Eq.(5.16).
Step 2: Exclude the candidates of α_{mic}, β_s, γ_s and η_s in **Step 1**.
Step 2.1: Exclude the candidates of α_{mic} with $\mathbf{s}_{2,1} > 0$ in Eq. (5.4).
Step 2.2: Exclude the candidates of β_s, γ_s and η_s with Eq. (5.14): $\mathbf{C}_{1,:}\mathbf{V}^*_{\mathbf{p}:,1} \geq 0$, $\mathbf{C}_{1,:}\mathbf{V}_{:,2} \geq 0$ and $\mathbf{C}_{1,:}\mathbf{V}^*_{\mathbf{p}:,3} \geq 0$.
Step 3: Test remaining candidates for α_{mic}, β_s, γ_s and η_s:
Step 3.1: Obtain values for $\mathbf{s}_{2,1}$, $\mathbf{s}_{3,j}$, $\mathbf{C}_{3,:}$, $\mathbf{C}_{2,:}$ and $\mathbf{C}_{1,:}$ with Eqs. (5.3), (5.4), (5.7), (5.11) (5.15);
Step 3.2: Obtain locations for all microphones and source with Eqs. (5.1);
Step 3.3: Calculate the Er in Eq. (5.20).
Step 4: Choose the optimal value for α_{mic}, β_s, γ_s and η_s with minimal value of Er in **Step 3**.
Step 5: Obtain the locations for all microphone and sources using **Step 3.1** and **Step 3.2**.

Note that when $N < 5$, the second term on the right-hand side of Eq. (5.20) should be discarded. The pseudo code for localizing both microphones and sources with the four unknowns α_{mic}, β_s, γ_s and η_s is presented in the *Algorithm*.

5.2.5 Extensions of proposed numerical method

To validate the proposed numerical method for confirming the locations of microphones and sources using four unknowns, and to facilitate the applications of JMSL, I extend the method to five different scenarios:

1) One co-located microphone and source and one known distance between a pair of microphones: In this scenario, an additional microphone $\mathbf{r_a}$ is co-located with the 1^{st} source s_1. This allows us to solve for three variables: β_s, γ_s and η_s , based on the known distances between the three sources ($\mathbf{s}_1$, $\mathbf{s}_2$ and $\mathbf{s}_3$) and the additional microphone $\mathbf{r_a}$. Additionally, knowing the distance between the 1^{st} microphone $\mathbf{r}_1$ and 2^{nd} microphone $\mathbf{r}_2$ allows us to determine α_{mic}. Consequently, the locations of all microphones and sources can be determined with closed-form solutions using the proposed method in this chapter.

2) On co-located microphone and sources: Here, an additional microphone $\mathbf{r_a}$ is co-located with the 1^{st} source $\mathbf{s}_1$ [23]. This setup allows us to solve for β_s, γ_s and η_s, leaving only α_{mic} as the unknown variable. The task of JMSL is to estimate the optimal numerical solution for α_{mic} within the boundaries defined in Eq. (5.16) using the proposed ***Algorithm*** in Section 5.2.4.

3) Two known distances for one pair of microphones and one pair of sources: In this scenario, the

distances between two microphones and two sources are known. Assuming these known distances are between the 1^{st} and 2^{nd} microphones and 1^{st} and 2^{nd} sources, the solutions of α_{mic} and β_s are determined. Therefore, the task of JMSL is to find the optimal numerical solutions for γ_s and η_s within the boundaries in Eq. (5.16) using the proposed ***Algorithm***.

4) One known distance between any two microphones: Here, the distance between any two microphones is known. Assuming the known distance is between the 1^{st} and 2^{nd} microphones, the solution for α_{mic} is obtained. The task of JMSL is then to find the optimal numerical solutions for β_s, γ_s and η_s within the boundaries in Eq. (5.16). In this setup, the number of known distances between microphones and sources is MN and the number of unknowns is $3(M+N)-7$ as $\alpha_{mic} = r_{3,2}$. Using the principle that the number of equations should be at least equal to the number of unknowns, $(M-3)(N-3) \geq 2$, it implies that configurations with four microphones and four sources are impossible for JMSL. However, proposed numerical method in the ***Algorithm*** demonstrates that estimating the three unknowns β_s, γ_s and η_s makes it possible to localize four microphones and four sources, relaxing the minimal configuration requirements of traditional methods and facilitating JMSL when the number of microphones and/or sources is limited.

5) No prior information: In this scenario, there are no known distance measurements between any pairs of microphones or sources. The task of JMSL is to estimate the optimal solutions for the four unknowns α_{mic}, β_s, γ_s and η_s within the boundaries in Eq. (5.16). Here, the number of known range measurements between microphones and sources is MN and the number of unknowns is $3(M+N)-6$, indicating that configurations with 1) $M=4$ and $N=4$; 2) $M=4$ and $N=5$; 3) $M=5$ and $N=4$ are impossible for localizing both microphones and sources. However, using proposed numerical method in the ***Algorithm***, the solutions for the four unknowns can be estimated even in these minimal configurations, reducing the number of required microphones and/or sources compared to traditional methods and relaxing the minimal configuration constraints of past state-of-the-art techniques.

5.3 Synthetic Experiments and Evaluations

In Section 5.3.1, the experimental settings are discussed. Following that, Section 5.3.2 presents the results of the proposed numerical method. Additionally, I demonstrate the robustness of proposed numerical method by adding Gaussian noise with zero mean and standard deviations σ of $\{10^{-6}, 10^{-4}, 10^{-3}, 10^{-2}\}$ [33] meters to both the range measurements $d_{i,j}$ and the four unknown variables, as described in [33]. Finally, Section 5.3.3 illustrates the limitations of the proposed

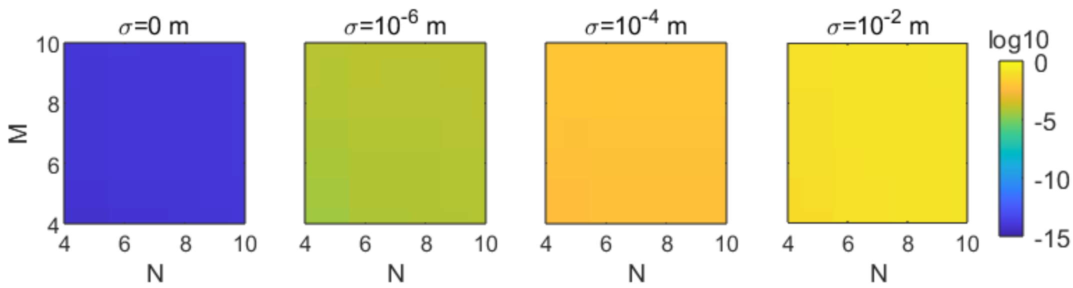


Figure 5.3: Results for proposed numerical method: scenario with one co-located microphone and source and one known distance between a pair of microphones (where log 10 represents $\log_{10}$ transformation; known values of α_{mic}, β_s, γ_s and η_s).

numerical method.

5.3.1 Setup

Simulation data:

The locations of microphones and sources are randomly generated in MATLAB using a uniform distribution within a unit cubic room of size $1 \times 1 \times 1\ m^3$ [33] with the speed of sound 340 m/s [33]. Consequently, the distance measurements between the microphones and sources can be derived from their respective locations.

Evaluation metric:

To evaluate and validate the proposed numerical method, I compare the localization errors between the ground truth and the estimated locations of microphones and sources. The error metric is defined as:

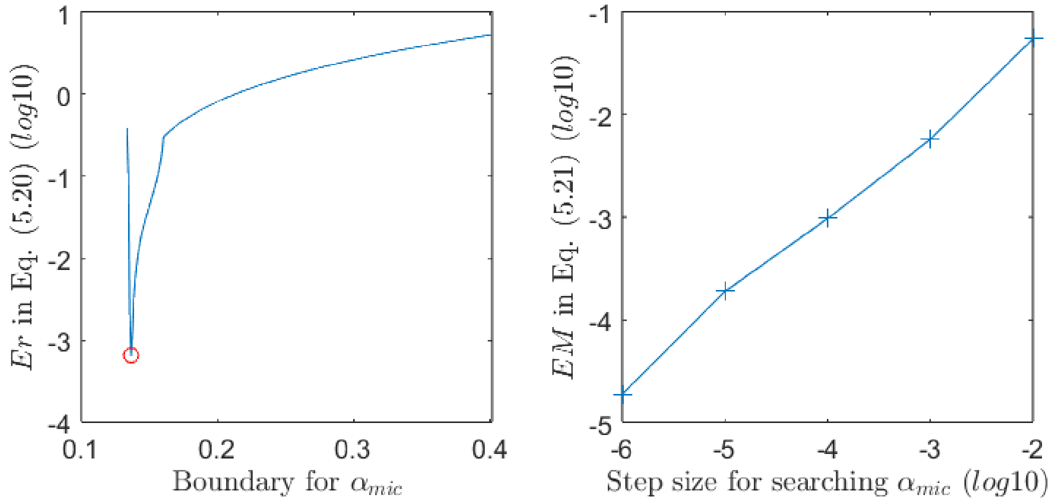
$$EM = \frac{\sum_{i=1}^{M} \|\mathbf{r}_i - \mathbf{r}_i^*\| + \sum_{j=1}^{N} \|\mathbf{s}_j - \mathbf{s}_j^*\|}{M+N}, \tag{5.21}$$

where $\mathbf{r}_i^*$ and $\mathbf{s}_j^*$ represent the estimated locations of the i^{th} microphone and j^{th} source, respectively.

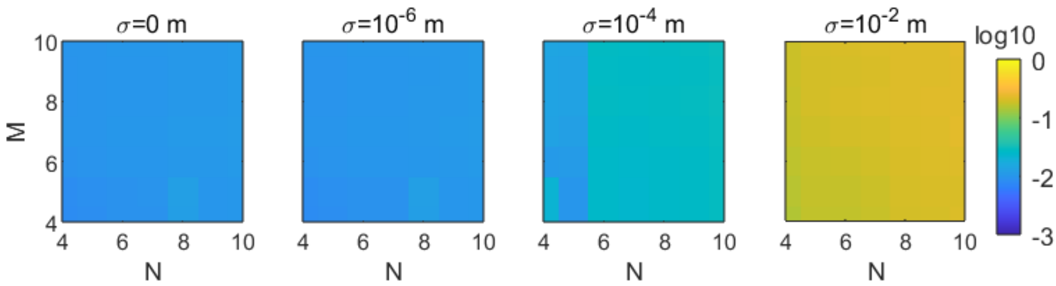
5.3.2 Results

Results for different configurations and noise intensity σ

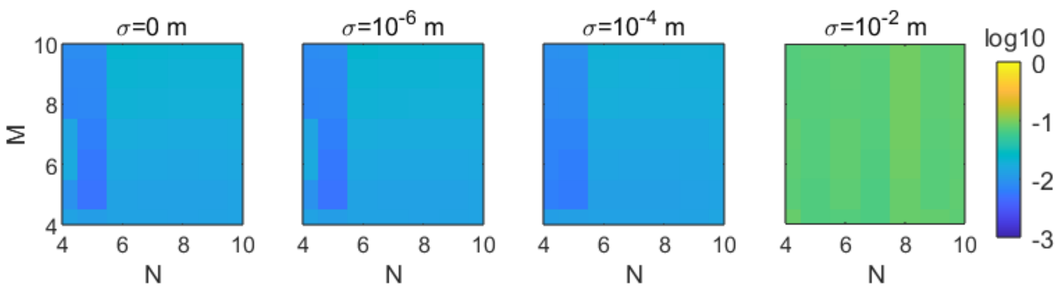
In this part, 30 different random configurations are conducted for given M microphones and N sources, then average the localization errors EM in Eq. (5.21) over these configurations. Figs. 5.3, 5.4 and 5.5 present the results using simulation data from these 30 configurations. Specifically, Figs.



(a) One co-located additional microphone and source ($M = 4$ and $N = 4$): the principle for searching α_{mic} by using Er in Eq. (5.20) (left), and the effect of α_{mic} on localization errors EM in Eq. (5.21) with different searching size of step (right).



(b) Localization error for proposed numerical method: scenario with one co-located additional microphone and source.



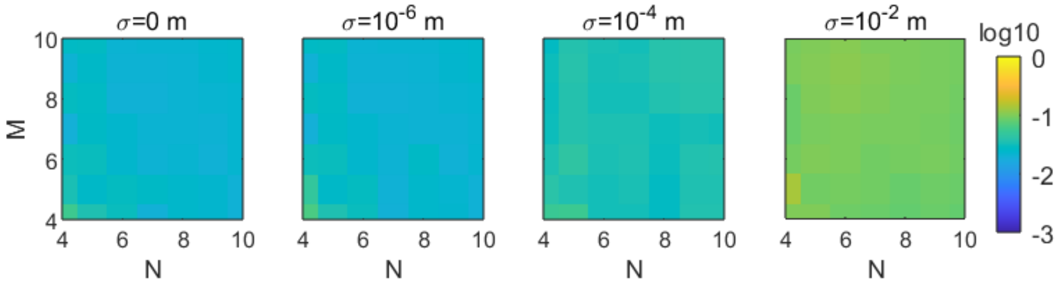
(c) Localization error for proposed numerical method: scenario with two known distances of any one pair of microphones and one pair of source.

Figure 5.4: Results for proposed numerical method; log 10 denotes the transformation of $\log_{10}$ for the corresponding values; known β_s, γ_s and η_s for figure (a); known β_s, γ_s and η_s for figure (b); known α_{mic} and β_s for figure (c).

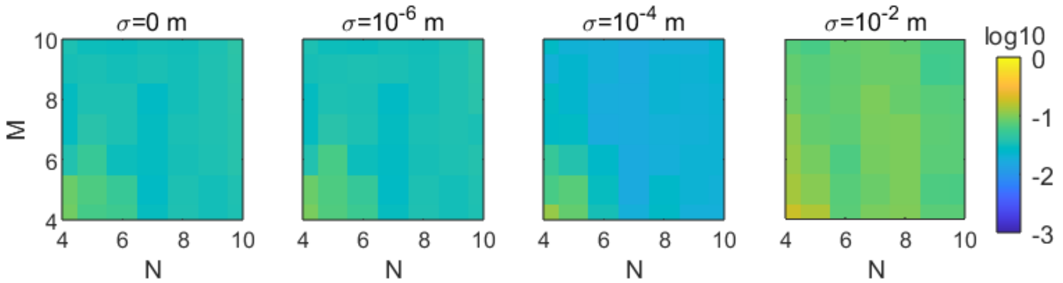
5.3 and 5.4 display the results for first three scenarios outlined in Section 5.2.5, validating proposed method's claim that the locations of microphones and sources can be determined by four unknown variables when $M \geq 4$ and $N \geq 4$. Figure 5.5 shows results for the last two scenarios in Section 5.2.5, further validating that four variables can represent the locations of all microphones and sources while also relaxing the minimal configurations established by state-of-the-art methods, thus enhancing the flexibility of JMSL configurations.

First, Fig. 5.3 presents the results for the scenario with one co-located additional microphone and source and one known distance between a pair of microphones. This scenario implies that the four variables α_{mic}, β_s, γ_s and η_s in proposed numerical method are known. As shown in Fig. 5.3, with a noise intensity of 0 m, the average localization error EM of 30 different configurations is approximately 10^{-15} m when both M and N vary from 4 to 10, validating proposed method's assertion that the locations of all microphones and sources can be represented by four unknown variables. Moreover, when noise is introduced into both the four variables (α_{mic}, β_s, γ_s and η_s) and the distances $d_{i,j}$ between microphones and sources, the average localization errors for the 30 configurations vary with the noise intensity σ. Specifically, for $\sigma = 10^{-6}$ m, the average localization error EM ranges from 10^{-5} m to 10^{-4} m when both M and N vary from 4 to 10. For $\sigma = 10^{-4}$ m, the average localization error EM ranges from 10^{-3} m to 10^{-2} m for the same range of M and N. For $\sigma = 10^{-2}$ m, the average localization error EM also ranges from 10^{-2} m to 10^{-1} m when both M and N vary from 4 to 10. Overall, the results in Fig. 5.3 confirm the proposed numerical method's claim that the locations of all microphones and sources can be represented by four unknown variables when $M \geq 4$ and $N \geq 4$, demonstrating its applicability in scenarios where one additional microphone and one source are co-located, and the distance between a pair of microphones is known.

Second, Fig. 5.4 shows the results for other two scenarios: one with an additional co-located microphone and source, and another with two known distances between any two microphones and any two sources. To determine the step size for searching the corresponding four unknowns α_{mic}, β_s, γ_s and η_s, the scenario with one additional co-located microphone and source with $M = 4$ and $N = 4$ are visualized. As shown in the left sub-figure of Fig. 5.4(a), when setting the step size to 10^{-3} m for searching the optimal α_{mic} within the boundaries in Eq. (5.16), the optimal α_{mic} can be found by choosing the minimal Er in Eq. (5.20). The right sub-figure of Fig. 5.4(a) illustrates that as the searching step size for α_{mic} varies from 10^{-6} m to 10^{-2} m, the value of average localization error EM for 30 different configurations is nearly linear. Specifically, with a searching step of α_{mic} is 10^{-3} m, the average localization error EM for 30 configurations is approximately 10^{-2} m. Therefore, to balance localization accuracy and the time required for searching unknowns, I set the step size



(a) Localization error for proposed numerical method: scenario with one known distance for a pair of microphones ($M = 4$ and $N = 4$ relaxes the minimal configuration of state-of-the-arts).



(b) Localization error for proposed numerical method: scenario without known distances for any pairs of microphones and any pairs of sources ($M = 4$ and $N = 4$, $M = 4$ and $N = 5$ and $M = 5$ and $N = 4$ relaxes the minimal configurations of state-of-the-arts).

Figure 5.5: Results for two scenarios: unknown three variables β_s, γ_s and η_s for figure (a); unknown four variables α_{mic}, β_s, γ_s and η_s for figure (b) ($\log 10$ denotes the transformation of $\log_{10}$ for the corresponding values).

for searching the four unknowns α_{mic}, β_s, γ_s and η_s to 10^{-3} m within the boundaries of Eq. (5.16), unless otherwise specified.

With a search step size of 10^{-3} m for α_{mic}, Fig. 5.4(b) shows the results for the scenario with one co-located additional microphone and source, implying that three variables β_s, γ_s and η_s are known. As seen in Fig. 5.4(b), if the noise intensity is 0 m, the average localization error EM across 30 different configurations ranges from 0.008 m to 0.012 m as both M and N vary from 4 to 10. This supports the claim that the proposed numerical method can represent the locations of all microphones and sources using four unknown variables. When noise is introduced to β_s, γ_s, η_s and the distance measurements $d_{i,j}$, the average localization error changes with the noise intensity σ. Specifically, with $\sigma = 10^{-6}$ m, the average localization error EM for 30 configurations is similar to that when $\sigma = 0$ m. If $\sigma = 10^{-4}$ m, the average localization error EM ranges from 0.009 m to 0.029 m for 30 configurations as both M and N vary from 4 to 10. When $\sigma = 10^{-2}$ m, the average localization error EM ranges from 0.14 m to 0.22 m. Overall, the results in in Fig. 5.4(b) validate the statement that the proposed numerical method can represent the locations of all microphones and sources with four unknown variables when $M \geq 4$ and $N \geq 4$, resulting in the conclusion that proposed numerical method can be applied to the scenario with one co-located additional microphone and source, demonstrating that a 10^{-3} m search step for α_{mic} is feasible for JMSL.

With a search step size of 10^{-3} m for both γ_s and η_s, Fig. 5.4(c) shows the results for the scenario with two known distances between any pair of microphones and any pair of sources, implying that α_{mic} and β_s are known. As shown in Fig. 5.4(c), when the noise intensity is 0 m, the average localization error EM for 30 configurations ranges from 0.005 m to 0.022 m as M and N vary from 4 to 10. This also validates the proposed numerical method, indicating that the locations of all microphones and sources can be represented by four unknown variables. When noise is added to α_{mic}, β_s and the distance measurements $d_{i,j}$, the average localization error EM varies with σ. Specifically, when $\sigma = 10^{-6}$ m and $\sigma = 10^{-4}$ m, the average localization errors EM of those two situations for 30 configurations are similar to those when $\sigma = 0$ m. If $\sigma = 10^{-2}$ m, the average localization error EM ranges from 0.06 m to 0.08 m for 30 configurations as M and N vary from 4 to 10. Overall, the results in Fig. 5.4(c) validate that the proposed numerical method can represent the locations of all microphones and sources with four unknown variables when $M \geq 4$ and $N \geq 4$, resulting in the conclusion that proposed numerical method can be applied to the scenario when two distances between any one pair of microphones and any one pair of sources are known, demonstrating that a 10^{-3} m search step for α_{mic} and β_s is feasible for JMSL.

Moreover, comparing the results from Figs. 5.4(b) and 5.4(c) when $\sigma = 10^{-2}$ m, it is evident that the scenario with two known distances (Fig. 5.4(c)) achieves better localization accuracy than the scenario with three known distances (Fig. 5.4(b)). Specifically, with fixed noise intensity in both distance measurements $d_{i,j}$ and known variables (i.e., β_s, γ_s, and η_s in Fig. 5.4(b) and γ_s and η_s in Fig. 5.4(c)), searching for the optimal solutions for the remaining unknown variables (α_{mic} in Fig. 5.4(b) and α_{mic} and β_s in Fig. 5.4(c)) allows the metric in Eq. (5.20) to choose the optimal values of unknown variables that minimize Er in Eq. (5.20). This results in better localization accuracy in Fig. 5.4(c) than in Fig. 5.4(b), due to the presence of noise in three variables (β_s, γ_s, and η_s) in Fig. 5.4(b), compared to noise in only two variables (γ_s and η_s) in Fig. 5.4(c).

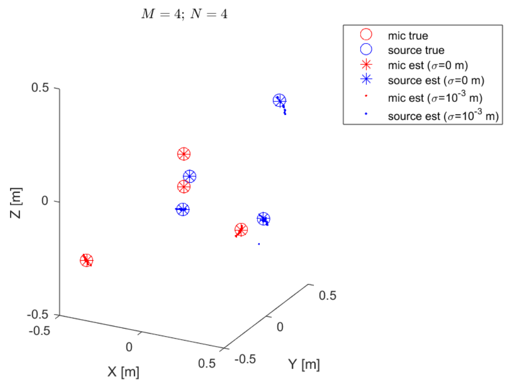
Third, Fig. 5.5(a) and (b) shows the results for two scenarios: 1) one known distance of a pair of microphones; 2) no prior information for any pairs of microphones or sources. With a search step size of 10^{-3} m for the three variables β_s γ_s and η_s, Fig. 5.5(a) shows the localization results when α_{mic} is known. As illustrated in Fig. 5.5(a), if the noise intensity σ is 0 m, 10^{-6} m, 10^{-4} m and 10^{-2} m, the average localization error EM of 30 different configurations ranges within

$\begin{bmatrix}0.01 & 0.05\end{bmatrix}$ m, $\begin{bmatrix}0.01 & 0.06\end{bmatrix}$ m, $\begin{bmatrix}0.02 & 0.06\end{bmatrix}$ m and $\begin{bmatrix}0.08 & 0.13\end{bmatrix}$ m, respectively. This validates that the proposed numerical method can represent the locations of all microphones and sources with four variables. Furthermore, when $M = 4$ and $N = 4$, the average localization errors EM of 30 different configurations for $\sigma = 0$ m $\sigma = 10^{-6}$ m, $\sigma = 10^{-4}$ m and $\sigma = 10^{-2}$ m are 0.05 m, 0.06 m 0.05 m and 0.09 m, respectively, validating that the proposed method can localize microphones and sources with $M = 4$ and $N = 4$, which is fewer than the minimal configurations reported in state-of-the-art methods over the past decades (see analysis of the fourth scenario in Section 5.2.5). Additionally, with a search step size of 10^{-3} m for four unknown variables α_{mic}, β_s, γ_s and η_s, Fig. 5.5(b) shows the corresponding localization results. As seen from Fig. 5.5(b), if the noise intensity σ is 0 m, 10^{-6} m, 10^{-4} m and 10^{-2} m, the average localization error EM for 30 different configurations ranges within $\begin{bmatrix}0.02 & 0.09\end{bmatrix}$ m, $\begin{bmatrix}0.02 & 0.09\end{bmatrix}$ m, $\begin{bmatrix}0.01 & 0.11\end{bmatrix}$ m and $\begin{bmatrix}0.05 & 0.17\end{bmatrix}$ m, respectively. This further validates that the locations of all microphones and sources can be represented by four unknown variables. More importantly, when $M = 4$ and $N = 4$, the average localization errors EM of 30 different configurations for $\sigma = 0$ m, $\sigma = 10^{-6}$ m, $\sigma = 10^{-4}$ m and $\sigma = 10^{-2}$ m are 0.09 m, 0.09 m, 0.11 m and 0.17 m, respectively. When $M = 4$ and $N = 5$, these errors are 0.06 m, 0.06 m 0.07 m and 0.14 m, respectively. When $M = 5$ and $N = 4$, the errors are 0.08 m, 0.08 m 0.06 m and 0.13 m, respectively. These results confirm that the proposed numerical method can be used for localizing both microphones and sources even when the number of microphones and sources is less than the requirement of minimal configurations reported in state-of-the-art methods over the past decades (see analysis of fifth scenario in Section 5.2.5).

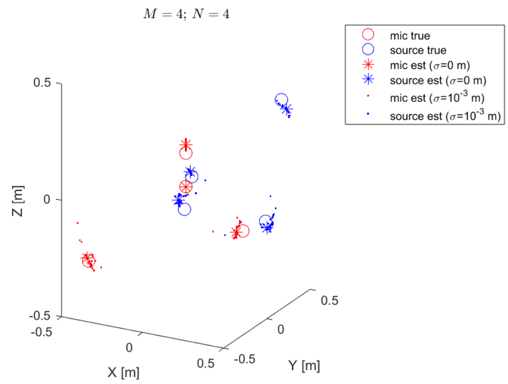
Results for fixed configurations and noise intensity σ

In this part, to further validate and evaluate that the proposed numerical method can relax the minimal configurations for JMSL, 30 independent experiments with fixed configurations and a noise intensity $\sigma = 10^{-3}$ are conducted. The step size for searching unknowns is set to 10^{-3} m within the boundaries defined in Eq. (5.16). Fig. 5.6 presents the results for the following cases: 1) known α_{mic} with $M = 4$ and $N = 4$; 2) unknown four variables with $M = 4$ and $N = 4$; 3) unknown four variables with $M = 4$ and $N = 5$; 4) unknown four variables with $M = 5$ and $N = 4$. Additionally, to visualize the impact of a noise intensity of $\sigma = 10^{-3}$ m on localizing microphones and sources, I display the ground truth and the estimated locations for $\sigma = 0$ m for both microphones and sources.

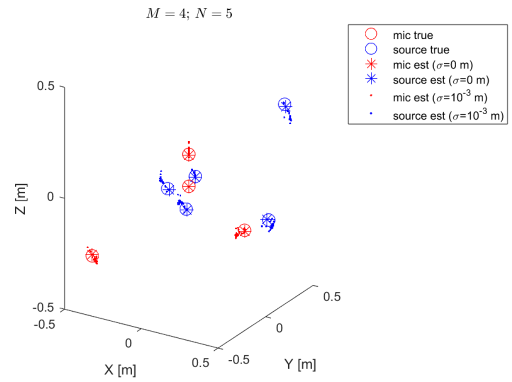
Using a step size of 10^{-3} m for searching three unknowns, Fig. 5.6(a) shows the results for the case when α_{mic} is known. From Fig. 5.6(a), it can be observed that the EM for $\sigma = 0$ m is 0.006



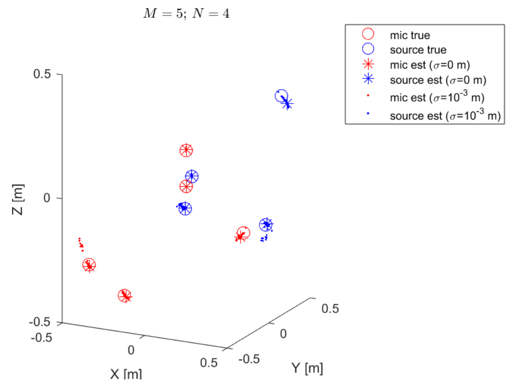
(a) Known α_{mic}; mean EM for $\sigma = 10^{-3}$ m: 0.04 m.



(b) Unknown four variables; mean EM for $\sigma = 10^{-3}$ m: 0.10 m.



(c) Unknown four variables; mean EM for $\sigma = 10^{-3}$ m: 0.06 m.



(d) Unknown four variables; mean EM for $\sigma = 10^{-3}$ m: 0.04 m

Figure 5.6: Localization results of the proposed numerical method for relaxing the minimal configurations of state-of-the-arts with the step size of 10^{-3} m.

m, indicating high localization accuracy. Even when noise is introduced to both α_{mic} and the distances between microphones and sources, the localization accuracy remains high, with an average localization error EM of just 0.04 m for 30 different experiments when $\sigma = 10^{-3}$ m. Then in Figs. 5.6(b) (c) and (d), using a step size of 10^{-3} m for searching four unknowns, I show the results for scenarios with four unknowns for $M = 4$ and $N = 4$, $M = 4$ and $N = 5$ and $M = 5$ and $N = 4$, respectively. For $\sigma = 0$ m, the localization errors EM for $M = 4$ and $N = 4$, $M = 4$ and $N = 5$ and $M = 5$ and $N = 4$ are 0.079 m, 0.016 m, 0.021 m, respectively. When noise ($\sigma = 10^{-3}$ m) is introduced into the distances between microphones and sources, the localization accuracy remains good, with average localization errors EM of 0.1 m, 0.06 m and 0.04 m for $M = 4$ and $N = 4$, $M = 4$ and $N = 5$ and $M = 5$ and $N = 4$, respectively. These results in Fig. (5.6) validate that proposed method effectively relaxes the minimal configurations for JMSL, making the configurations more flexible and enhancing the field of signal processing.

5.3.3 Discussions

The proposed numerical method converts the localization problems into the task of estimating four unknown variables. The lower and upper boundaries of these variables are determined using triangle inequalities. The solutions for these variables are then found by using the known range measurements between microphones and sources. This method challenges the long-standing consensus on the minimal configurations required by state-of-the-art methods over the past decades, thereby making the configurations of JMSL more flexible, enhancing the efficiency of JMSL.

In a simulation room of $1 \times 1 \times 1$ m^3 with a search step of 10^{-3} m, the proposed numerical method achieved accurate localization results. Although these results are from simulations, the only requirement for the proposed method is having the range measurements between microphones and sources, suggesting it can also achieve accurate localization in real environments. In addition, the results in Fig. (5.3), (5.4), (5.5) and (5.6) also show the robustness of proposed numerical method. When different levels of noise with varying intensities σ are introduced into the range measurements, the proposed numerical method demonstrates consistent and accurate localization results. This indicates that the method is robust to external noise and not significantly sensitive to errors in the initial range measurements. While the proposed numerical method demonstrates theoretical robustness, its practical applicability in real-world scenarios requires consideration of varying acoustic conditions and hardware limitations. In environments with noise, reverberation, or occlusions, pre-processing techniques such as noise reduction, signal enhancement, and adaptive filtering can be employed to improve the accuracy of time-of-arrival TOA measurements. The method's effectiveness has been validated through simulations with added noise, indicating its potential for deployment in practical settings. Furthermore, the design of JMSL systems must account for the specific characteristics of different microphone types, such as capacitor and MEMS microphones, which vary in sensitivity, resolution, and range. To ensure accurate localization, careful attention to microphone synchronization and calibration is crucial. The microphone arrangement should also be optimized based on the environment—indoor settings may benefit from strategic placement and acoustic treatment to minimize reflections, while outdoor applications may require robust systems capable of handling environmental interferences like wind noise. These considerations demonstrate that the proposed method is adaptable to a wide range of real-world conditions and provides practical guidelines for effective JMSL system design.

In addition, the proposed numerical method is applicable to larger rooms since room size is not a prerequisite. However, it has a significant limitation: the time required to search for optimal solutions for the unknown variables. For example, when performing JMSL using the proposed method, once the distance matrix D^* in Eq. (2.11) and the two singular matrices $\mathbf{U}_\mathbf{p}^*$ and $\mathbf{V}_\mathbf{p}^*$ in Eq. (2.13) are obtained, the key task is to determine the values of eleven unknowns: $s_{2,1}$, $s_{3,1}$, and nine variables in C. From Eqs. (5.3), (5.4), (5.6), (5.11), and (5.15), it is evident that the time complexity for computing these nine variables is constant, i.e., $O(1)$, indicating independence from the data scale. However, the computation of these eleven variables depends on the precise values of four additional unknowns (α_{mic}, β_s, γ_s and η_s). Solving for these four variables requires an iterative search within the boundaries defined in Eq. (5.16). For instance, if the boundaries for all four unknowns span one meter, the number of iterations needed to find the optimal numerical solutions for JMSL is approximately 10^9 when α_{mic} is known and 10^{12} when all four variables are unknown. This illustrates the computational intensity and time-consuming nature of the proposed numerical method.

5.4 Summary

In this chapter, the main focus is to relax the minimal configurations of state-of-the-arts during past decades by using the synchronized TOA/range measurements between microphones and sources. By formulating the localization problems with several triangles and applying the laws of cosine to those triangles, the localization problems of microphones and sources can be transformed to the estimation of four unknown distances pertaining to a pair of microphones and three pairs of sources. Then the triangle inequalities have been used for obtaining the lower and upper boundaries for the four unknown distances, so that I can search the optimal numerical solutions for those four unknowns under the corresponding lower and upper boundaries, given an appropriate step. Experimental results validate the feasibility of proposed numerical method: 1) the locations of all microphones and sources can be represented by four variables when the number of both microphones and sources is larger than or equal to four; 2) four microphones and four sources are enough to localize both microphones and sources when the distance of one pair of microphones is known. 3) even without prior information about the distances between a pair of microphones and three pairs of sources, all microphones and sources can still be localized when the number of microphones and sources is four/four/five and four/five/four, respectively. This is fewer than the minimal number typically required for localizing both microphones and sources according to state-of-the-art methods. Therefore, proposed research output in this chapter is a milestone for the topic of JMSL, as it not only challenges the long-term

assumption that the task of JMSL must need at least six/five/four microphones and four/five/six sources, respectively, but also making the configurations of JMSL more flexible and efficient.
In addition, the research output in this chapter opens an new research gate for JMSL by challenging the consensus of minimal configurations during past decades. By breaking the consensus of minimal configurations for JMSL in state-of-the-arts during the past decades, finding more efficient way to obtain the solutions of four unknowns can be an new research sub-topic of JMSL, though further reducing the number of microphones and sources might be another alternative research direction for JMSL.

Chapter 6

Conclusion and Future Work

6.1 Summary

This PhD thesis addresses the asynchronous self-localization of microphones and audio sources by proposing a TSF, enhancing the efficiency of accuracy for self-localization. First, I demonstrate the equivalence of TOA and TDOA, proving that microphone signals alone suffice for self-localization, thereby simplifying the task, making the task of self-localization more efficient. Then a CLRA method is introduced to estimate the UTIm for asynchronous TOA/TDOA, which helps obtain the accurate range measurements between microphones and sources. Finally, the primary focus of this thesis is to relax the minimal configurations required by state-of-the-art methods from the past decades by using the range measurements between microphones and audio sources, enhancing the efficiency of self-localization .

In Chapter 2, a comprehensive literature review on audio self-localization is presented, covering TOA-based, TDOA-based, AOA-based, enerygy-based and pairwise distance estimation methods. Despite advancements in these areas, several research gaps remain. Firstly, TOA measurements require both microphone and source signals, whereas TDOA measurements need only microphone signals. After estimating the UTIm for TOA and TDOA, TOA provides range measurements between microphones and sources, while TDOA only offers range differences relative to the source. Range measurements are more informative for self-localization. Therefore, before estimating the UTIm, unifying TOA and TDOA is crucial so that range measurements can be obtained using only microphone signals. I propose a solution for this in Chapter 3. Secondly, after unifying TOA and TDOA, the next task is to estimate the UTIm. Existing LRP is in a risk of getting stuck in local minima if the UTIm initialization is inappropriate, which is often random. Thus, it is urgent to explore more low-rank structure information of UTIm. I propose a solution for this in Chapter 4. Third, once the range measurements between microphones and sources are available, traditional methods have aimed to achieve minimal configurations for localization using either closed-form or iterative solutions, typically involving six/five/four microphones and four/five/six sources. The consensus on minimal configurations is based on having more or equal number of equations (range measurements) than unknowns (microphone and source locations). However, it remains unexplored to localize both microphones and sources with fewer than the minimal configurations stated by state-of-the-art methods, which could make the localization task more efficient. Therefore, I present a numerical solution in Chapter 5. This chapter also introduced the problem formulation and preliminaries related to the three challenges mentioned above.

In Chapter 3, I proposed a mapping function for the transformation of TOA and TDOA. With proposed mapping function, I proved that the transformation of TOA and TDOA can be identical to each other, challenging the long-term assumptions that TOA measurements needs to be estimated with both microphones and sources signals, obtaining the range measurements between microphones and sources once the UTIm in TOA/TDOA is estimated, making the task of self-localization more efficient.

In Chapter 4, I proposed a CLRA method for UTIm estimation. Traditional methods use LRP for UTIm estimation but risk local minima due to random initialization. I introduced three new LRP variants, supported by mathematical proof, enriching the pool of LRP solutions. I then introduced four linear constraints for UTIm based on LRP and the proposed variants, and applied the Gaussian Newton method to solve for the UTIm, achieving a global optimal solution, enhancing the accuracy of range measurements between microphones and sources for self-localization.

In Chapter 5, I presented a numerical method that transforms the localization problem into the estimation of four unknowns, deriving their lower and upper boundaries. This allows obtaining solutions within these boundaries, relaxing the minimal configurations required by state-of-the-art methods over the past decades. This method shows that fewer microphones and sources are needed for JMSL than previously thought, making self-localization more flexible and efficient.

6.2 Future Research Plan

In this section, I outline potential future research directions for asynchronous TOA and TDOA-based self-localization.

- Enhancing Low-Rank Property Variants: The proposed LRPV1 and LRPV2 methods in Chapter 4 have limitations related to the number of microphones and sources. Specifically, LRPV1 maintains the low-rank property only when $M-1 > N-1+3$ and $N-1 > 3$, and LRPV2 maintains it when $N-1 > M-1+3$ and $M-1 > 3$. These constraints limit the applicability of the methods for UTIm estimation. To overcome this, future research could focus on developing new variants of LRPV1 and LRPV2. By reforming their structures using advanced mathematical operations, their low-rank properties are retained while eliminating the current limitations.

- Improving the CLRA Method: The proposed CLRA method in Chapter 4 has certain limitations in its denoising capabilities. One way to address this is to revise the objective function in Eq. (4.78) to better account for noise in TOA/TDOA measurements. This adjustment could enhance the CLRA's denoising performance, thereby improving the accuracy of both UTIm estimation and the resulting range measurements.

- Optimizing the Numerical Method for JMSL: The numerical method proposed in Chapter 5 aims to relax the minimal configurations required for JMSL in state-of-the-art methods. However, the method's search process for the four unknowns is time-consuming. Future research could explore incorporating energy information from both microphone and source signals to derive pairwise distances pertaining to pairs of microphones or sources more efficiently. This approach could significantly speed up the localization process and make the configurations more flexible and efficient.

- Integrating UTIm Estimation and Localization: Despite the advancements in the proposed TSF for audio self-localization in this PhD thesis, there remains a gap between UTIm estimation in the second stage and the localization of microphones and sources in the third stage. The third stage assumes known range measurements between microphones and sources, which suggests that four microphones and four sources are sufficient for self-localization. However, in the second stage, the minimal requirement for the number of microphones and sources for UTIm estimation is $(M-5)(N-5) \geq 8$, making four microphones and four sources insufficient. A potential solution is to design a method that simultaneously estimates UTIm and performs localization. This would enable more flexible self-localization configurations for both synchronous and asynchronous scenarios.

- Improved Temporal Estimation in Asynchronous Environments Based on the Proposed CLRA

Method: Building upon the CLRA technique introduced in this thesis for UTIm in asynchronous environments, future research could further enhance the accuracy and robustness of temporal estimation. While the proposed method shows significant improvements over traditional approaches, there is potential for further refinement using advanced machine learning or optimization techniques. By leveraging the augmented low-rank structural information from various data sources, future work could reduce errors even in noisy or highly dynamic environments. The impact of these improvements would be profound in applications such as human-robot interaction, where accurate temporal estimation is crucial for seamless coordination, or in elderly monitoring systems, where reliability in localization could significantly improve safety and response times.

- Multimodal Sensory Fusion Leveraging the Proposed TOA-TDOA Mapping Function: The novel mapping function between TOA and TDOA developed in this research demonstrates the potential for using only microphone signals for TOA-based self-localization. Future work could extend this by integrating other sensory modalities, such as visual or LiDAR data, with the proposed audio-based framework. This multimodal approach would combine the advantages of audio self-localization in poor lighting or obstructed environments with visual or range-based sensors to provide even more robust and accurate localization. The potential impact would be transformative for autonomous robots navigating in complex environments, where multimodal fusion could enable more reliable decision-making and interaction. For applications in elderly care, this could enhance monitoring systems to function in a broader range of conditions, improving overall safety and coverage.

- Further Reduction in Device Requirements Using the Proposed Numerical Method: This thesis presents a novel numerical method that reduces the minimum number of microphones and sources required for effective self-localization. Future research could explore additional optimizations to further minimize hardware requirements while maintaining or improving accuracy. By refining the proposed method, it may be possible to achieve accurate localization with even fewer devices, making the technology more accessible and affordable for practical deployment. The potential impact of this work is significant for low-cost applications, such as home security or wearable devices, where minimal hardware is essential. This reduction in complexity could broaden the scope of audio self-localization, enabling its use in more resource-constrained environments without compromising performance.

- Real-Time Self-Localization Systems Based on the Proposed TSF Framework: The TSF introduced in this dissertation has already demonstrated improvements in both the accuracy and efficiency of self-localization. Future work could focus on enhancing the real-time capabilities of this framework by optimizing computational efficiency. This could involve further developing the algorithms to process data faster or implementing parallel processing techniques to handle larger datasets in real-time scenarios. The impact of such developments would be particularly important for time-critical applications, such as autonomous driving or emergency response systems, where rapid and accurate localization is essential. By achieving real-time performance, the proposed methods could expand the practical applications of audio self-localization to dynamic, fast-paced environments where instant responses are crucial.

Bibliography

[1] M. V. Moreno-Cano, M. A. Zamora-Izquierdo, J. Santa, and A. F. Skarmeta. An indoor localization system based on artificial neural networks and particle filters applied to intelligent buildings. *Neurocomputing*, 122:116–125, 2013.

[2] Z. Wang, L. Liu, and M. Zhou. Protocols and applications of ad-hoc robot wireless communication networks: An overview. *future*, 10(20), 2005.

[3] T. Oshida, K. Nakadai, and H. G. Okuno. Automatic speech recognition improved by two-layered audio-visual integration for robot audition. *2009 9th IEEE-RAS International Conference on Humanoid Robots*, pages 604–609, 2009.

[4] K. L. Su. Automatic fire detection system using adaptive fusion algorithm for fire fighting robot. *2006 IEEE International Conference on Systems, Man and Cybernetics*, 2.

[5] H. D. Kim, J. S. Choi, and M. Kim. Human-robot interaction in real environments by audio-visual integration. *International Journal of Control, Automation, and Systems*, 5(1):61–69, 2007.

[6] X. Wu, H. Gong, P. Chen, Z. Zhong, and Y. Xu. Surveillance robot utilizing video and audio information. *Journal of Intelligent and Robotic Systems*, 55(4-5):403–421, 2009.

[7] M. Bernard, K. Kondak, I. Maza, and A. Ollero. Autonomous transportation and deployment with aerial robots for search and rescue missions. *Journal of Field Robotics*, 28(6):914–931, 2011.

[8] I. P. Alonso, D. F. F. Llorca, M. Gavilan, S. A. A. Pardo, M. A. Garcia-Garrido, L. Vlacic, and M. A. Sotelo. Accurate global localization using visual odometry and digital maps on urban environments. *IEEE Transactions on Intelligent Transportation Systems*, 13(4):1535–1545, 2012.

[9] A. Couturier and M. A. Akhloufi. A review on absolute visual localization for uav. *Robotics and Autonomous Systems*, 135(103666), 2021.

[10] Y. Wu, R. Ayyalasomayajula, M. J. Bianco, D. Bharadia, and P. Gerstoft. Sslide: Sound source localization for indoors based on deep learning. *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4680–4684, 2021.

[11] J. Y. Lee, S. Y. Chi, J. Y. Lee, M. Hahn, and Y. J. Cho. Real-time sound localization using time difference for human robot interaction. *IFAC Proceedings Volumes*, 38(1):54–57, 2005.

[12] K. Nakamura, K. Nakadai, F. Asano, and G. Ince. Intelligent sound source localization and its application to multimodal human tracking. *In 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 143–148, 2011.

[13] M. H. Korayem, S. Azargoshasb, A. H. Korayem, and S. Tabibian. Design and implementation of the voice command recognition and the sound source localization system for human–robot interaction. *Robotica*, 39(10):1779–1790, 2021.

[14] A. Bertrand. Applications and trends in wireless acoustic microphone networks: A signal processing perspective. *Proc. IEEE SCVT*, pages 1–6, 2011.

[15] M. H. Hennecke and G. A. Fink. Towards acoustic self-localization of ad hoc smartphone arrays. *Proc. HSCMA*, pages 127–132, 2011.

[16] C. Mahapatra and A. R. Mohanty. Explosive sound source localization in indoor and outdoor environments using modified levenberg marquardt algorithm. *Meas.*, 187(110362), 2022.

[17] S. Tervo, J. Pätynen, and T. Lokki. Acoustic reflection path tracing using a highly directional loudspeaker. *Proc. IEEE WASPAA*, pages 245–24, 2009.

[18] M. S. Brandstein and H. F. Silverman. A robust method for speech signal time-delay estimation in reverberant rooms. *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, 1:375–378, 1997.

[19] C. Knapp and G. Carter. The generalized correlation method for estimation of time delay. *IEEE transactions on acoustics, speech, and signal processing*, 24(4):320–327, 1976.

[20] M. Azaria and D. Hertz. Time delay estimation by generalized cross correlation methods. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(2):280–285, 1984.

[21] G. A. Darbellay and I. Vajda. Estimation of the information by an adaptive partitioning of the observation space. *IEEE Transactions on Information Theory*, 45(4):1315–1321, 1999.

[22] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Physical review E*, 69(6):066138, 2004.

[23] M. Crocco, A. D. Bue, and V. Murino. A bilinear approach to the position self-calibration of multiple microphones. *IEEE Trans. Signal Process.*, 60(2):660–673, 2011.

[24] A. Contini, A. Canclini, F. Antonacci, M. Compagnoni, A. Sarti, and S. Tubaro. Self-calibration of microphone arrays from measurement of times of arrival of acoustic signals. *Proc. IEEE ISCCSP*, pages 1–6, 2012.

[25] S. D. Valente, M. Tagliasacchi, F. Antonacci, P. Bestagini, A. Sarti, and S. Tubaro. Geometric calibration of distributed microphone arrays from acoustic source correspondences. *Proc. IEEE MMSP*, pages 13–18, 2010.

[26] P. D. Jager, M. Trinkle, and A. Hashemi-Sakhtsari. Automatic microphone array position calibration using an acoustic sounding source. *Proc. IEEE ICIEA*, pages 2110–2113, 2009.

[27] P. Pertilä, M. Mieskolainen, and M. S. Hämäläinen. Closed-form self-localization of asynchronous microphone arrays. *Proc. IEEE HSCMA*, pages 139–144, 2011.

[28] X. Li, Z. D. Deng, L. T. Rauchenstein, and T. J. Carlson. Contributed review: Source-localization algorithms and applications using time of arrival and time difference of arrival measurements. *Rev. Sci. Instrum.*, 87(4):041502, 2016.

[29] I. R. Urazghildiiev and D. E. Hannay. Localizing sources using a network of synchronized compact arrays. *IEEE J. Ocean. Eng.*, 46(4):1302–1312, 2021.

[30] F. Meyer, A. Tesei, and M. Z. Win. Localization of multiple sources using time-difference of arrival measurements. *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, pages 3151–3155, 2017.

[31] B. Laufer-Goldshtein, R. Talmon, and S. Gannot. A hybrid approach for speaker tracking based on tdoa and data-driven models. *IEEE/ACM Trans. Audio, Speech, Language Process.*, 26(4):725–735, 2018.

[32] D. Bechler, M. Grimm, and K. Kroschel. Speaker tracking with a microphone array using kalman filtering. *Adv. Radio Sci.*, 1(B.3):113–117, 2003.

[33] T. K. Le and N. Ono. Closed-form and near closed-form solutions for toa-based joint source and sensor localization. *IEEE Transactions on Signal Processing*, 64(18):4751–4766, 2016.

[34] T. K. Le and K. C. Ho. Algebraic complete solution for joint source and microphone localization using time of flight measurements. *IEEE Trans. Signal Process.*, 68:1853–1869, 2020.

[35] T. K. Le and N. Ono. Closed-form and near closed-form solutions for tdoa-based joint source and microphone localization. *IEEE Trans. Signal Process.*, 65(5):1207–1221, 2016.

[36] C. Evers and P. A. Naylor. Acoustic slam. *IEEE/ACM Trans. Audio, Speech, Language Process.*, 26(9):1484–1498, 2018.

[37] M. Crocco, A. Trucco, and A. D. Bue. Uncalibrated 3d room geometry estimation from sound impulse responses. *J. Franklin Inst.*, 354(18):8678–8709, 2017.

[38] F. Cao, Y. Cheng, A. M. Khan, Z. Yang, and Z. Guo. Are microphone signals alone sufficient for joint microphones and sources localization? *Accepted by 29th Internation Conference on Automation and Computing*, 2024.

[39] T. K. Le and K. C. Ho. Joint source and sensor localization by angles of arrival. *IEEE Transactions on Signal Processing*, page 6521–6534, 2020.

[40] T. K. Le and K. C. Ho. Uncovering source ranges from range differences observed by sensors at unknown positions: Fundamental theory. *IEEE Trans. Signal Process.*, 67(10):2665–2678, 2019.

[41] D. E. Badawy, V. Larsson, M. Pollefeys, and I. Dokmanic. Localizing unsynchronized sensors with unknown sources. *IEEE Trans. Signal Process.*, 71:641–654, 2023.

[42] P. H. Schönemann. On metric multidimensional unfolding. *Psychometrika*, 35(3):349–366, 1970.

[43] I. McCowan, M. Lincoln, and I. Himawan. Microphone array shape calibration in diffuse noise fields. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(3):666–670, 2008.

[44] M. J. Taghizadeh, P. N. Garner, and H. Bourlard. Enhanced diffuse field model for ad hoc microphone array calibration. *Signal Processing*, 101:242–255, 2014.

[45] W. S. Torgerson. Multidimensional scaling: I. theory and method. *Psychometrika*, 17(4):401–419, 1952.

[46] J. B. Kruskal. Multidimensional scaling. *Sage*, (11), 1978.

[47] I. Dokmanic, R. Parhizkar, J. Ranieri, and M. Vetterli. Euclidean distance matrices: essential theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 32(6):12–30, 2015.

[48] P. Pertilä, M. S. Hämäläinen, and M. Mieskolainen. Passive temporal offset estimation of multichannel recordings of an ad-hoc microphone array. *IEEE transactions on audio, speech, and language processing*, 21(11):2393–2402, 2013.

[49] V. Deleskog, H. Habberstad, G. Hendeby, D. Lindgren, and N. Wahlstrom. Robust nls sensor localization using mds initialization. *in Proc. Int. Conf. Inf. Fusion*, pages 1–7, 2014.

[50] M. Parviainen, P. Pertila, and M. S. Hamalainen. Self-localization of wireless acoustic sensors in meeting rooms. *in Proc. Joint Workshop Hands-Free Speech Communication and Microphone Arrays*, page 152–156, 2014.

[51] T. K. Hon, L. Wang, J. D. Reiss, and A. Cavallaro. Audio fingerprinting for multi-device self-localization. *IEEE/ACM Trans. Audio, Speech, Language Process.*, 23(10):1623–1636, 2015.

[52] M. Chen, Z. Liu, L. W. He, P. Chou, and Z. Zhang. Energy-based position estimation of microphones and speakers for ad hoc microphone arrays. *in Proc. IEEE Workshop Appl. Signal Processing Audio Acoustics*, page 22–25, 2007.

[53] Z. Liu, Z. Zhang, L. W. He, and P. Chou. Energy-based sound source localization and gain normalization for ad hoc microphone arrays. *in Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, page 761–764, 2007.

[54] X. Sheng and Y. H. Hu. Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks. *IEEE Trans. Signal Process.*, 53(1):44–53, 2005.

[55] S. Woźniak and K. Kowalczyk. Passive joint localization and synchronization of distributed microphone arrays. *IEEE Signal Process. Lett.*, 26(2):292–296, 2018.

[56] D. Hu, Z. Chen, and F. Yin. Passive geometry calibration for microphone arrays based on distributed damped newton optimization. *IEEE/ACM Trans. Audio, Speech, Language Process.*, 29:118–131, 2020.

[57] R. Biswas and S. Thrun. A passive approach to sensor network localization. *in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, page 1544–1549, 2004.

[58] V. C. Raykar, I. V. Kozintsev, and R. Lienhart. Position calibration of microphones and loudsources in distributed computing platforms. *IEEE Trans. Speech, Audio Process.*, 13(1):70–83, 2004.

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

[60] N. D. Gaubitch, W. B. Kleijn, and R. Heusdens. Auto-localization in ad-hoc microphone arrays. *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, page 106–110, 2013.

[61] N. D. Gaubitch, W. B. Kleijn, and R. Heusdens. Calibration of distributed sound acquisition systems using toa measurements from a moving acoustic source. *2014 IEEE International Conference on Acoustics, Speech and SignalProcessing (ICASSP)*, pages 7455–7459, 2014.

[62] F. Jiang and Y. Kuang. Time delay estimation for tdoa self-calibration using truncated nuclear norm regularization. *In 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3885–3889, 2013.

[63] Y. Kuang and K. Åström. Stratified sensor network self-calibration from tdoa measurements. *21st European Signal Processing Conference*, pages 1–5, 2013.

[64] L. Wang, T. K. Hon, J. D. Reiss, and A. Cavallaro. Self-localization of ad-hoc arrays using time difference of arrivals. *IEEE Trans. Signal Process.*, 64(4):1018–1033, 2015.

[65] R. Heusdens and N. Gaubitch. Time-delay estimation for toa-based localization of multiple sensors. *2014 IEEEInternational Conference on Acoustics, Speech and Signal Processing*, pages 609–613, 2014.

[66] N. D. Gaubitch, W. B. Kleijn, and R. Heusdens. Auto-localization in ad-hoc microphone arrays. *In 2013 IEEE international conference on acoustics, speech and signal processing*, pages 106–110, 2013.

[67] S. Thrun. Affine structure from sound. *Advances in Neural Information Processing Systems*, 18, 2005.

[68] S. H. C. and A. Z. Robinson. Passive source localization employing intersecting spherical surfaces from time-of-arrival differences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 35(8):1223–1225, 1987.

[69] J. Smith and J. Abel. Passive source localization employing intersecting spherical surfaces from time-of-arrival differences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 35(12):1661–1669, 1987.

[70] Y. T. Chan and K. C. Ho. A simple and efficient estimator for hyperbolic location. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 42(8):1905–1915, 1994.

[71] L. Yang and K. C. Ho. An approximately efficient tdoa localization algorithm in closed-form for locating multiple disjoint sources with erroneous sensor positions. *IEEE Transactions on Signal Processing*, 57(12):4598–4615, 2009.

[72] M. S. Brandstein, J. E. Adcock, and H. F. Silverman. A closed-form location estimator for use with room environment microphone arrays. *IEEE transactions on Speech and Audio Processing*, 5(1):45–50, 1997.

[73] J. R. I. Dokmanic, R. Parhizkar and M. Vetterli. Euclidean distance matrices: essential theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 32(6):12–30, 2015.

[74] L. Liberti and C. Lavor. Euclidean distance geometry. *Berlin: Springer*, 3, 2017.

[75] J. C. Gower. Properties of euclidean and non-euclidean distance matrices. *Linear algebra and its applications*, 67:81–97, 1985.

[76] T. Le and N. Ono. Numerical formulae for toa-based microphone and source localization. *In 2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pages 178–182, 2014.

[77] T. Le and N. Ono. Reference-distance estimation approach for tdoa-based source and sensor localization. *In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2549–2553, 2015.

[78] M. Pollefeys and D. Nister. Direct computation of sound and microphone locations from time-difference-of-arrival data. *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2445–2448, 2008.

[79] E. Bolker and B. Roth. When is a bipartite graph a rigid framework? *Pacific Journal of Mathematics*, 90(1):27–44, 1980.

[80] W. Adams and P. Loustaunau. An introduction to gröbner bases. *American Mathematical Society*, 3, 2022.

[81] D. Grayson and M. Stillman. Macaulay2. *http://www.math.uiuc.edu/ Macaulay2/.*

[82] Y. Kuang, S. Burgess, A. Torstensson, and K. Åström. A complete characterization and solution to the microphone position self-calibration problem. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3875–3879, 2013.

[83] S. Burgess, Y. Kuang, and K. Åström. Toa sensor network calibration for receiver and transmitter spaces with difference in dimension. *21st European Signal Processing Conference (EUSIPCO 2013)*, pages 1–5, 2013.

[84] S. Burgess, Y. Kuang, and K. Åström. Toa sensor network self-calibration for receiver and transmitter spaces with difference in dimension. *Signal Processing*, 107:33–42, 2015.

[85] S. Zhayida, S. Burgess, Y. Kuang, and K. Åström. Minimal solutions for dual microphone rig self-calibration. *In 2014 22nd European Signal Processing Conference*, pages 2260–2264, 2014.

[86] J. Zhang, R. C. Hendriks, and R. Heusdens. Structured total least squares based internal delay estimation for distributed microphone auto-localization. *Proc. Int. Workshop Acoustic Signal Enhancement*, pages 1–5, 2016.

[87] X. Wang and D. Hu. Distributed self-localization for acoustic transceiver networks. *IEEE Signal Process. Lett.*, 2023.

[88] R. Biswas and S. Thrun. A passive approach to microphone network localization. *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, page 1544–1549, 2004.

[89] D. E. Badawy, V. Larsson, M. Pollefeys, and I. Dokmanic. Localizing unsynchronized sensors with unknown sources. *IEEE Trans. Signal Process.*, 71:641–654, 2023.

[90] Y. Kuang and K. Åström. Stratified microphone network self-calibration from tdoa measurements. *Proc. EUSIPCO*, pages 1–5, 2013.

[91] D. Hu, Z. Chen, and F. Yin. Geometry calibration for acoustic transceiver networks based on network newton distributed optimization. *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, 29:1023–1032, 2021.

[92] K. Batstone, G. Flood, T. Beleyur, V. Larsson, H. R. Goerlitz, M. Oskarsson, and K. Åström. Robust self-calibration of constant offset time-difference-of-arrival. *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, pages 4410–4414, 2019.

[93] https://www2.math.upenn.edu/ ryblair/math240/papers/lec1_20.pdf.

[94] *https://www.math.tamu.edu/ fnarc/psfiles/rank2005.pdf*.

[95] S. Lipschutz and M. L. Lipson. Linear algebra. *MacGraw-Hill*, 2009.

[96] T. K. Le, K. Ho, and T. H. Le. Rank properties for matrices constructed from time differences of arrival. *IEEE Trans. Signal Process.*, 66(13):3491–3503.

[97] J. Nocedal and S. Wright. Numerical optimization. *Springer Science and Business Media*, 2006.