Imperial College London
United Kingdom

Super-resolved Localisation in Multipath Environments

Hengyan Liu

Supervised by Dr. Wei Dai

The Thesis Submitted to the Electrical and Electronic Engineering Department
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy, August 2021

Abstract

In the last few decades, the localisation problems have been studied extensively. There are still some open issues that remain unresolved. One of the key issues is the efficiency and preciseness of the localisation in presence of non-line-of-sight (NLoS) path. Nevertheless, the NLoS path has a high occurrence in multipath environments, but NLoS bias is viewed as a main factor to severely degrade the localisation performance. The NLoS bias would often result in extra propagation delay, and angular bias. Numerous localisation methods have been proposed to deal with NLoS bias in various propagation environments, but they are tailored to some specific scenarios due to different prior knowledge requirements, accuracies, computational complexities, and assumptions.

To super-resolve the location of mobile device (MD) without prior knowledge, we address the localisation problem by super-resolution technique due to its favourable features, such as working on continuous parameter space, reducing computational cost and good extensibility. Besides the NLoS bias, we consider an extra array directional error which implies the deviation in the orientation of the array placement. The proposed method is able to estimate the locations of MDs and self-calibrate the array directional errors simultaneously. To achieve joint localisation, we directly map MD locations and array directional error to received signals. Then the group sparsity based optimisation is proposed to exploit the geometric consistency that received paths are originating from common MDs. Note that the super-resolution framework cannot be directly applied to our localisation problems. Because the proposed objective function cannot be efficiently solved by semi-definite programming.

Typical strategies focus on reducing adverse effect due to the NLoS bias by separating line-of-sight (LoS)/NLoS path or mitigating NLoS effect. The LoS path is well studied for localisation and multiple methods have been proposed in the literature. However, the number of LoS paths are typically limited and the effect of NLoS bias may not always be reduced completely. As a long-standing issue, the suitable solution of using NLoS path

is still an open topic for research. Instead of dealing with NLoS bias, we present a novel localisation method that exploits both LoS and NLoS paths in the same manner. The unique feature is avoiding hard decisions on separating LoS and NLoS paths and hence relevant possible error. A grid-free sparse inverse problem is formulated for localisation which avoids error propagation between multiple stages, handles multipath in a unified way, and guarantees a global convergence. Extensive localisation experiments on different propagation environments and localisation systems are presented to illustrate the high performance of the proposed algorithm compared with theoretical analysis. In one of the case studies, single antenna access points (APs) can locate a single antenna MD even when all paths between them are NLoS, which according to the authors knowledge is the first time in the literature.

Statement of originality

I herewith declare that the content of this thesis is the product of my own work under the guidance of my PhD supervisor Dr. Wei Dai. All material that is not my own work has been properly acknowledged.

Copyright Declaration

Acknowledgement

First of all, I would like to express my sincere gratitude and appreciation to my PhD supervisor, Dr. Wei Dai. I have been so lucky to have a supervisor who was always by my side and willing to give help. During the past four years, he provided me with his professional guidance, selfless support, earnest encouragement to pursue my dreams. He also gave me many opportunities to attend conferences and summer school, and to co-supervise the Master students. I have learned a lot from him, not only about how to do the academic research, but also how to be a good scholar.

I would like to thank Prof. Yuan Shen who gave me valuable advices. I would like to thank Dr. Tianyao Huang allowing me to visit Tsinghua University research in summer 2021. I also would like to thank Prof. Cong Ling who is a senior lecturer in the Department of Electrical and Electronic Engineering at Imperial College London for his valuable and constructive comments and suggestions during my Master studying.

Thanks to all my friends who made my PhD journey unforgettable. I would particularly like to thank Cheng and Junjie for their kindness and discussion on the DSTL project, Guangbin for cooperative work, Tianhong for his wise advices. I would like to thank my girlfriend, Yanlin for all her love and support.

Finally, I would like to acknowledge my parents, who shared all the ups and downs of my entire PhD study career, for their countless support and encouragement in life. Without their trust, support and love, I would not have finished this thesis.

Abbreviations

ADCG	Alternating Descent Conditional Gradient
ANM	Atomic Norm Minimization
AP	Access Point
AWGN	Additive White Gaussian Noise
CAGR	Compound Annual Growth Rate
CRB	Cramér-Rao Bound
CS	Compressed Sensing
DoA	Direction-of-Arrival
DoD	Direction-of-Departure
FIM	Fisher Information Matrix
FLS	Feasible Locations of Scatter
FLMD	Feasible Locations of Mobile Device
GPS	Global Positioning System
GSE – SC	Group Sparsity Exploitation for Self-Calibration
LoS	Line-of-Sight
LS	Least Square
MD	Mobile Device
MF	Matched Filtering
ML	Maximum Likelihood
MMV	Multiple Measurement Vectors
MuG	Multipath-Exploited and Grid-Free
NLoS	Non-Line-of-Sight

OFDM	Orthogonal Frequency Division Multiplexing
OLoS	Obstructed-Line-of-Sight
PDF	Probability Density Function
RSS	Received-Signal-Strength
RMS	Root Mean Square
RMSE	Root Mean Square Error
SDP	Semi-Definite Programming
SNR	Signal-to-Noise Ratio
TDoF	Time-Different-of-Flight
ToF	Time-of-Flight
ULA	Uniform Linear Arrays
w.r.t.	with respect to

Notations

$\mathbb{R}$	the set of real numbers
$\mathbb{C}$	the set of complex numbers
$\nabla_{\circ} C(\cdot)$	the partial derivative of the cost function $C(\cdot)$ w.r.t. $\circ$
Re	the real part of a complex number
Im	the imaginary part of a complex number
x^*	the conjugate of x
$\boldsymbol{x}^{\mathrm{T}}, \boldsymbol{X}^{\mathrm{T}}$	the transpose of vector $\boldsymbol{x}$, matrix $\boldsymbol{X}$
$\boldsymbol{x}^{\mathrm{H}}, \boldsymbol{X}^{\mathrm{H}}$	the conjugate transpose of vector $\boldsymbol{x}$, matrix $\boldsymbol{X}$
$\boldsymbol{X}[q, w]$	the [q,w]-th entry of matrix $\boldsymbol{X}$
$\boldsymbol{x}[n]$	the [n]-th entry of vector $\boldsymbol{x}$
$\mathrm{sign}(\cdots)$	the sign of a non-zero complex number, defined by $\forall z \in \mathbb{C}^*, \quad \mathrm{sign}(z) = \frac{z}{\lvert z \rvert}$
$\lfloor x \rfloor$	the largest integer less than real number x
$\odot$	the Hadamard product
$\binom{a}{b} = \frac{a!}{b!(a-b)!}$	the combination is the choice of b things from a set of a things without replacement and where order does not matter

Contents

List of Figures

Chapter 1

Localisation in Multipath Environments: Methods and Challenges

1.1 Introduction

Localisation has found many commercial, civilian, and military applications including robotics, mixed reality and emergency systems. There is a wide variety of localisation techniques, which makes localisation a lucrative research field. There is an urgent need to address the precise localisation problems caused by location-based service (LBS). The LBS is a suite of services that includes features such as indoor location, local search and information, asset tracking, and others offered via software, which helps users to access location-specific data, and other online services [1, 7]. For localisation in an outdoor environment, a Global Positioning System (GPS) is an ideal method. However, the signal from GPS is unreliable in the interior space. Since the satellite signal is seriously degraded and the localisation accuracy cannot meet people's performance requirements of indoor localisation [8, 9]. The indoor location technology facilitates the navigation and tracking

of targets and people in a building using sensory information captured by mobile devices. It has received a high amount of attention in the last decades. Therefore, we urgently need localisation methods that can perform well in various environments.

Some of the major factors driving the growth of the LBS market include high penetration of smart devices such as sensors, Internet of Things and others. The LBS industry is becoming more popular due to its numerous benefits. The global indoor LBS market size accounted for 7.0 billion dollars in 2021, and is expected to reach 19.7 billion dollars in 2026, with the projected Compound Annual Growth Rate (CAGR) of 22.9 % during the forecast period. The global LBS market size to grow from 36.35 billion dollars in 2020 to 318.64 billion dollars by 2030, at a CAGR of 24.3 % during the forecast period. The growing demand for location technology to support robotic process and lean automation, and the rising focus on Industry 4.0 smart cities and smart manufacturing, and the growing need for contact-tracing solutions due to the COVID-19 pandemic would provide opportunities for the growth of organizations in the market.

1.1.1 Multipath Environments

In the last few decades, the localisation problems have been studied extensively. There are still some open issues that remain unresolved. One of the key issues is the efficiency and preciseness of the localisation in presence of non-line-of-sight (NLoS) path. The multipath happens when the signal takes two or more paths from the mobile device (MD) to the access points (APs). As shown in Fig 1.1, a MD spreads a radio signal that directly arriving at the AP is referred as line-of-sight (LoS) path. The other parts of transmitted signal may encounter reflecting surfaces, and the signal will scatter off theses objects which is referred as NLoS path. The scattering can occur from the buildings, cars and airplanes which result in extra NLoS bias at APs. In modern wireless system, the NLoS path is commonly encountered in both indoor e.g. residential, office, etc. and outdoor e.g. urban area, shopping plaza, etc. In these scenarios, the LoS path between MD and

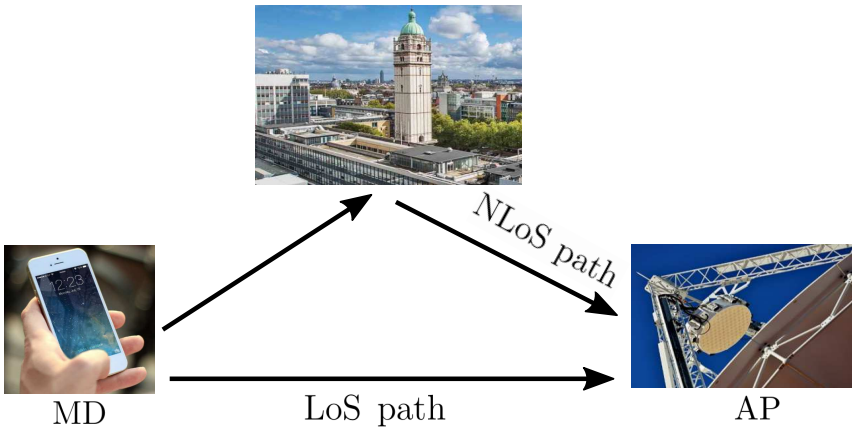


Figure 1.1: Example of multipath, a signal might bounce off several different scatters before it arrives at an AP.

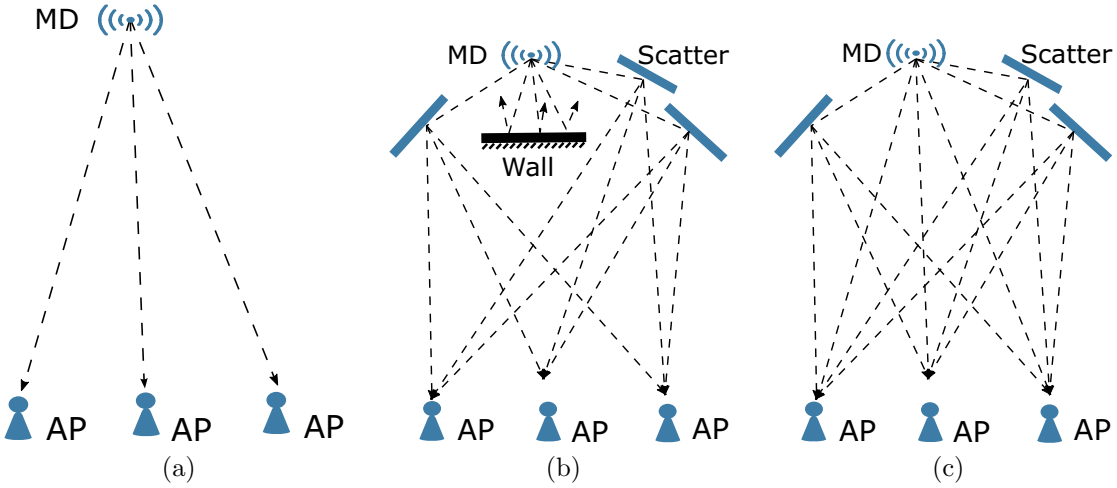


Figure 1.2: Three propagation environments: (a) LoS environment is defined as the case where only LoS paths present; (b) OLoS environment is defined as the case where all LoS paths are blocked and only signals reflected by scatters can be received; (c) In NLoS environment both LoS paths and scattered paths are present.

AP might be blocked or not. Moreover, the exact propagation environment is unknown to estimator which makes localisation problem become more difficulty to handle.

In practice, the localisation methods are tailored to some specific scenarios due to some special assumptions or the need of a priori knowledge [1–4, 10–15]. Based on type of received path, multipath environments can be divided into three categories [3] 1) the *LoS environment* where all propagation paths are LoS; 2) the *obstructed-line-of-sight (OLoS) environment* where all paths are NLoS; and 3) the *NLoS environment* where both LoS and NLoS paths are present. See Fig 1.2 for details of LoS, OLoS and NLoS environments.

In the literature, LoS paths and NLoS paths are processed by different manners. Since the theoretical study of Cramér-Rao bound (CRB) [16, 17] shows that the performance of localisation only depends on the LoS paths. Without prior knowledge, the NLoS bias carry no information for localisation, and it is assumed to be unknown and arbitrary. In order to achieve better localisation performance, this conclusion suggests to reduce adverse effect of NLoS path and only exploit LoS paths for localisation. Typical environments including residential, office, and urban area have a high occurrence of NLoS paths. In such environments, the number of LoS paths may not always be sufficient for localisation, which degrades the performance of these localisation methods [10–15,18–20]. On the other hand, the use of the GPS might be impractical. Therefore, it is critical to understand the impact of NLoS paths on localisation systems and to develop methods that exploit them for localisation.

The LoS path is well studied for localisation and a variety of methods have been proposed in the literature [10–15, 18–22]. In general, typical localisation methods involve two steps for highly accurate location awareness applications. Firstly, intermedia parameters are estimated such as time-of-flight (ToF), time-difference-of-flight (TDoF), direction-of-arrival (DoA), direction-of-departure (DoD), and received-signal-strength (RSS) techniques. Secondly, the localisation algorithms are used for estimating location of an unknown MD from known APs.

- The ToF based methods are developed based on trilateration algorithm [10, 12–15, 18–21]. The propagation distance is firstly computed based on measurements. Then location of MD is reconstructed by finding intersection of circles. The circles are centered at more than three APs (in 2-dimensional space), and the radii are equal to corresponding propagation distances.

- For the TDoF based methods [11,22], the difference between ToFs in several APs are used to estimate a MD's position. This could either be based on the difference in the times at which a single signal originating from a MD arrives at three or more APs,

or based on the difference in the times at which multiple signals originating from a MD arrive at another AP. This algorithm requires highly precise synchronisation of the APs, but not precise synchronisation between MD and APs. TDoF based methods are more commonly used in wireless sensor networks.

- For the RSS based methods [23, 24], the propagation distance is measured based on the attenuation model. The advantage is that the methods can be applied to most localisation systems with low cost. However, the distance estimation is sensitive to noise and interference. The MD mobility and unpredictable variation in channel behaviour may cause large bias in distance evaluation. As a coarse estimation method, the RSS technique is not applicable to super-resolved localisation, its application is limited to some special scenarios.

- For the DoA based methods [25], the angles of LoS paths are used for estimating location of MD. The major disadvantage of this technique is high requirement on angular resolution i.e., adoption of antenna arrays and a minimum distance between the antenna elements which results additional costs and larger antenna sizes. Furthermore, the performance of this technique is highly sensitive to directional errors i.e., deviation of orientation in array placement and NLoS bias.

The NLoS paths are viewed as a main factor to severely degrade the performance of conventional localisation methods. The adverse effect of NLoS path is referred as NLoS bias which is caused by scattering on obstacles or scatters.

- In angular localisation methods (e.g., DoA based method), the NLoS bias results in extra directional error in angular measurements, making the estimated DoA larger or less than the true value. In Chapter 3, we propose a super-resolved localisation for estimating location of MD and self-calibrating the array directional errors simultaneously. In our formulation, a group sparsity based optimisation is proposed to exploit the geometric consistency that received LoS paths are originating from

MDs. Therefore, the proposed method also can be used for reducing effect of NLoS bias by discarding all NLoS paths.

- In range measurement based methods (i.e., ToF, TDoF and RSS based methods), the NLoS bias result in longer propagation distance than the actual direct length. The location of MD may not be estimated uniquely. In Chapter 4, we propose a multipath-exploited and grid free (MuG) method to exploit both LoS and NLoS path for localisation, without dealing with NLoS bias.

1.2 Popular Localisation Methods

In this section, we provide a brief overview on ToF, DoA and DoD based localisation methods. For other techniques such as RSS and fingerprint-based methods, interested readers are referred to [23, 24, 26] and the references therein.

We focus on the following localisation problem by adopting a commonly assumed setup in the literature [1–4, 10–12, 14, 15, 27–31]. Consider wireless communication systems as shown in Fig 1.2 where J cooperative APs jointly estimate the location of a MD. The location of j-th AP is known and it is denoted by $\boldsymbol{AP}_j = [AP_j^x, AP_j^y]^{\mathrm{T}}$, where T denotes the transpose of a matrix. The unknown location of MD is denoted by $\boldsymbol{m} = [m^x, m^y]^{\mathrm{T}}$. For the sake of simplicity, this thesis will provide analysis for 2-dimensional localisation, since the extension to higher dimensions can be easily obtained. It is typically assumed that via control signalling, the MD and the involved APs are synchronised [1, 28] and the transmitted waveform from the MD is known to APs. In the signal model, we only consider either LoS or NLoS with single-bounced signals from the MD to the APs. This is motivated by the fact that signals scattered twice or more times typically suffer from great propagation losses and are thus less perceptible [4, 5]. It is noteworthy that the above setup is a simplification of actual systems. For example, the assumption of synchronisation and the complete discard of multiple-bounced signals may be problematic in practice.

Nevertheless, the above setup is widely adopted in the literature [1–4, 10–12, 14, 15, 27–31] for the purpose of highlighting the approach/idea without being drowned into great technical details.

In the literature, the common signal model of multipath is to treat the effect of reflection on range measurements as a positive stochastic NLoS bias [12]. The range measurement d_j at the j-th AP is extracted from the ToF τ_j between MD $\boldsymbol{m} = [m^x, m^y]^{\mathrm{T}}$ and the j-th AP $\boldsymbol{AP}_j = [AP^x, AP^y]^{\mathrm{T}}$

$$\begin{aligned} d_j &= c\tau_j && (1.1)\\ &= f_j(\boldsymbol{m}) + b_j + v_j && (1.2)\\ &= \left(\|\boldsymbol{m} - \boldsymbol{AP}_j\|_2 + b_j + v_j\right), && (1.3) \end{aligned}$$

where $f_j(\boldsymbol{m}) = \|\boldsymbol{m} - \boldsymbol{AP}_j\|_2$ is the true direct propagation distance between MD and the j-th AP; c is the speed of light; b_j is a positive NLoS bias introduced due to scattering on obstacles or scatters; more precisely,

$$b_j \begin{cases} = 0, & \text{if the path is LoS,} \\ > 0, & \text{if the path is NLoS,} \end{cases} \tag{1.4}$$

In the literature, there are different ways to model the NLoS bias b_j, e.g. Gaussian distributed [32, 33], constant along a time window [10], exponentially distributed [34–36] , uniformly distributed [18] , or based on an empirical model from measurements [37, 38]; $v_j \sim \mathcal{N}(0, \sigma^2)$ is the additive white Gaussian noise(AWGN) with variance σ^2.

The NLoS bias b and measurement noise v are two dominant factors of existing localising methods. As shown in Fig 1.3, the basic idea of range measurement based methods is the trilateration algorithm which requires at least three LoS paths in three participating APs (in 2-dimensional space). Each range measurement generates a circle which is centered at the measuring AP. The corresponding radius equal to the distance from MD to the

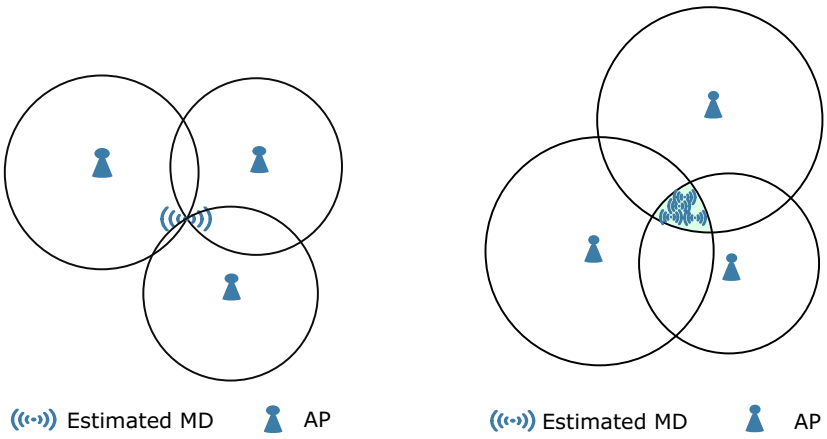


Figure 1.3: Trilateration algorithm in 2-dimensional space. (a) In case of LoS paths, the estimated MD is located at intersection point of three circles; (b) When the LoS paths are interfered by NLoS bias b or measurement noise v, the estimated MDs are located at intersection area of three circles.

measuring AP. As shown in Fig 1.3(a), in the absence of any range measurement error, the intersection of these three circles unambiguously determines the location of MD. However, when the range measurements i.e., radii of circles are increased by NLoS bias b or noise v, then the intersection area contains infinite many estimations. This adverse effect is detailed by geometrical demonstration in Fig 1.3(b). Therefore, NLoS bias b and noise v results in unreliable estimation and significantly decreases the localising accuracy if its effects are not taken into account.

1.2.1 LoS Environment

In a LoS environment, the common strategy is incorporating range measurements to determine the location of MD. Two steps are required for localising MD. The first step is obtaining measured propagation distances $\{d_j\}_{j=1}^{J}, J \geqslant 3$. Considering existing of measurement error v in the measured ranges d, the trilateration algorithm can only provide a uncertain region of possible estimations, rather than single estimation. In the second step, a variety of methods have been developed for solving this problem, including least square (LS) methods [20,39,40] and iterative statistical positioning algorithms [10,41,42].

1.2.1.1 LoS Environment: Least Square (LS) Method

The LS method solves the localisation problem by formulating a set of non-linear equations [20,39,40]. In the absence of NLoS bias (i.e. $b = 0$ for multipath), the MD can be estimated as follows,

$$\min_{\boldsymbol{m}\in\mathbb{R}^2} \sum_{j\in\mathcal{L}} \left(d_j - \|\boldsymbol{m} - \boldsymbol{AP}_j\|_2\right)^2, \tag{1.5}$$

where the set $\mathcal{L} := \{j \mid j\text{-th path is LoS}\}$ consists of LoS paths.

For the sake of simplicity, we will provide an analysis for 2-dimensional localisation and an extension to higher dimensions can be easily obtained. Let $\boldsymbol{m} = [m^x, m^y]^{\mathrm{T}}$ be the unknown MDs x- and y-coordinates and let $\boldsymbol{AP}_j = [AP_j^x, AP_j^y]^{\mathrm{T}}$ denote the coordinates of the known j-th AP, where $j \in \{1, \ldots, J\}$.

In order to obtain a LS solution, the nonlinear equation $f_j(\boldsymbol{m}) = \|\boldsymbol{m} - \boldsymbol{AP}_j\|_2$ is firstly linearised. Linearising $f(\boldsymbol{m})$ can be achieved by the first-order Taylor series expansion around reference point $\boldsymbol{m}_0$ and keeping the first two terms [43], i.e.,

$$f_j(\boldsymbol{m}) \approx f_j\left(\boldsymbol{m}_0\right) + \boldsymbol{h}_j\left(\boldsymbol{m}_0\right) \cdot \left(\boldsymbol{m} - \boldsymbol{m}_0\right), \tag{1.6}$$

where the Jacobian vector $\boldsymbol{h}_j\left(\boldsymbol{m}\right)$ around $\boldsymbol{m} = \boldsymbol{m}_0$ is given by

$$\boldsymbol{h}_j\left(\boldsymbol{m}_0\right) = \left[\begin{array}{cc} \frac{\partial f_j}{\partial m^x} & \frac{\partial f_j}{\partial m_y} \end{array} \right]_{\boldsymbol{m}=\boldsymbol{m}_0}. \tag{1.7}$$

The matrix form of range measurements in (1.3) can be approximated as follows:

$$\boldsymbol{d} \approx \boldsymbol{F}(\boldsymbol{m}_0) + \boldsymbol{H}\left(\boldsymbol{m}_0\right) \cdot \left(\boldsymbol{m} - \boldsymbol{m}_0\right) + \boldsymbol{v}, \tag{1.8}$$

where $\boldsymbol{d} = [d_1, \cdots, d_J]^{\mathrm{T}}$ is a vector form of range measurements at J APs. The matrix $\boldsymbol{F}(\boldsymbol{m}_0)$ consists of a set nonlinear equations given by

$$\boldsymbol{F}(\boldsymbol{m}_0) = \begin{bmatrix} f_1(\boldsymbol{m}_0) \\ \vdots \\ f_J(\boldsymbol{m}_0) \end{bmatrix} = \begin{bmatrix} \|\boldsymbol{m}_0 - \boldsymbol{A}\boldsymbol{P}_1\|_2 \\ \vdots \\ \|\boldsymbol{m}_0 - \boldsymbol{A}\boldsymbol{P}_J\|_2 \end{bmatrix} \tag{1.9}$$

$$= \begin{bmatrix} \sqrt{(m_0^x - AP_1^x)^2 + (m_0^y - AP_1^y)^2} \\ \vdots \\ \sqrt{(m_0^x - AP_J^x)^2 + (m_0^y - AP_J^y)^2} \end{bmatrix}. \tag{1.10}$$

The Jacobian matrix $\boldsymbol{H}(\boldsymbol{m})$ around $\boldsymbol{m} = \boldsymbol{m}_0$ is given by

$$\boldsymbol{H}(\boldsymbol{m}_0) = \begin{bmatrix} \boldsymbol{h}_1 \\ \vdots \\ \boldsymbol{h}_J \end{bmatrix}_{\boldsymbol{m}=\boldsymbol{m}_0} = \begin{bmatrix} \frac{\partial f_1}{\partial m^x} & \frac{\partial f_1}{\partial m_y} \\ \vdots & \vdots \\ \frac{\partial f_J}{\partial m^x} & \frac{\partial f_J}{\partial m^y} \end{bmatrix}_{\boldsymbol{m}=\boldsymbol{m}_0} \tag{1.11}$$

$$= \begin{bmatrix} \frac{m^x - AP_1^x}{\sqrt{\left(m^x - AP_1^x\right)^2 + \left(m^y - AP_1^y\right)^2}} & \frac{m^y - AP_1^y}{\sqrt{\left(m^x - AP_1^x\right)^2 + \left(m^y - AP_1^y\right)^2}} \\ \vdots & \vdots \\ \frac{m^x - AP_J^x}{\sqrt{\left(m^x - AP_J^x\right)^2 + \left(m^y - AP_J^y\right)^2}} & \frac{m^y - AP_J^x}{\sqrt{\left(m^x - AP_J^x\right)^2 + \left(m^y - AP_J^y\right)^2}} \end{bmatrix}_{\boldsymbol{m}=\boldsymbol{m}_0}. \tag{1.12}$$

The location of MD is estimated by minimizing the cost function given by

$$E[\hat{\boldsymbol{m}}] = [\boldsymbol{d} - \boldsymbol{F}(\boldsymbol{m}_0) - \boldsymbol{H}(\boldsymbol{m}_0) \cdot (\hat{\boldsymbol{m}} - \boldsymbol{m}_0)]^{\mathrm{T}} [\boldsymbol{d} - \boldsymbol{F}(\boldsymbol{m}_0) - \boldsymbol{H}(\boldsymbol{m}_0) \cdot (\hat{\boldsymbol{m}} - \boldsymbol{m}_0)], \tag{1.13}$$

The linearised LS solver is then given by

$$\hat{\boldsymbol{m}} = \boldsymbol{m}_0 + \left(\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{m}_0)\boldsymbol{H}(\boldsymbol{m}_0)\right)^{-1} \boldsymbol{H}^{\mathrm{T}}(\boldsymbol{m}_0)\left[\boldsymbol{d} - \boldsymbol{F}(\boldsymbol{m}_0)\right]. \tag{1.14}$$

This LS solver may introduce errors when the first-order Taylor series expansion does not accurately approximate the nonlinear function $f(\boldsymbol{m})$. In order to improve performance of

LS solution (1.14), a proper initial estimation of the unknown parameters is required, i.e., the reference point $\boldsymbol{m}_0$ should be chosen sufficiently close to the true MD location. With a random initial estimation of the unknown parameters, the LS solution may converge to a local optimal solution, instead of a global optimal solution.

To obtain a more accurate estimation, an iterative strategy is applied to LS solver (1.14). At the $(i+1)$-th iteration, the estimation at the i-th iteration is substituted into (1.8) to re-linearise the system around it, then iterate each successive estimate being closer to the optimal estimation [44],

$$\hat{\boldsymbol{m}}_{i+1} = \hat{\boldsymbol{m}}_i + \left(\boldsymbol{H}^{\mathrm{T}}\left(\hat{\boldsymbol{m}}_i\right)\boldsymbol{H}\left(\hat{\boldsymbol{m}}_i\right)\right)^{-1}\boldsymbol{H}^{\mathrm{T}}\left(\hat{\boldsymbol{m}}_i\right)\left[\boldsymbol{d} - \boldsymbol{F}\left(\hat{\boldsymbol{m}}_i\right)\right]. \tag{1.15}$$

The iteration can be terminated by some criterion. For example, the algorithm can be stopped after a maximum number of iterations has been performed. Alternatively, given a small threshold ϵ, the iterative algorithm must stop if $\left|E\left(\hat{\boldsymbol{m}}_{i+1}\right) - E\left(\hat{\boldsymbol{m}}_i\right)\right| < \epsilon$.

1.2.1.2 LoS Environment: Maximum Likelihood (ML) Method

The common hypothesis used in statistical method is assuming that the measurement noise v is a zero mean Gaussian noise [33]. Then the conditional probability density function (PDF) of range measurements $\boldsymbol{d} = [d_1, \cdots, d_J]^{\mathrm{T}}$ given estimated $\boldsymbol{m} = [m^x, m^y]^{\mathrm{T}}$ can be expressed as follows [10, 41, 42]:

$$P(\boldsymbol{d} \mid \boldsymbol{m}) = \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left\{-\frac{\left(d_j - f_j(\boldsymbol{m})\right)^2}{2\sigma_i^2}\right\} \tag{1.16}$$

$$= \frac{1}{\sqrt{(2\pi)^J \det(\boldsymbol{Q})}} \exp\left\{-\frac{\boldsymbol{B}}{2}\right\}, \tag{1.17}$$

where

$$\boldsymbol{B} = [\boldsymbol{d} - \boldsymbol{F}(\boldsymbol{m})]^{\mathrm{T}}\boldsymbol{Q}^{-1}[\boldsymbol{d} - \boldsymbol{F}(\boldsymbol{m})], \tag{1.18}$$

and the $\boldsymbol{Q}$ is given by

$$\boldsymbol{Q} = \mathrm{E}\left[\boldsymbol{v}\boldsymbol{v}^T\right] = \mathrm{diag}\left[\sigma_1^2, \sigma_2^2, \ldots, \sigma_J^2\right]^{\mathrm{T}}. \tag{1.19}$$

Then the ML solution for $\hat{\boldsymbol{m}}$ is determined by maximising the PDF $P(\boldsymbol{d} \mid \boldsymbol{m})$

$$\hat{\boldsymbol{m}} = \arg\max_{\boldsymbol{m}} P(\boldsymbol{d} \mid \boldsymbol{m}) \tag{1.20}$$

Solving (1.20) requires a search over all possible MD locations which is computational expensive.

1.2.2 NLoS Environment

The mixture of LoS and NLoS path often occurs in an urban or indoor environment. As introduced in previous section, the performance of localisation methods is effected by NLoS bias due to obstacles in the direct paths. NLoS bias results in unreliable localisation and significantly decreases the location accuracy if its effects are not taken into account. In the view of the CRB, NLoS bias does not contribute useful information for localisation [16, 17]. Therefore, the optimal strategy is to reduce the adverse effect of NLoS bias as much as possible, and exploit information carried by LoS paths to infer the location of MD. Some localisation methods that cope with the existence of NLoS bias have been proposed in [10–15, 18–20]. In those works, there are two categories to deal with the localisation problem in the presence of NLoS path:

* *LoS/NLoS separation*: These methods focus on separating LoS/NLoS paths, and estimating the location of MD by using LoS paths. Note that the NLoS paths are either discarded [10–14, 18] or used for restricting the feasible region [15, 19].

* *NLoS mitigation*: The second category is typically adopted to mitigate of the adverse effect NLoS paths by estimating NLoS biases or assigning proper weighted factors

[20, 36, 45]. Note that most methods assume that NLoS paths have been separated or statistical information is available.

Remark that the LoS/NLoS separation play a curial role in both two categories.

1.2.2.1 LoS/NLoS separation

In order to reduce adverse effect of NLoS bias, the optimal strategy is to discard NLoS paths and only use LoS paths for localising MD. The LoS/NLoS separation can be achieved by statistical approach or geometrical approach.

- Under the statistical methods, the NLoS bias is considered as a random variable that can be constant [10], a random process of a Gaussian [11, 13, 14], or uniform distribution [18] with given parameters. In [12, 46], the LoS/NLoS path is separated by comparing the estimated variance of measurement with the prior historical information of LoS/NLoS path. To build an accurate statistical model, these methods require prior information about the distribution of NLoS bias or historical information.

- In the geometrical method, the geometric relationship between the MD and AP is exploited to separate LoS paths based on measured time-of-flight (ToF) or direction-of-arrival (DoA). In [47], the method exploits the fact that LoS paths typically arrive with a shorter ToF than NLoS paths. In [15], a compressed sensing (CS) based framework is proposed for separating LoS path by exploiting the fact that DoAs of LoS paths must originate from the same MD, but DoAs of NLoS paths is arbitrary.

In some situations, the statistical information of NLoS bias may be obtained a priori based on surveyed data [10–14, 18, 46]. As a consequence, the statistical method is widely adopted to separate LoS paths or NLoS paths from multipath. In [10], the ML detector is based on historical information which assumes T ToF measurements are available for

each one of J APs. The common model of the t-th ToF measurement ($t = 1, ..., T$) at the j-th AP is expressed as

$$\tau_{j,t} = f_j(\boldsymbol{m})/c + b_j + v_j \tag{1.21}$$

$$= \left(\|\boldsymbol{m} - \boldsymbol{AP}_j\|_2\right)/c + b_j + v_j, \tag{1.22}$$

where the NLoS bias b_j is assumed as a constant along the time window [10], and v_j is a zero-mean Gaussian noise independent in time and independent for each AP. In case of NLoS path at the j-th AP, the conditional PDF over T ToF measurements can be formulated as follows:

$$p_j^{\mathrm{NLoS}}\left(\boldsymbol{\tau}_j \mid \boldsymbol{m}, b_j\right) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{1}{2}\left(\frac{\tau_{j,t} - f_j(\boldsymbol{m})/c - b_j}{\sigma_j}\right)^2}, \tag{1.23}$$

where $\boldsymbol{\tau}_j = [\tau_{j,1} \dots \tau_{j,T}]^{\mathrm{T}}$. Similarly, the conditional PDF of LoS path at the j-th AP over T ToF measurements can be formulated as:

$$p_j^{\mathrm{LoS}}\left(\boldsymbol{\tau}_j \mid \boldsymbol{m}, b_j\right) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{1}{2}\left(\frac{\tau_{j,t} - f_j(\boldsymbol{m})/c}{\sigma_j}\right)^2}. \tag{1.24}$$

The LoS/NLoS path is further separated by different hypothesis based on PDFs in (1.23), (1.24). The number of hypothesis (N^{hyp}) depends on the number of APs (J) for localisation and the maximum number of APs (H^{NLoS}, for $H^{\mathrm{NLoS}} \leqslant J$) suffering from NLoS paths. The number of hypothesis characterised by the subset of APs with NLoS paths are

$$N^{\mathrm{hyp}} = \binom{J}{0} + \cdots + \binom{J}{H^{\mathrm{NLoS}}}, \tag{1.25}$$

where the operator $\binom{a}{b} = \frac{a!}{b!(a-b)!}$ is the choice of b things from a set of a things without replacement and where order does not matter; H^{NLoS} is the maximum number of APs suffering the NLoS paths. Note that the computational cost of the algorithm is growing

with the number of hypothesis N^{hyp}.

In the l-th hypothesis, the subset of APs with LoS paths can be characterised by S_l^{LoS}, and the subset of APs with NLoS paths can be characterised by S_l^{NLoS}. The conditional PDF associated with JT ToF measurements can be expressed as:

$$q_l\left(\boldsymbol{\tau} \mid \boldsymbol{m}, \boldsymbol{b}_{S_l^{\text{NLoS}}}\right) = \prod_{j \in S_l^{\text{NLoS}}} p_j^{\text{NLoS}}\left(\boldsymbol{\tau}_j \mid \boldsymbol{m}, b_j\right) \prod_{j \in S_l^{\text{LoS}}} p_j^{\text{LoS}}\left(\boldsymbol{\tau}_j \mid \boldsymbol{m}\right), \tag{1.26}$$

where $\boldsymbol{\tau} = [{\boldsymbol{\tau}_1}^{\text{T}}, \ldots, {\boldsymbol{\tau}_J}^{\text{T}}]^{\text{T}}$ is the matrix form of JT ToF measurements and $\boldsymbol{b}_{S_l^{\text{NLoS}}}$ denotes vector form of NLoS bias of the NLoS paths in the l-th hypothesis.

In (1.26), the value of conditional PDF is determined by different subset of unknown parameters: $\boldsymbol{m}, \boldsymbol{b}_{S_l^{\text{NLoS}}}$. The first step is maximising the l-th hypothesis PDF with respect to corresponding unknown parameters:

$$q_l^{ML}(\boldsymbol{\tau}) = \max_{\boldsymbol{m}, \boldsymbol{b}_{S_l^{\text{NLoS}}}} q_l\left(\boldsymbol{\tau} \mid \boldsymbol{m}, \boldsymbol{b}_{S_l^{\text{NLoS}}}\right), \tag{1.27}$$

where $\boldsymbol{m}$ is location of the MD, and $\boldsymbol{b}_{S_l^{\text{NLoS}}}$ is vector form of NLoS bias in the l-th hypothesis.

Then LoS/NLoS separation can be achieved by selecting proper hypothesis index $\widehat{l}$:

$$\widehat{l} = \arg\max_l \gamma_l \cdot q_l^{ML}(\boldsymbol{\tau}), \tag{1.28}$$

where γ_l is constant assigned based on prior probability of the hypothesis. Without knowledge of the probability, the γ_l is chosen as $\gamma_l = 1, \forall l$.

1.2.2.2 NLoS Mitigation

Discarding NLoS paths may not be a viable option, as the number of available LoS paths may be limited due to obstacles. To solve this problem, NLoS mitigation methods consider

another strategy to deal with NLoS bias [36, 45]. The effect of NLoS bias is reduced by estimating the bias or assigning weights. The performance is depending on how much prior information is available, and the estimation may be unreliable in presence of NLoS bias, since the effect can not be reduced completely.

Given different prior statistics information, different NLoS mitigation methods are selected for localisation [36]. For the ease of demonstration, in this section, the NLoS bias is assumed exponentially distributed with the parameter b_j, therefore, the mean is b_j and variance is b_j^2 , and the measurement noise v_j at the $j-$th AP is Gaussian with zero mean and variance σ_j^2.

Known LoS/NLoS Status and distribution Parameter

As we mentioned in Section 1.2.2, LoS/NLoS separation play an important role in localisation. In NLoS environment, the ideal case that the NLoS paths have been separated from LoS paths and distribution parameters are known. Without loss of generality, given J APs, the set of APs with LoS paths is denoted by S^{LoS}, the complementary set of APs are assumed to be suffering from NLoS paths S^{NLoS}. The corresponding conditional PDF of range measurements $\boldsymbol{d} = [d_1, \cdots, d_J]^{\mathrm{T}}$ given estimated $\hat{\boldsymbol{m}} = [\hat{m}^x, \hat{m}^y]^{\mathrm{T}}$ can be expressed as follows

$$p(\boldsymbol{d} \mid \boldsymbol{m}) = \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi}\tilde{\sigma}_j} \exp\left(-\frac{\left(\tilde{d}_j - f_j(\boldsymbol{m})\right)^2}{2\tilde{\sigma}_j^2} \right) \tag{1.29}$$

where

$$\tilde{\sigma}_j = \begin{cases} \sigma_j, & j \in S^{\mathrm{LoS}} \\ \sqrt{\sigma_j^2 + b_j^2}, & j \in S^{\mathrm{NLoS}} \end{cases} \tag{1.30}$$

and

$$\tilde{d}_j = \begin{cases} d_j, & j \in S^{\mathrm{LoS}} \\ d_j - b_j, & j \in S^{\mathrm{NLoS}} \end{cases} \tag{1.31}$$

The log likelihood can be obtained by taking logarithm on both sides of (1.29) and ignoring the irrelevant constants,

$$\Lambda(\boldsymbol{d} \mid \boldsymbol{m}) = -\sum_{j=1}^{J} \frac{\left(\tilde{d}_i - f_j(\boldsymbol{m})\right)^2}{2\tilde{\sigma}_i^2}. \tag{1.32}$$

The location of MD can be determined by maximising the log likelihood in (1.32). Equivalently, the ML location estimate is obtained by

$$\hat{\boldsymbol{m}} = \arg\max_{\boldsymbol{m}} (\Lambda(\boldsymbol{d} \mid \boldsymbol{m})). \tag{1.33}$$

Known NLoS Status and σ_j But Unknown Parameter b_j

In practical scenarios, some of the parameters might be unknown for the solver. Assuming that LoS or NLoS paths have been separated (known NLoS status) and variance of measurement noise σ_j^2 is known, but distribution parameter b_j is unknown. With known NLoS path set S^{NLoS}, the unknown parameter b_j can be bounded as

$$0 \leq b_j \leq d_j, \quad j \in S^{\mathrm{NLoS}}, \tag{1.34}$$

where d_j is the range measurement at the j-th AP; S^{NLoS} denotes the set of APs with NLoS paths.

Let $\boldsymbol{b} = \begin{bmatrix} \cdots & b_j & \cdots \end{bmatrix}^{\mathrm{T}}$, $j \in S^{\mathrm{NLoS}}$ denote parameter of NLoS bias. Similarly as (1.33),

the location estimate can be determined by maximising the log likelihood given by

$$\Lambda(\boldsymbol{d} \mid \boldsymbol{m}, \boldsymbol{b}) = -\sum_{j=1}^{J} \frac{\left(\tilde{d}_j - f_j(\boldsymbol{m})\right)^2}{2\tilde{\sigma}_i^2}, \tag{1.35}$$

w.r.t. $\boldsymbol{m}$, $\boldsymbol{b}$. The location of estimated MD $\hat{\boldsymbol{m}}$ can be determined by

$$\begin{aligned} \hat{\boldsymbol{m}} &= \arg\max_{\boldsymbol{m},\boldsymbol{b}} \left(\Lambda\left(\boldsymbol{d} \mid \boldsymbol{m}, \boldsymbol{b}\right)\right) \\ &\text{s.t. } 0 \leq b_j \leq d_j, \quad j \in S^{\mathrm{NLoS}}. \end{aligned} \tag{1.36}$$

Known NLoS Probability P^{NLoS_j} and Distribution Parameters b_j

With known NLoS probability P^{NLoS_j} and distribution parameters b_j, the joint conditional PDF at the j-th AP can be expressed as

$$p'_j = P^{\mathrm{NLoS}_j} p_{b_j} + \left(1 - P^{\mathrm{NLoS}_j}\right) p_j, \tag{1.37}$$

where P^{NLoS_j} is the probability that the j-th AP has a NLoS path; p_{b_j} is the Gaussian density function with mean b_j and variance $\tilde{\sigma}_j^2 = \sigma_j^2 + b_j^2$,

$$p_{b_j} = \frac{1}{\sqrt{2\pi}\tilde{\sigma}_j} \exp\left(-\frac{(d_j - f_j(\boldsymbol{m}) - b_j)^2}{2\tilde{\sigma}_j^2}\right); \tag{1.38}$$

p_j is the Gaussian density function with zero mean and variance σ_j^2,

$$p_j = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(d_j - f_j(\boldsymbol{m}))^2}{2\sigma_j^2}\right). \tag{1.39}$$

Then, the joint conditional PDF over J APs is given by:

$$p(\boldsymbol{d} \mid \boldsymbol{m}) = \prod_{j=1}^{J} p'_j, \tag{1.40}$$

and the log likelihood of the measurements is given by:

$$\Lambda(\boldsymbol{d} \mid \boldsymbol{m}) = \sum_{j=1}^{J} \ln p'_j. \tag{1.41}$$

Finally, the location of MD is estimated by

$$\hat{\boldsymbol{m}} = \arg\max_{\boldsymbol{m}} \left(\Lambda\left(\boldsymbol{d} \mid \boldsymbol{m}\right)\right). \tag{1.42}$$

Known NLoS Probability P^{NLoS_j} and Variance σ_j^2 But Unknown Distribution Parameters b_j

A more general case is that both probability of NLoS path P^{NLoS_j} and noise variance σ_j^2 are known based on historical data. However, LoS paths or NLoS paths have not been separated, and the distribution parameter of NLoS path b_j is unknown. Similar as (1.34), the distribution parameter b_j can be bounded. The location of MD can be estimated by maximising the log likelihood in (1.42) with respect to unknown location $\boldsymbol{m}$ and distribution parameter b_j:

$$\hat{\boldsymbol{m}} = \arg\max_{\boldsymbol{m},\boldsymbol{b}} \Lambda\left(\boldsymbol{d} \mid \boldsymbol{m}, \boldsymbol{b}\right) \tag{1.43}$$

$$= \arg\max_{\boldsymbol{m},\boldsymbol{b}} \left(\sum_{j=1}^{J} \ln p'_j\right). \tag{1.44}$$

Without Any Prior Statistical Information

In most practical scenarios, it is very difficult to separate LoS or NLoS paths from multipath environments since prior statistical information of NLoS bias may be unavailable. These NLoS paths will severely degrade the accuracy of localisation, since they have large positive biases which make the measured propagation distances much larger than the true values. To deal with NLoS bias without any prior information, one of NLoS mitiga-

tion methods is jointly estimating location of MD and NLoS biases. The corresponding objective function is formulated as

$$\begin{aligned} \min_{\boldsymbol{m}\in\mathbb{R}^2,\{b_j\}_{j=1}^J} & \sum w_j \left(d_j - f_j(\boldsymbol{m}) - b_j\right)^2 \\ \text{s.t. } & f_j(\boldsymbol{m}) = \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{P}_j\|_2, \\ & b_j > 0, \end{aligned} \tag{1.45}$$

where $\{w_j\}_{j=1}^J$ are the weighting elements for reducing effect of NLoS path and emphasizing more reliable LoS paths; Without any prior information, w_j can be chosen as one.

In [36], the author proposed a Taylor-series expansion based linear quadratic programming for the problem (1.45). In order to have LS solution, the function $f_j(\boldsymbol{m}) = \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{P}_j\|_2$ is expanded in the first order Taylor series at the initial point $\boldsymbol{m}_0$ which is expressed as:

$$f_j(\boldsymbol{m}) \approx f_j\left(\boldsymbol{m}_0\right) + \boldsymbol{h}_j(\boldsymbol{m}_0) \cdot \left(\boldsymbol{m} - \boldsymbol{m}_0\right), \tag{1.46}$$

where the Jacobian vector $\boldsymbol{h}_j(\boldsymbol{m})$ around $\boldsymbol{m} = \boldsymbol{m}_0$ is given by

$$\boldsymbol{h}_j(\boldsymbol{m}_0) = \left[\begin{array}{cc} \frac{\partial f_j}{\partial m^x} & \frac{\partial f_j}{\partial m_y} \end{array} \right]_{\boldsymbol{m}=\boldsymbol{m}_0}. \tag{1.47}$$

The matrix form of range measurements in (1.3) can be approximated as follows:

$$\boldsymbol{d} \approx \boldsymbol{F}(\boldsymbol{m}_0) + \boldsymbol{H}(\boldsymbol{m}_0) \cdot (\boldsymbol{m} - \boldsymbol{m}_0) + \boldsymbol{b} + \boldsymbol{v}, \tag{1.48}$$

where $\boldsymbol{d} = [d_1, \cdots, d_J]^{\mathrm{T}}$ is a vector form of range measurements between the MD and APs; The vector $\boldsymbol{b} = [b_1, \cdots, b_J]^{\mathrm{T}}$ contains NLoS bias for multipath; The $\boldsymbol{F}(\boldsymbol{m}_0)$ contains

a set of nonlinear equations given by

$$\boldsymbol{F}(\boldsymbol{m}_0) = \begin{bmatrix} f_1(\boldsymbol{m}_0) \\ \vdots \\ f_J(\boldsymbol{m}_0) \end{bmatrix} = \begin{bmatrix} \sqrt{(m_0^x - AP_1^x)^2 + (m_0^y - AP_1^y)^2} \\ \vdots \\ \sqrt{(m_0^x - AP_J^x)^2 + (m_0^y - AP_J^y)^2} \end{bmatrix}; \tag{1.49}$$

The Jacobian matrix $\boldsymbol{H}(\boldsymbol{m})$ around $\boldsymbol{m} = \boldsymbol{m}_0$ is given by

$$\boldsymbol{H}(\boldsymbol{m}_0) = \begin{bmatrix} \boldsymbol{h}_1 \\ \dots \\ \boldsymbol{h}_J \end{bmatrix}_{\boldsymbol{m}=\boldsymbol{m}_0} = \begin{bmatrix} \frac{\partial f_1}{\partial m^x} & \frac{\partial f_1}{\partial m_y} \\ \dots & \dots \\ \frac{\partial f_J}{\partial m^x} & \frac{\partial f_J}{\partial m^y} \end{bmatrix}_{\boldsymbol{m}=\boldsymbol{m}_0}. \tag{1.50}$$

Note that the problem (1.45) is a constrained nonlinear underdetermined problem, in which there might be infinite number of solutions and finding its solution is computationally difficult. In [36], the unknown parameters are approximately solved by using one parameter b_s to represent all unknown NLoS biases $\{b_j\}_{j=1}^J$. As explained by author, variable b_s plays the role as a "balancing" parameter to partially reduce the effect of NLoS biases. Obviously, the localisation performance is degraded by one NLoS constant b_s, and this one NLoS-bias-based method can be used to tackle localisation problem with similar biases. Then the measurement in (1.48) can be expressed as:

$$\boldsymbol{d}_L = \boldsymbol{A}\boldsymbol{z} + \boldsymbol{v} \tag{1.51}$$

where $\boldsymbol{d}_{\boldsymbol{m}_0} = \boldsymbol{d} - \boldsymbol{F}(\boldsymbol{m}_0) - \boldsymbol{H}(\boldsymbol{m}_0)$, and $\boldsymbol{A}$ is written as:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{h}_1 & 1 \\ \vdots & \vdots \\ \boldsymbol{h}_J & 1 \end{bmatrix}, \tag{1.52}$$

and $\boldsymbol{z}$ is defined as the unknown parameters vector as follows:

$$\boldsymbol{z} = \begin{bmatrix} \boldsymbol{m}^{\mathrm{T}} & b_s \end{bmatrix}^{\mathrm{T}}, \tag{1.53}$$

The location of MD and bias are estimated by minimizing the cost function given by

$$E[\boldsymbol{z}] = [\boldsymbol{d}_{\boldsymbol{m}_0} - \boldsymbol{A}\boldsymbol{z}]^{\mathrm{T}}[\boldsymbol{d}_{\boldsymbol{m}_0} - \boldsymbol{A}\boldsymbol{z}], \tag{1.54}$$

The linearised LS solution is then given by

$$\hat{\boldsymbol{z}} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\boldsymbol{d}_{\boldsymbol{m}_0} \tag{1.55}$$

This algorithm introduces errors when the linearised function does not accurately approximate the original nonlinear function. In order to make (1.46) valid, it requires an initial estimate of the unknown parameters, i.e., the reference point $\boldsymbol{m}_0$ should be chosen sufficiently close to the true location of MD. With a random initial estimate of the unknown parameters, this algorithm may converge to a local optimal solution, instead of a global optimal solution. On the other hand, the problem is simplified by replacing $\{b_j\}_{j=1}^{J}$ with one NLoS bias b_s which would result in performance degradation.

1.2.3 OLoS Environment

Reducing effect of NLoS bias may not be a viable option, as the prior knowledge may be limited. In order to exploit NLoS paths for localisation without dealing with NLoS bias, existing methods are proposed under some special settings. In order to eliminate consequence of different effect of different path and use multipath for localisation, the methods in [1–4, 48] require extra angle information besides ToF information. These methods follow the common two-step localisation strategy, first the intermediate parameters, such as ToF, DoA and DoD [1–4], then finding the MD by various localisation methods. In

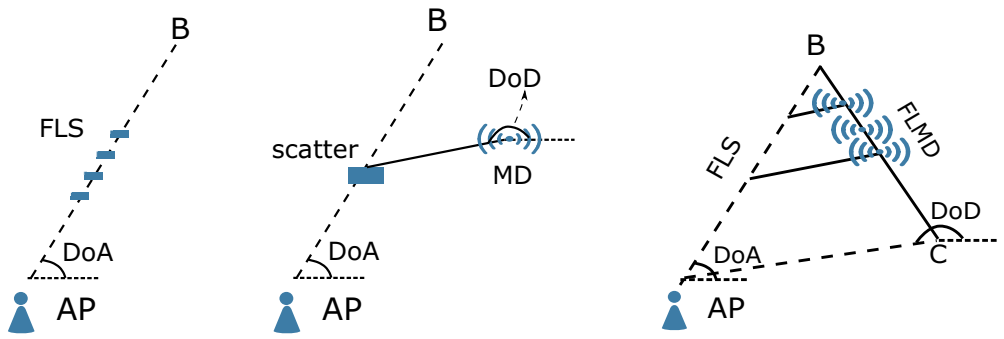


Figure 1.4: The basic idea behind [1–5] is interpreted into three steps: (a) the FLS is determined by ToF τ and DoA θ information which contains infinite scatter estimations; (b) given one estimated scatter and corresponding DoD α information, one estimated MD can be further determined; (c) As a result, the FLMD determined by ToF τ, DoA θ and DoD α which is the straight line between B and C. The exact location can then be determined by intersecting of more than two FLMDs.

the literature, the location of MD is determined by ML non-linear programming [1], LS algorithm [2, 4], and combination of ML and LS [3]. Note that [1–4] assume that the parameters are obtainable at either the MD or the AP, and the parameters associated with different propagation paths are assumed known. In order to address data association problem, [5] proposes a novel ML estimator to directly localise MD with prior on the number of multipath, but the problem cannot be solved efficiently. This comes at the cost of increased computational complexity, since the many unknown parameters including MD location, DoD components and DoA components need to be jointly estimated.

The basic idea behind [1–5] is interpreted into three steps. See Fig 1.4 for geometrical demonstration. As shown in Fig 1.4(a), given measured ToF and DoA, the feasible locations of the scatter (FLS) can be determined which is along the line between AP and B. Next step is shown in Fig 1.4(b). For each feasible scatter, the location of MD can be further determined with assistance of DoD information. The FLS contains infinite number of scatter estimations which results in the feasible locations of the MD (FLMD) lie on the piecewise line BC, see Fig 1.4(c) for demonstration. The exact location can then be determined by intersecting of more than two FLMDs.

Mathematically, the coordinate of the point B (B^x, B^y) in Fig 1.4 can be denoted by

$$B^x = AP^x + c\tau\cos\theta \tag{1.56}$$

$$B^y = AP^y + c\tau\sin\theta, \tag{1.57}$$

where θ is DoA of received path, and τ is ToF. Similarly, the coordinate of the point C (C^x, C^y) can be denoted by

$$C^x = AP^x - c\tau\cos\alpha \tag{1.58}$$

$$C^y = AP^y - c\tau\sin\alpha, \tag{1.59}$$

where α is the DoD of path. Based on the equations (1.56) - (1.59), the equation of the FLMD in Fig. 1.4 can be expressed as:

$$am^x + bm^y = l, \tag{1.60}$$

where

$$a = \sin\theta + \sin\alpha, \tag{1.61}$$

$$b = -\left(\cos\theta + \cos\alpha\right) \tag{1.62}$$

$$l = AP^x\left(\sin\theta + \sin\alpha\right) - AP^y\left(\cos\theta + \cos\alpha\right) + c\tau\left(\sin\alpha\cos\theta - \cos\alpha\sin\theta\right) \tag{1.63}$$

Given J APs, we use the index j, k denoting parameter at k-th signal path of the j-th AP. Then the matrix form of equation (1.60) is expressed as:

$$\boldsymbol{Am} = \boldsymbol{L}, \tag{1.64}$$

where

$$\boldsymbol{A} = \begin{bmatrix} \sin\theta_{11} + \sin\alpha_{11} & -(\cos\theta_{11} + \cos\alpha_{11}) \\ \vdots & \vdots \\ \sin\theta_{JK} + \sin\alpha_{JK} & -(\cos\theta_{JK} + \cos\alpha_{JK}) \end{bmatrix} \tag{1.65}$$

$$\boldsymbol{m} = \begin{bmatrix} m^x & m^y \end{bmatrix}^T \tag{1.66}$$

$$\boldsymbol{L} = \begin{bmatrix} AP_1^x\left(\sin\theta_{11} + \sin\alpha_{11}\right) - AP_1^y\left(\cos\beta_{11} + \cos\alpha_{11}\right) \\ +c\tau_{11}\left(\sin\alpha_{11}\cos\theta_{11} - \cos\alpha_{11}\sin\theta_{11}\right) \\ \vdots \\ AP_J^x\left(\sin\theta_{JK} + \sin\alpha_{JK}\right) - AP_J^y\left(\cos\beta_{JK} + \cos\alpha_{JK}\right. \\ \left.+c\tau_{JK}\left(\sin\alpha_{JK}\cos\theta_{JK} - \cos\alpha_{JK}\sin\theta_{JK}\right)\right. \end{bmatrix} \tag{1.67}$$

Considering the additive measurement noise $\boldsymbol{v}$, the matrix form can be expressed as:

$$\boldsymbol{Am} + \boldsymbol{v} = \boldsymbol{L}, \tag{1.68}$$

In order to estimate location of MD, the objective function can be formulated as

$$\min_{\boldsymbol{m}} \| \boldsymbol{L} - \boldsymbol{Am}\|^2 \tag{1.69}$$

The above approach works when there are more than two paths with measured parameters for localisation.

1.3 Challenges in Multipath Environment

The localisation performance of LoS/NLoS separation based methods depends on the ability to separate LoS/NLoS paths. In case of perfect separation, these methods are able to provide well estimations that close to theoretical bound. For instance, the LoS paths based method in [10] can localise MD in meter level. As for NLoS mitigation, the localisation performance of methods depends on how much prior information is available. As shown in [36], given less statistical information, the precision of localisation is around meter level. Given the detail statistical information, its precision is improved to decimeter which is close to CRB. Therefore, the difficulty of localisation is increased by assuming that neither LoS/NLoS separation information nor statistical characteristics of NLoS bias is available. The technical difficulties of existing methods to deal with NLoS bias are as following.

1. As for LoS/NLoS separation, these methods are effective when sufficient LoS paths are available for localisation. In the urban or indoors scenarios, the number of LoS paths is typically limited. There is no guarantee that the a LoS path always exists from the MD to a particular AP, which degrades the performance of these localisation algorithms.

2. As for NLoS mitigation, the performance is depending on how much prior information is available, and the estimation may be unreliable because the presence of NLoS bias, since complete reduction of the NLoS bias effect may be impractical.

3. The localisation performance depends on the ability to accurately separate LoS/NLoS paths. However, the mistaken separation cannot be avoided completely. There is always the possibility that NLoS path is incorrectly separated as a LoS path and vice versa, which may significantly degrade the localisation accuracy [49].

4. In practical application, the prior statistical information may not always be available at AP. There is extra latency incurred due to collection of range estimates to

establish a historical information. On the other hand, the statistics of NLoS bias b are not necessarily identically distributed. In indoor scenario, the NLoS biases have different variances and means due to propagation distance and condition of scatter.

5. Generally it is difficult to determine the joint probability distributions of the features required by many statistical approaches.

Therefore, the efforts for finding an efficient method to incorporate both LoS and NLoS paths for localisation are still needed.

1.4 Contributions of the Thesis

This thesis focuses on super-resolved localisation algorithms and the main contributions are as follows:

1. A super-resolution based convex optimisation is formulated for localising MDs. Then, we provide primal and dual solvers for the localisation problem. An example is provided to show how to transform a received signal into superposition of exponentials, which can be recovered by super-resolution technique. Numerical experiment is performed to illustrate the preciseness of the proposed method. Details are provided in Chapter 2.

2. We propose a novel super-resolved localisation method that can estimate multiple MDs and perform self-calibration to correct array directional errors simultaneously. To achieve joint localisation, we directly map MD locations and array directional errors to received signals. Then the group sparsity based optimisation is proposed to exploit the geometric consistency. Unlike the NLoS paths, the received LoS paths at different APs are originating from common MDs. Thus, the proposed method also can be applied to separate LoS/NLoS paths. Details are provided in Chapter 3.

3. Localisation is challenging in present of NLoS path. Traditional methods focus on dealing with NLoS bias. We propose a novel localisation method that can exploit both LoS and NLoS paths for localisation without dealing with NLoS bias. Another unique feature is avoiding hard decisions on separating LoS/NLoS path and hence relevant possible error. A grid-free sparse inverse problem is formulated for localisation which avoids error propagation between multiple stages, handles multipath in a unified way, and guarantees a global convergence. Extensive localisation experiments on different propagation environments and localisation systems are presented to illustrate the high performance of the proposed algorithm compared with theoretical analysis. Especially, in one of the case studies, single antenna AP can locate a single antenna MD even when all paths are NLoS, which according to the authors knowledge is the first time in the literature. Details are provided in Chapter 4.

1.5 Organization of the Thesis

The rest of the thesis is organized as follows. In Chapter 2, we introduce the super-resolved localisation method based on the modern super-resolution technique. In Chapter 3, we discuss a novel super-resolved localisation method with self-calibrating the array directional errors. In Chapter 4, we propose an end-to-end multipath-exploited and grid-free (MuG) framework which can be directly applied to various system configuration and propagation environments. In Chapter 5, we discuss results of the present work and gives potential future research directions of the localisation.

1.6 Publications

The material presented in this thesis has led to the following publications:

1. H. Liu, W. Dai, and Y. Shen, " MuG: A Multipath-Exploited and Grid-free Localisation

Method," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE, 2021, pp.4665-4669.

2. H. Liu, W. Dai, and Y. Shen, " Super-resolved Localisation in Multipath Environment," (submitted to IEEE Transactions on Cybernetics)

3. H. Liu, W. Dai, and Y. Shen, " Super-resolved Localisation without Identifying LoS/NLoS Paths," *arXiv preprint arXiv: 1910.12662, 2019.*

Chapter 2

Super-Resolved Localisation

In this Chapter, we briefly describe the background of the super-resolution technique and it's application in localisation problem. Unlike previous compressed sensing (CS) works, the unknown parameters of super-resolution technique are not assumed to locate on pre-defined grids, but can be any values in the continuous parameter space. A super-resolved localisation method is proposed to recover the signal from it noisy observation and identify the unknown parameters. The reset of this Chapter is organized as follows. Section 2.1 introduces the basic concepts of the super-resolution technique. Section 2.2 formulates an optimisation problem for localisation problem, followed by two solvers. In Section 2.3, we present some numerical simulation results to illustrate the preciseness of the proposed method. The conclusion is drawn in Section 2.4.

2.1 Super-Resolution Technique

Our proposed super-resolved localisation is based on super-resolution technique due to its favourable properties. Super-resolution technique is designed to recover a superposition of point sources with high precision [6,50,51]. In settings of various super-resolution problems, one aims to recover the frequencies components from a superposition of complex

exponential functions. In fact, many problems arising in science and engineering with high dimensional signals that can be formulated as the same model. Examples of applications include localisation [15,19,52] , medical imaging [53], astronomy [54] and microscopy [55]. In localisation problem, the delay estimation can be transformed into frequency estimation problem by signal processing methods (e.g., Fourier transform and Finite Rate of Innovation (FRI) theory [56]) which encourage us applying the super-resolution technique to resolve localisation problem. In this section, we show a basic concept of super-resolution technique. The super-resolution technique is motivated by the idea of atomic norm for continuous parameter space. On the other hand, the atomic norm minimisation (ANM) is a well defined convex optimisation to recover the signal with superposition of atoms. The exact recovery is guaranteed when parameters are reasonably well separated.

2.1.1 Atomic Norm

To begin, we formulate the observed signal $\boldsymbol{y}$ as the superposition of atoms $\{\boldsymbol{a}_k\}_{k=1}^K$ from an atomic set $\mathcal{A}$. This form is widely used in signal processing, more precisely,

$$\boldsymbol{y} = \sum_{k=1}^{K} \gamma_k \boldsymbol{a}_k \ \in \mathbb{C}^N, \tag{2.1}$$

where γ_k is complex coefficient of k-th atom $\boldsymbol{a}_k$; the atom $\boldsymbol{a}_k \in \mathcal{A}$, is defined as simple building block of the signal; $\mathcal{A} = \{\cdots, \boldsymbol{a}_k, \cdots\}$ is called atomic set that includes the finite or infinite atoms.

In many scenarios, the dimension of signal N can be much less than the size of atomic set $\mathcal{A}$, leading to various combinations of atoms for representing signal. Then, how to select proper representation among these combinations is the key issue in signal processing. In the literature, one meaningful criterion is to seek parsimonious decomposition of signal $\boldsymbol{y}_j$ with the minimal number of atoms which is referred to as sparse representation.

The sparse representation is one meaningful criterion since many synthetic and natural

signals inherently satisfy this assumption [57,58], such as signal in localisation [15,19,52], audio and music signals [59], medical imaging [53], and astronomy [54]. The atomic norm proposed in [6, 51] aims to decompose the signal $\boldsymbol{y} \in \mathbb{C}^N$ into a linear combination of a few atoms. The definition is shown as

$$\|\boldsymbol{y}\|_{\mathcal{A}} = \inf \left\{ t > 0 \,:\, \boldsymbol{y} \in t \operatorname{conv}(\mathcal{A}) \right\}. \tag{2.2}$$

Given the definition in (2.2), atomic norm identifies the unit ball with the convex hull of atomic set $\mathcal{A}$. In view of sparsity recovery, the atomic norm enforces the signals involving only a few atoms. Alternatively, the atomic norm $\|\cdot\|_{\mathcal{A}}$ can be rewritten as [6]:

$$\|\boldsymbol{y}\|_{\mathcal{A}} = \inf \left\{ \sum_{\boldsymbol{a}_k \in \mathcal{A}} \gamma_k \,:\, \boldsymbol{y} = \sum_{k=1}^{K} \gamma_k \boldsymbol{a}_k \,,\ \gamma_k \geqslant 0, \quad \forall \boldsymbol{a}_k \in \mathcal{A} \right\}. \tag{2.3}$$

The dual atomic norm is defined as:

$$\|\boldsymbol{q}\|_{\mathcal{A}}^{*} = \sup \left\{ \langle \boldsymbol{q}, \boldsymbol{a} \rangle \,:\, \boldsymbol{a} \in \mathcal{A} \right\}, \tag{2.4}$$

where $\langle \boldsymbol{q}, \boldsymbol{a} \rangle = (\boldsymbol{a}^{\mathrm{H}} \boldsymbol{q})$.

The atomic norm $\|\cdot\|_{\mathcal{A}}$ is firstly proposed and analysed in [6] to enforce the sparsity in atomic set $\mathcal{A}$. As a natural regularizer, it can be alternatively viewed as the continuous counterpart of the L_1 norm in discrete CS methods [60–62] , and the nuclear norm for rank minimisation [6,63,64]. In the works of CS, the sparse atom is viewed as the foundation for signal compressing and recovering [65–68]. Discrete CS methods suppose that the signal is sparse under known transform atom such as Fourier transform matrix and wavelet basis, these atoms can be taken from finite dictionary by dividing the continuous parameter space into discrete grid. This discrete methods provide a well estimation on the condition that unknown parameters are perfectly located on the discrete grid. However, there are some unavoidable drawbacks in practical scenarios. As noticed in [69,70], the signal cannot be sparsely represented under a finite dictionary when the true parameters are not falling

onto the discrete grid. The finer grid can help for improving performance by minimising the difference between true parameter and imaginary discrete grids. However, a denser grids may degrade the performance of recovery. Since the dictionary will become very coherent and computational cost is increased by increasing the number of grid points [71]. To address mentioned problems, the super-resolution technique is proposed motivated by the idea of atomic norm for the continuous space.

2.1.2 ANM Method

The ANM is a well defined convex programming to recover the signal with superposition of atoms. Without considering measurement noise, the goal is to reconstruct $\boldsymbol{y} \in \mathbb{C}^N$ from the given observed signal $\boldsymbol{r} = \boldsymbol{y} \in \mathbb{C}^N$. As suggested in [6], the general form of atomic minimisation is written as

$$\begin{aligned} \min_{\boldsymbol{y}} \quad & \|\boldsymbol{y}\|_{\mathcal{A}} \\ \text{s.t.} \quad & \boldsymbol{r} = \boldsymbol{y}. \end{aligned} \tag{2.5}$$

The Lagrangian duality theory provides an important information for using atomic norm. The corresponding Lagrange dual problem of (2.5) is given as [72]

$$\max_{\boldsymbol{q}} \quad \operatorname{Re}\langle \boldsymbol{q}, \boldsymbol{r} \rangle \tag{2.6}$$

$$\text{s.t.} \quad \|\boldsymbol{q}\|_{\mathcal{A}}^{*} \leqslant 1, \tag{2.7}$$

where $\operatorname{Re}\langle \boldsymbol{q}, \boldsymbol{r} \rangle = \operatorname{Re}(\boldsymbol{r}^{\mathrm{H}} \boldsymbol{q})$ is the real part of a inner product, $\|\boldsymbol{q}\|_{\mathcal{A}}^{*}$ is dual atomic norm of vector $\boldsymbol{q} \in \mathbb{C}^N$. The dual atomic norm is defined as

$$\|\boldsymbol{q}\|_{\mathcal{A}}^{*} \triangleq \sup_{\boldsymbol{a} \in \mathcal{A}} \operatorname{Re}\ \langle \boldsymbol{a}, \boldsymbol{q} \rangle. \tag{2.8}$$

In case of $\boldsymbol{a}(\tau) = \left[e^{-i2\pi\tau 0}, e^{-i2\pi\tau}, \cdots, e^{-i2\pi\tau(N-1)}\right]^{\mathrm{T}}$, the dual atomic norm can be expressed as

$$\|\boldsymbol{q}\|_{\mathcal{A}}^{*} = \sup_{\tau\in[0,1)} |\underbrace{\langle \boldsymbol{a}(\tau), \boldsymbol{q}\rangle}_{Q(\tau)}|, \tag{2.9}$$

which indicates that $\|\boldsymbol{q}\|_{\mathcal{A}}^{*}$ can be written in a complex trigonometric polynomial $Q(\tau) = \langle \boldsymbol{a}(\tau), \boldsymbol{q}\rangle = \sum_{n=0}^{N-1} q_n e^{i2\pi\tau n}$. This gives an important information that the constraint (2.7) over infinite dimension can be transformed as linear matrix inequalities with some Hermitian matrices. Therefore, in case of $\boldsymbol{a}(\tau) = \left[e^{-i2\pi\tau 0}, e^{-i2\pi\tau}, \cdots, e^{-i2\pi\tau(N-1)}\right]^{\mathrm{T}}$, the dual problem of (2.6) can be rewritten as

$$\max_{\boldsymbol{q}} \quad \mathrm{Re}\,\langle \boldsymbol{q}, \boldsymbol{r}\rangle \tag{2.10}$$

$$\text{s.t} \quad \begin{bmatrix} \boldsymbol{H} & \boldsymbol{q} \\ \boldsymbol{q}^{*} & 1 \end{bmatrix} \succeq 0, \tag{2.11}$$

$$\sum_{p=1}^{N-l} H[p, p+l] = \begin{cases} 1, & l = 0 \\ 0, & l = 1, ..., N-1. \end{cases}, \tag{2.12}$$

where $H[p, p+l]$ denotes $[p, p+l]$-th entry of matrix $\boldsymbol{H}$. The problem (2.10) is semi-definite programming (SDP) involving finite constrains which can be solved using generic off-the-shelf convex solvers [73].

In the noise free case, the analysis of exact recovery guarantees is given in [51],

Theorem 2.1. *[51, 72] Suppose we observe the time samples of*

$$y[n] = \sum_{k=1}^{K} \gamma_k e^{i2\pi n\tau_k}\ ,\ n = 0, \ldots, N-1,$$

with unknown ToFs $\{\tau_1, \tau_2, ..., \tau_K\} \subset [0, 1]$ *on the index set* $T \subset \{0, ..., N-1\}$ *of size* L *selected uniformly at random. Additionally, assume* $\mathrm{sgn}(\gamma_k) := \gamma_k / |\gamma_k|$ *are drawn i.i.d from a symmetric distribution on the complex unit circle.*

$$\triangle\tau = \min_{a\neq b} |\tau_a - \tau_b|$$

If $\triangle\tau \geqslant \frac{4}{N-1}$*, then there exists a numerical constant* C *such that:*

$$L \geqslant C \max\left\{\log^2\frac{N}{\delta}, s\log\frac{K}{\delta}\log\frac{N}{\delta}\right\},$$

is sufficient to guarantee that we can recover $\boldsymbol{r}$ *and localize the ToFs* $\{\tau_k\}_{k=1}^K$ *via a semi-definite program with probability at least* $1-\delta$*.*

Theorem 2.1 shows that the unobserved samples and unknown parameters can be recovered when unknown parameters are reasonably well separated. Note that the continuous dictionary of super-resolution is globally coherent. This theorem shows that the performance of super-resolution technique is not determined by global coherence of the dictionary, but the local coherence between the atoms composing the true signal.

2.2 Super-Resolved Localisation

In the localisation problem, we first focus on a simple but widely used signal model. Without loss of generality, we assume the observed signal $\boldsymbol{r}_j$ at the j-th AP can be expressed in the following form,

$$\boldsymbol{r}_j = \boldsymbol{y}_j \tag{2.13}$$

where

$$\boldsymbol{y}_j = \sum_{k=1}^{K}\gamma_{j,k}\boldsymbol{a}(\tau_{j,k})\,, \quad \gamma_{j,k} > 0,\ \boldsymbol{a}(\tau_{j,k}) \in \mathcal{A}, \tag{2.14}$$

where $\gamma_{j,k}$ is weight of the atom to model attenuation of the k-th path at the j-th AP, atom $\boldsymbol{a}(\tau_{j,k}) \in \mathbb{C}^N$ is defined as $\left[e^{-i2\pi\tau_{j,k}0}, e^{-i2\pi\tau_{j,k}}, \cdots, e^{-i2\pi\tau_{j,k}(N-1)}\right]^{\mathrm{T}}$; $\tau_{j,k} \in [0,1]$ denotes propagation delay of the k-th path at the j-th AP, which is defined over continuous

space. The $\mathcal{A}$ is atomic set that includes the simple building blocks of the signal $\boldsymbol{y}_j$ which includes infinite number of atoms

$$\mathcal{A} = \left\{ \left[e^{-i2\pi\tau 0}, e^{-i2\pi\tau}, \cdots, e^{-i2\pi\tau(N-1)}\right]^{\mathrm{T}}, \tau \in [0,1] \right\}. \tag{2.15}$$

This is because the parameter τ can be arbitrary values across the continuous parameter space $[0,1]$.

The goal of localisation problem is estimating location of MD $\boldsymbol{m}$ based on observed signal $\{\boldsymbol{r}_j\}_{j=1}^{J}$ from J APs. Considering the ANM of (2.13), (2.14) can be reformulated as SDP, and solved by off-the-shelf solvers [51, 74, 75]. Therefore, in the first step, we reconstruct signal from observed signal $\{\boldsymbol{r}_j\}_{j=1}^{J}$ by primal solver in Section 2.2.2 or dual solver in Section 2.2.3. Then, the unknown parameters $\{\tau_{j,k}\}_{j=1,k=1}^{J,K}$ are retrieved for inferring location of MD. The parameter retrieving can be achieved by matrix pencil in Section 2.2.2.1 or dual certification in Section 2.2.3.1.

2.2.1 Why Atomic Norm

While there are numerous traditional methods in the literature, the atomic norm carries a number of favourable properties that are useful for helping recovering signal [6]. The motivations of using atomic norm for localisation are shown as follows:

- General framework. Atomic norm provides a general convex penalty function for linear inverse problems. Based on the definition of atomic norm in (2.2), the unit ball of atomic norm $\|\cdot\|_{\mathcal{A}}$ is the convex hull of the atomic set $\mathcal{A}$. Some examples are shown in Figure 2.1. When $\mathcal{A}$ is the collection of unit-Euclidean-norm one-sparse vectors, the convex hull is unit ball of the L_1 norm. When $\mathcal{A}$ is the set of unit-Euclidean-norm rank-one matrices, the convex hull is nuclear norm ball. In summary, atomic norm generalizes the L_1 norm for sparse vector estimation problems [60–62], and nuclear norm used for rank minimisation problems [6, 63, 64].

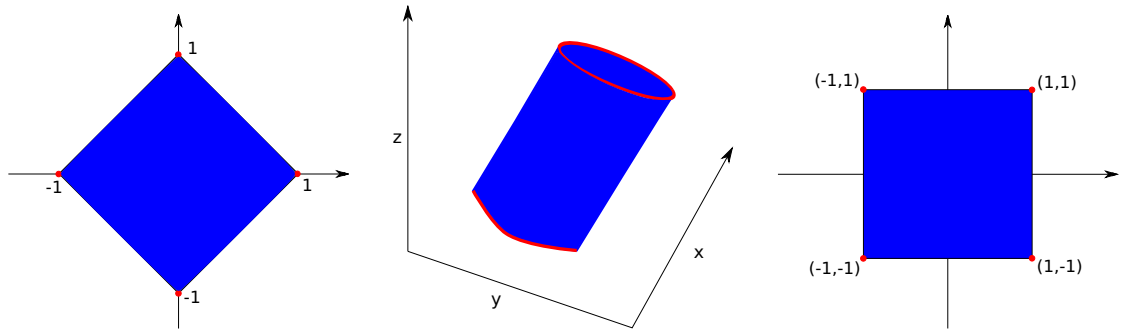


Figure 2.1: Unit balls of atomic norms. In the figures, red points indicate the set of atoms, and the unit ball of the associated atomic norm is shown in blue. In (a), the atoms are the unit-Euclidean-norm one-sparse vectors, and the atomic norm is the L_1 norm. In (b), the atoms are the 2×2 symmetric unit-Euclidean-norm rank-one matrices, and the atomic norm is the nuclear norm. In (c), the atoms are the vectors $\{-1,+1\}^2$, and the atomic norm is the L_∞ norm [6].

- Exact recovery guarantee: As introduced in [51, 72], the unknown parameters can be recovered with performance guarantee when parameters are reasonably well-separated $\triangle\tau \geqslant \frac{4}{N-1}$. See Theorem 2.1 for more information.

- Robust to noisy measurement: In practical scenarios, the observed signal are usually corrupted by noise [74,75]. There is no estimator that can exactly recover the signal from its noisy measurements. The atomic norm provides near-optimal recovery of $\boldsymbol{y}$, such as $\boldsymbol{r} = \boldsymbol{y} + \boldsymbol{v}$, where $\boldsymbol{v}$ is additive complex Gaussian noise with zero mean and variance σ^2 [76] . When the unknown parameters $\{\tau_k\}_{k=1}^K$ satisfy the separation condition $\triangle\tau \geqslant \frac{4}{N-1}$, the mean square error (MSE) of estimated $\hat{\boldsymbol{y}}$ given by lasso formulation $\|\boldsymbol{r} - \boldsymbol{y}\|_2^2 + \lambda\|\boldsymbol{r}\|_{\mathcal{A}}$ is bounded by [76]

$$\frac{1}{N}\|\widehat{\boldsymbol{y}} - \boldsymbol{y}_{\text{true}}\|_2^2 = O\left(\sigma^2 \frac{K \log N}{N}\right). \tag{2.16}$$

- A framework tailored to various applications. The signal recovered by atomic minimisation can be sum of sparse vectors, low-rank matrices, low-rank tensors and orthogonal matrices, etc. It provides an efficient solver for real-world application not only for localisation [15, 19, 52], but also audio and music signals [59], medical imaging [53], astronomy [54] and microscopy [55].

2.2.2 Primal Solver

In the first step, we consider to reconstruct $\boldsymbol{y}_j$ from its observed signal $\boldsymbol{r}_j = \boldsymbol{y}_j$. To achieve this, ANM based localisation can be formulated in following form:

$$\min_{\{\boldsymbol{y}_j\}_{j=1}^{J}} \quad \sum_j \|\boldsymbol{y}_j\|_{\mathcal{A}} \tag{2.17}$$

$$\text{s.t.} \qquad \boldsymbol{r}_j = \boldsymbol{y}_j, \ \text{ for } j = 1, ..., J. \tag{2.18}$$

Solving primal problem (2.17) require a high computational cost. Because the corresponding atomic set contains infinite number of atoms. It is worth noticing that the paper [51, 74, 75] has demonstrated that the ANM (2.17) is equivalently to semi-definite program

$$\min_{\{\boldsymbol{y}_j, t_j\}_{j=1}^{J}} \quad \sum_j \frac{1}{2N} \text{trace}\left(\text{Toep}\left(\boldsymbol{y}_j\right)\right) + \frac{1}{2} t_j \tag{2.19}$$

$$\text{s.t.} \qquad \begin{bmatrix} \text{Toep}\left(\boldsymbol{y}_j\right) & \boldsymbol{r}_j \\ \boldsymbol{r}_j^H & t_j \end{bmatrix} \succcurlyeq 0, \ \text{ for } j = 1, ..., J. \tag{2.20}$$

The semi-definite program is the well studied topic, and $\{\boldsymbol{y}_j\}_{j=1}^{J}$ can be obtained by CVX tool box [73]. Then the next step is retrieving ToFs $\{\tau_{j,k}\}_{j=1,k=1}^{J,K}$ from Toeplitz matrices $\{\text{Toep}\left(\boldsymbol{y}_j\right)\}_{j=1}^{J}$ by performing the matrix pencil method [77].

2.2.2.1 Exponential Retrieving Using Matrix Pencil

Solving SDP in (2.19), the signals $\{\boldsymbol{y}_j\}_{j=1}^{J}$ have been reconstructed from the observed $\{\boldsymbol{r}_j\}_{j=1}^{J}$. In order to localise MD, the next step is retrieving unknown ToFs $\{\tau_{j,k}\}_{j=1,k=1}^{J,K}$ from reconstructed signal $\{\boldsymbol{y}_j\}_{j=1}^{J}$. In this section, we present an example of using the matrix pencil method.

Firstly, matrix pencil with order $P \in \mathbb{Z}^+$ is defined as:

$$\boldsymbol{L}_P(\lambda) = \boldsymbol{A}_0 + \boldsymbol{A}_1\lambda + \cdots + \boldsymbol{A}_P\lambda^P, \tag{2.21}$$

where $\{\boldsymbol{A}_i\}_{i=0}^P$ are square matrices with the same size.

In particular, a linear matrix pencil has the form of $\boldsymbol{L}_1(\lambda) = \boldsymbol{A}_0 - \lambda\boldsymbol{A}_1$. An eigenvalue λ of a linear matrix pencil is a rank reducing number of the matrix pencil, i.e. $\det(\boldsymbol{A}_0 - \lambda\boldsymbol{A}_1) = 0$. The problem of retrieving all eigenvalues of a matrix pencil is called generalized eigenvalue problem.

Let us consider a sequence of signals $\{y[n]\}_{n=0}^{N-1}$. Specifically, we assume the element $y[n]$ is given by a sum of K exponentials:

$$y[n] = \sum_{k=1}^{K} \gamma_k a_k^n, \tag{2.22}$$

where all γ_k are non-zeros and $a_k^n = e^{-i2\pi\tau_k n}$.

A Toeplitz matrix has constant values along its diagonals. Let us denote $\boldsymbol{T}_{\{n,L,M\}}$ as a $L \times M$ Toeplitz matrix which is built using $\{y[n]\}_{n=0}^N$ and starts with element $y[n]$:

$$\boldsymbol{T}_{\{n,L,M\}} = \begin{bmatrix} y[n] & y[n-1] & \cdots & y[n-M+1] \\ y[n+1] & y[n] & \cdots & y[n-M+2] \\ \vdots & \vdots & \ddots & \vdots \\ y[n+L-1] & y[n+L-2] & \cdots & y[n-M+L] \end{bmatrix}. \tag{2.23}$$

The Toeplitz matrix $\boldsymbol{T}_{\{n,L,M\}}$ can be constructed using $L + M - 1$ consecutive elements. The number of rows L and columns M are assumed to be larger than the number of exponentials K. Based on the constraint (2.20), $\boldsymbol{T}_{\{n,L,M\}}$ is a positive semi-definite Toeplitz matrix. Thanks to classical Vandermonde decomposition [78], the positive Toeplitz matrix

$\boldsymbol{T}_{\{n,L,M\}}$ can be decomposed as follows:

$$\begin{aligned}\boldsymbol{T}_{\{n,L,M\}} &= \begin{bmatrix} 1 & \cdots & 1 \\ a_1 & \cdots & a_K \\ \vdots & \ddots & \vdots \\ a_1^{L-1} & \cdots & a_K^{L-1} \end{bmatrix} \begin{bmatrix} \gamma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \gamma_K \end{bmatrix} \begin{bmatrix} a_1^n & a_1^{n-1} & \cdots & a_1^{n-M+1} \\ \vdots & \vdots & \ddots & \vdots \\ a_K^n & a_K^{n-1} & \cdots & a_K^{n-M+1} \end{bmatrix} \\ &= \boldsymbol{V}_{\{0,L,K\}} \mathrm{diag}\left(\boldsymbol{\gamma}\right) \boldsymbol{V}_{\{n,M,K\}}^{\mathrm{T}}, \end{aligned} \tag{2.24}$$

where $\boldsymbol{V}_{\{a,b,c\}}$ is a Vandermonde matrix starting with order a and with b rows and c columns, and $\mathrm{diag}\left(\boldsymbol{\gamma}\right)$ is a diagonal matrix with diagonal values being $[\gamma_1, \cdots, \gamma_K]$.

Let us build Toeplitz matrices $\boldsymbol{T}_{\{0,K,K\}}$ and $\boldsymbol{T}_{\{1,K,K\}}$ using $\{y\left[n\right]\}_{n=0}^{N}$ where the elements are sum of K exponentials. From (2.24), we have that:

$$\boldsymbol{T}_{\{0,K,K\}} = \boldsymbol{V}_{\{0,K,K\}} \mathrm{diag}\left(\boldsymbol{\gamma}\right) \boldsymbol{V}_{\{0,K,K\}}^{\mathrm{T}} \tag{2.25}$$

and

$$\begin{aligned}\boldsymbol{T}_{\{1,K,K\}} &= \boldsymbol{V}_{\{0,K,K\}} \mathrm{diag}\left(\boldsymbol{\gamma}\right) \boldsymbol{V}_{\{1,K,K\}}^{\mathrm{T}} \\ &= \boldsymbol{V}_{\{0,K,K\}} \mathrm{diag}\left(\boldsymbol{\gamma}\right) \mathrm{diag}\left(\boldsymbol{a}\right) \boldsymbol{V}_{\{0,K,K\},}^{\mathrm{T}} \end{aligned} \tag{2.26}$$

where $\mathrm{diag}(\boldsymbol{a}) = \mathrm{diag}([a_1, a_2, \cdots, a_k])$. In case of the $\boldsymbol{V}_{\{0,K,K\}}$ is invertible, based on (2.25) and (2.26), we have:

$$\boldsymbol{T}_{\{1,K,K\}} \boldsymbol{V}_{\{0,K,K\}}^{\text{-T}} = \boldsymbol{T}_{\{0,K,K\}} \boldsymbol{V}_{\{0,K,K\}}^{\text{-T}} \mathrm{diag}\left(\boldsymbol{a}\right). \tag{2.27}$$

From (2.27), $\boldsymbol{V}_{\{0,K,K\}}^{\text{-T}}$ can be considered as the generalised eigenvectors for matrix pair $\boldsymbol{T}_{\{0,K,K\}}$ and $\boldsymbol{T}_{\{1,K,K\}}$ with the corresponding generalised eigenvalues $\mathrm{diag}\left(\boldsymbol{a}\right)$. The roots of the exponentials can be obtained by solving a generalized eigenvalue problem for a

linear matrix pencil $\boldsymbol{T}_{\{1,K,K\}} - \mu \boldsymbol{T}_{\{0,K,K\}}$:

$$\left(\boldsymbol{T}_{\{1,K,K\}} - \mu \boldsymbol{T}_{\{0,K,K\}}\right) \boldsymbol{v} = 0, \tag{2.28}$$

where the eigenvalues $\mu \in \{a_1, ..., a_k, ..., a_K\}$ are exponentials composing the true signal (2.22); $\boldsymbol{v} \in \{\boldsymbol{v}_1, ..., \boldsymbol{v}_K\}$ is column vector composing the matrix $\boldsymbol{V}^{\text{-T}}_{\{0,K,K\}} = [\boldsymbol{v}_1, ..., \boldsymbol{v}_K]$.

The generalized eigenvalue of the matrix pencil is equivalent to the eigenvalue of matrix $\boldsymbol{T}^{-1}_{\{0,K,K\}} \boldsymbol{T}_{\{1,K,K\}}$:

$$\left(\boldsymbol{T}^{-1}_{\{0,K,K\}} \boldsymbol{T}_{\{1,K,K\}} - \mu \boldsymbol{I}\right) \boldsymbol{v} = 0. \tag{2.29}$$

When the observed data is corrupted by noise, directly taking the inversion of $\boldsymbol{T}^{-1}_{\{0,K,K\}}$ may not be a good numerical approach. To be more noise resilient, we can generate two bigger Toeplitz matrices as $\boldsymbol{T}_{\{0,N-L,L\}}$ and $\boldsymbol{T}_{\{1,N-L,L\}}$ with $K \leq L \leq N - K$. They are obtained by dropping the first and last row of $\boldsymbol{T}_{\{0,N-L+1,L\}}$, respectively. Let us consider Toeplitz matrix $\boldsymbol{T}_{\{0,N-L+1,L\}}$ and take the truncated singular value decomposition of it and only retain the K largest singular values:

$$\boldsymbol{T}_{\{0,N-L+1,L\}} = \boldsymbol{U}_K \boldsymbol{\Sigma}_K \boldsymbol{V}_K^T, \tag{2.30}$$

where $\boldsymbol{\Sigma}_K$ is the diagonal matrix containing the K largest singular values of $\boldsymbol{T}_{\{0,N-L+1,L\}}$, and $\boldsymbol{U}_K$ and $\boldsymbol{V}_K$ are the corresponding left and right singular vectors.

$\boldsymbol{T}_{\{0,N-L,L\}}$ and $\boldsymbol{T}_{\{1,N-L,L\}}$ can be represented as:

$$\boldsymbol{T}_{\{1,N-L,L\}} = \overline{\boldsymbol{U}}_K \boldsymbol{\Sigma}_K \boldsymbol{V}_K^T, \tag{2.31}$$

$$\boldsymbol{T}_{\{0,N-L,L\}} = \underline{\boldsymbol{U}}_K \boldsymbol{\Sigma}_K \boldsymbol{V}_K^T, \tag{2.32}$$

where $\overline{\boldsymbol{U}}_K$ and $\underline{\boldsymbol{U}}_K$ are obtained by dropping the first and last row of $\boldsymbol{U}_K$.

Algorithm 1: Matrix Pencil Method

Input: Input noisy samples $\{y[n]\}_{n=0}^{N-1}$ and number of exponentials K.
Output: Retrieved exponentials $\{a_k\}_{k=1}^{K}$.
1: Build the Toeplitz matrix $\boldsymbol{T}$ from noisy samples.
2: Perform SVD of $\boldsymbol{T}$: $\boldsymbol{T} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$.
3: Keep the K left-singular vectors that correspond to the K largest singular values $\boldsymbol{U}_K$.
4: Build $\overline{\boldsymbol{U}}_K$ and $\underline{\boldsymbol{U}}_K$ from $\boldsymbol{U}_K$.
5: Retrieve K exponentials $\{a_k\}_{k=1}^{K}$ by solving the eigenvalue problem in (2.34).

Based on (2.28), (2.31) and (2.32), we have:

$$\left(\overline{\boldsymbol{U}}_K - \mu\underline{\boldsymbol{U}}_K\right)\boldsymbol{\Sigma}_K\boldsymbol{V}_K^T\boldsymbol{v} = 0. \tag{2.33}$$

The exponentials can be obtained by solving the eigenvalue problem:

$$\left(\underline{\boldsymbol{U}}_K^{\dagger}\overline{\boldsymbol{U}}_K - \mu\boldsymbol{I}\right)\boldsymbol{v} = 0. \tag{2.34}$$

where $\underline{\boldsymbol{U}}_K^{\dagger}$ is the Moore-Penrose pseudo-inverse of $\underline{\boldsymbol{U}}_K$; the eigenvalues $\mu \in \{a_1, ..., a_k, ..., a_K\}$ are exponentials composing the true signal (2.22). **Algorithm 1** summarizes the matrix pencil method for retrieving exponentials.

2.2.2.2 Stability Robustness of Matrix Pencil

The stability robustness analysis of matrix pencil has been studied in [79], and this problem is to estimate the distance of a stable matrix to the set of all unstable matrices. In the literature, a matrix is defined to be stable when all the eigenvalues are contained in the open left half of the complex plane [79–81]. In [79], the stability robustness analysis of matrix is extended to matrix pencil. A linear matrix pencil has the form of

$$\boldsymbol{L}_1(\lambda) = \boldsymbol{A}_0 - \lambda\boldsymbol{A}_1. \tag{2.35}$$

The matrix pencil (2.35) is said to be stable if all its generalized eigenvalues are located in the open left half of the complex plane [79]. The generalised eigenvalues of (2.35) are defined as the roots of the polynomial in λ:

$$\det\left(\boldsymbol{A}_0 - \lambda \boldsymbol{A}_1\right). \tag{2.36}$$

Theorem 2.2. *Given $\boldsymbol{A}_0$ and $\boldsymbol{A}_1 \in \mathbb{C}^{n\times n}$ such that $\rho(\boldsymbol{A}_1) = r^{rank}$, the matrix pencil $\boldsymbol{L}_1(\lambda)$ is stable and $|\Lambda(\boldsymbol{L}_1(\lambda))| = r^{rank}$, then*

- $\mu(\boldsymbol{L}_1(\lambda)) > 0$;
- $\alpha\mu(\boldsymbol{L}_1(\lambda)) = \mu(\alpha\boldsymbol{A}_0 - \lambda\beta\boldsymbol{A_1})$, *for any* $\alpha > 0$, $\beta > 0$;
- $\mu(\boldsymbol{L}_1(\lambda)) = \mu(\boldsymbol{W}_1\boldsymbol{A}_0\boldsymbol{W}_2 - \boldsymbol{W}_1\boldsymbol{A_1}\boldsymbol{W}_2)$, *for any orthogonal matrices* $\boldsymbol{W_1}$, $\boldsymbol{W}_2 \in \mathbb{R}^{n\times n}$;
- $\mu(\boldsymbol{L}_1(\lambda)) \leq \underline{\sigma}(\boldsymbol{A}_0)$;
- $\mu(\boldsymbol{L}_1(\lambda)) \leq \underline{\sigma}\left(\boldsymbol{U}_2^T\boldsymbol{A}_0\boldsymbol{V}_2\right)$;

where $\rho(\boldsymbol{A}_1)$ is rank of matrix $\boldsymbol{A}_1$; $\Lambda(\boldsymbol{L}_1(\lambda)$ is the set of all generalised eigenvalues; $|\Lambda(\boldsymbol{L}_1(\lambda))|$ is the number of elements in $\Lambda(\boldsymbol{L}_1(\lambda))$; $\mu(\boldsymbol{L}_1(\lambda))$ is the distance of $\boldsymbol{L}_1(\lambda)$ from instability by

$$\mu(\boldsymbol{L}_1(\lambda)) = \inf\{\| \triangle \|_s : \triangle \in \mathbb{R}^{n\times n} \text{ and } \Lambda(\boldsymbol{L}_1(\lambda) + \triangle) \not\subset \mathbb{C}^-\}, \tag{2.37}$$

where $\|\cdot\|_s$ is spectral norm; $\underline{\sigma}$ is the minimum singular value of the matrix $\boldsymbol{A}_0$; $\mathbb{C}^-$ denotes the sets $\{\lambda \in \mathbb{C} : Re(\lambda) < 0\}$; Let a singular value decomposition of $\boldsymbol{A}_1$ is

$$\boldsymbol{A}_1 = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T = \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{S}_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1 & \boldsymbol{V}_2 \end{bmatrix}^T, \tag{2.38}$$

where $\boldsymbol{U}, \boldsymbol{V} \in \mathbb{R}^{n\times n}$ are orthogonal matrices and $\boldsymbol{S}_{11} \in \mathbb{R}^{n\times n}$ is a diagonal matrix with positive elements [79].

Remark that the Theorem 2.1 also indicates the matrices $\boldsymbol{A_0}, \boldsymbol{A}_1$ can be recovered with performance guarantee when parameters are well-separated. Because the parameters recovered by ANM can be used for constructing matrices $\boldsymbol{A_0}, \boldsymbol{A}_1$. Therefore, in the practice, it is reasonably to apply **Algorithm 1** for the matrix pencil problem.

2.2.3 Dual Solver

As for atomic norm minimisation, the dual problem provides a useful structure for analysis and implementation. The first step is to reconstruct dual feasible $\{\boldsymbol{q}_j\}_{j=1}^{J}$ from observed signal $\{\boldsymbol{r}_{j=1}^{J}\}$. Then, retrieve the unknown time dealys $\{\tau_{j,k}\}_{j=1,k=1}^{J,K}$ from dual feasible $\{\boldsymbol{q}_j\}_{j=1}^{J}$.

In particular, it is shown in [6] that the dual problem of program (2.17) can be written as

$$\max_{\{\boldsymbol{q}_j\}_{j=1}^{J}} \quad \sum_j \mathrm{Re}\left\langle \boldsymbol{q}_j, \boldsymbol{r}_j \right\rangle \tag{2.39}$$

$$\text{s.t} \qquad \text{for } j = 1, ..., J,$$

$$\left|\left\langle \boldsymbol{q}_j, \boldsymbol{a}\left(\tau\right)\right\rangle\right| \leq 1,\ \tau \in [0,1]\,. \tag{2.40}$$

which is the optimisation program with the infinite many constrains (2.40). Solving the problem (2.39) is computational cost. As presented in Section 2.1.2, the dual problem (2.39) can be characterized as semi-definite program with finite many constrains (2.42),

(2.43). This dual problem of localisation also has a semi-definite formulation:

$$\max_{\{\boldsymbol{q}_j\}_{j=1}^{J}} \quad \sum_{j} \mathrm{Re}\left\langle \boldsymbol{q}_j, \boldsymbol{r}_j \right\rangle \tag{2.41}$$

$$\text{s.t} \quad \text{for } j = 1, ..., J,$$

$$\begin{bmatrix} \boldsymbol{H}_j & \boldsymbol{q}_j \\ \boldsymbol{q}_j^* & 1 \end{bmatrix} \succeq 0, \tag{2.42}$$

$$\sum_{p=1}^{N-l} H_j[p, p+l] = \begin{cases} 1, & l = 0 \\ 0, & l = 1, ..., N-1, \end{cases} \tag{2.43}$$

where $H_j[p, p+l]$ denotes $[p, p+l]$-th entry of matrix $\boldsymbol{H}_j$. As shown in [50, 51, 75], the linear matrix inequalities (2.42), (2.43) are equivalent to the dual norm constraint $|\langle \boldsymbol{q}_j, \boldsymbol{a}(\tau) \rangle| \leq 1$, $\tau \in [0, 1]$. The SDP in (2.41) can be solved off-the-shelf solvers, such as CVX [73]. Then the next step is extracting ToFs $\{\tau_{j,k}\}_{j=1,k=1}^{J,K}$ from dual problem solution $\{\boldsymbol{q}_j\}_{j=1}^{J}$ by performing the dual certification, as detailed next.

2.2.3.1 Exponential Retrieving Using Dual Certification

Given the exponential atom ($a[n] = e^{-i2\pi\tau n}$), the dual certification provides useful approach for retrieving unknown delays. In the dual certification, the true parameters $\{\boldsymbol{\tau}'_{j,k}\}_{k=1}^{K}$ that comprise the optimal $\boldsymbol{y}_j$ is located at peaks of the dual polynomial $Q_j(\tau) = \langle \boldsymbol{q}_j, \boldsymbol{a}(\tau) \rangle$:

$$\{\boldsymbol{\tau}'_{j,k}\}_{k=1}^{K} = \{\tau : |Q_j(\tau)| = 1, \, \tau \in [0, 1]\}, \tag{2.44}$$

This method is shown in Fig 2.2, for the observed signal $\boldsymbol{r} = \sum_{k=1}^{5} \gamma_k \boldsymbol{a}(\tau_k)$ with five ToFs, where peak of the dual polynomial $Q_j(\tau)$ is matching to the location of true ToFs.

Under strong duality, one implies that equality between primal objective value $\|\boldsymbol{y}_j\|_{\mathcal{A}} = \sum_{k=1}^{K} |\gamma_{j,k}|$, where $\boldsymbol{y}_j = \sum_{k=1}^{K} \gamma_{j,k} \boldsymbol{a}(\tau_{j,k})$ and dual objective value $\mathrm{Re}\left\langle \boldsymbol{q}_j, \boldsymbol{r}_j \right\rangle$. Given $\boldsymbol{r}_j =$

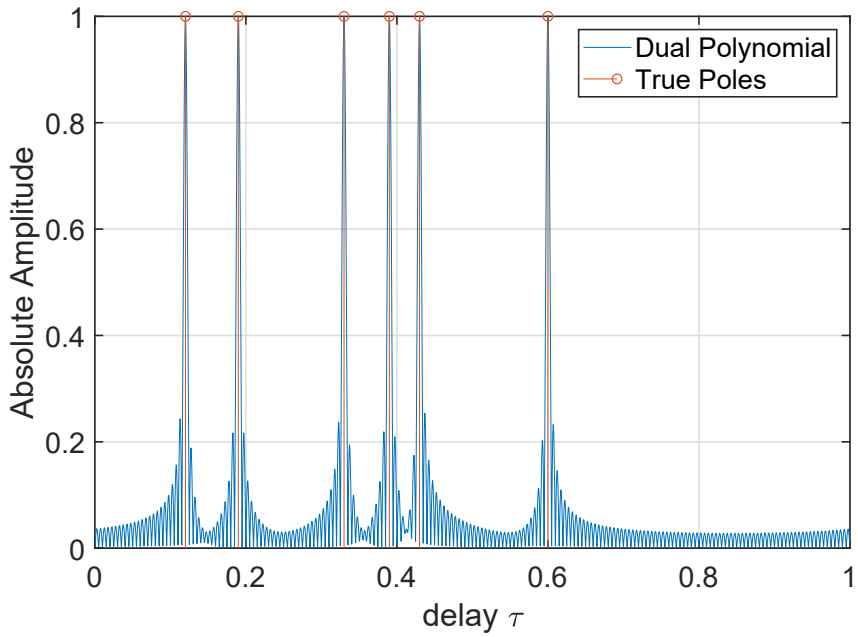


Figure 2.2: Retrieving ToF τ via identifying the peaks of the dual polynomial (in blue).

$\boldsymbol{y}_j$, the dual objective value can be expressed as

$$\mathrm{Re}\,\langle \boldsymbol{q}_j, \boldsymbol{y}_j \rangle = \mathrm{Re}\left\langle \boldsymbol{q}_j, \sum_{k=1}^{K} \gamma_{j,k} \boldsymbol{a}(\tau_{j,k}) \right\rangle = \mathrm{Re} \sum_{k=1}^{K} \gamma_{j,k}^{*} Q_j(\tau_{j,k}) = \sum_{k=1}^{K} |\gamma_{j,k}|. \tag{2.45}$$

Therefore, we have $Q_j(\tau_{j,k}) = \mathrm{sign}(\gamma_{j,k}) = \frac{\gamma_{j,k}}{|\gamma_{j,k}|}$ and $|Q_j(\tau_{j,k})| = 1$.

Note that the dual certification does not require any prior knowledge, and can estimate the unknown parameters with infinitesimal precision. One the other hand, it can be efficiently performed by Fast Fourier transform.

2.3 Numerical Simulation

In this section, we will first introduce the configuration of the system and the shape of the transmitted waveform. Then, we will present a super-resolution based localisation method and the simulation results.

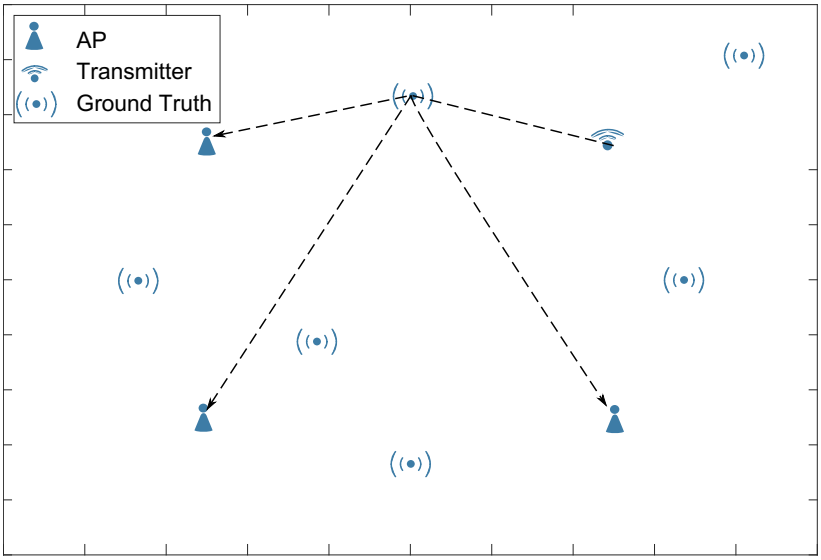


Figure 2.3: The system setup for simulation

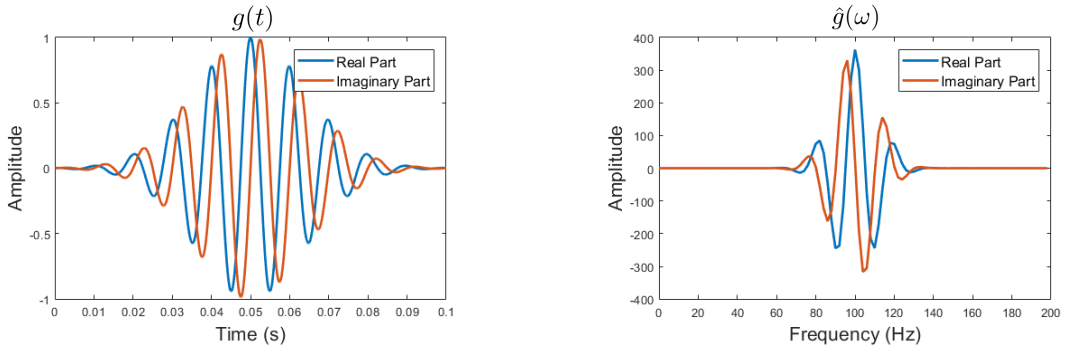
2.3.1 System Configuration

As shown in Fig 2.3, we consider the localisation system with one transmitter, $J = 3$ APs to localise $K = 6$ MDs. The transmitted signal is firstly sent from transmitter, then passing through the MD, finally arriving at AP. The signal speed is assumed to be constant in the propagation medium. We make the assumptions in the problem setting that the coordinates of transmitter $\boldsymbol{T}_{rans}$, APs $\{\boldsymbol{AP}_j\}_{j=1}^{J}$ are known and they work with synchronized clocks. The transmitted signal is also known to the APs. This localisation system has emerged in many novel applications such as target detection and localisation, ultrasound imaging and Internet of Things.

2.3.2 Transmitted Signal

We assume that transmitters transmit pulses with the same shape, and the transmitted signal $g(t)$ is assumed to be a Gaussian modulated sinusoidal pulse:

$$g(t) = \exp\left(-\left(t - \frac{T_0}{2}\right)^2 \sigma_g^{-2}\right) \exp\left(j2\pi f_c \left(t - \frac{T_0}{2}\right)\right), \tag{2.46}$$



(a) Transmitted waveform in time domain $g(t)$.

(b) Transmitted waveform in frequency domain $\hat{g}(\omega)$.

Figure 2.4: An example of transmitted waveform and its Fourier transform with $T_0 = 0.1\,\text{s}$, $\sigma_g = 0.02$, and $f_c = 100\,\text{Hz}$.

where T_0 is the duration of the transmitted waveform, σ_g is the standard deviation of the Gaussian pulse, and f_c is the frequency of the sinusoidal signal.

The transmitted waveform $g(t)$ from transmitter is passed by the k-th MD $\boldsymbol{m}_k$ and is then received by the j-th AP $\boldsymbol{AP}_j$. Therefore, the received signal in the j-th AP is superposition of multiple delayed versions of $g(t)$, and it takes the form of:

$$x_j(t) = \sum_{k=1}^{K} \gamma_{j,k} g(t - \tau_{j,k}) + v_j, \text{ for } j = 1, 2, ..., J, \tag{2.47}$$

where $\gamma_{j,k}$ is the unknown coefficients modelling the signal attenuation of k-th path at j-th AP; $\tau_{j,k}$ is the ToF for the signal transmitted by $\boldsymbol{T}_{rans}$, reflected by $\boldsymbol{m}_k$ and received by j-th AP $\boldsymbol{AP}_j$. The propagation delay $\tau_{j,k} = d_{j,k}/c$ is determined by the transmission distance $d_{j,k} = \|\boldsymbol{T}_{rans} - \boldsymbol{m}_k\|_2 + \|\boldsymbol{m}_k - \boldsymbol{AP}_j\|_2$ and the speed of signal propagation in the space c; v_j is additive Gaussian noise with zero mean and variance σ_j^2.

In practice, the received signal is uniform time-samples of $x_j(t)$ in the form of

$$x_j[n] = \sum_{k=1}^{K} \gamma_{j,k} g(nT_s - \tau_{j,k}) + v_j[n], \tag{2.48}$$

where $n = 0, 1, ..., N-1$ and N is the number of time instances for taking samples; T_s is the sampling period.

In order to apply ANM framework for estimating unknown ToFs, the first step is transforming received signal into superposition of exponentials as (2.13), (2.14). The Fourier transform of received signal $x_j(t)$ defined in (2.47) is given by:

$$\hat{x}_j(\omega) = \sum_{k=1}^{K} \gamma_{j,k} \hat{g}(\omega) e^{-j\omega\tau_{j,k}}, \tag{2.49}$$

where $\hat{x}_j(\omega)$ and $\hat{g}(\omega)$ is the Fourier transform of $x_j(t)$ and $g(t)$, respectively. The Fourier transform of the transmitted signal $g(t)$ is expressed as follows:

$$\hat{g}(\omega) = \sigma_g \sqrt{\pi} \exp\left(-\left(\sigma_g \pi (\omega - f_c)\right)^2\right) \exp\left(j2\pi \frac{T_0}{2} (\omega - f_c)\right), \tag{2.50}$$

where T_0 is the duration of the transmitted waveform, σ_g is the standard deviation of the Gaussian pulse in time domain, and f_c is the frequency of the sinusoidal signal. Figure 2.4 shows an example of the transmitted waveform in time $g(t)$ and frequency domain $\hat{g}(\omega)$, respectively. Note that the center of frequency domain signal $\hat{g}(\omega)$ is located at f_c and the standard deviation of the Gaussian term in $\hat{g}(\omega)$ is $(2\sigma_g\pi)^{-\frac{1}{2}}$. That is, if the transmitted pulse $g(t)$ is wider in time domain, its spectral domain representation $\hat{g}(\omega)$ has a narrower pulse. In order to retrieve sufficient samples of the sum of exponentials, it would be better to have a smaller standard deviation σ_g in the transmitted waveform.

From (2.50) and (2.49), we can therefore rewrite the observed signal as sum of K exponentials:

$$r_j[n] = \frac{\hat{x}_j(n\omega_0)}{\hat{g}(n\omega_0)} = \sum_{k=1}^{K} \gamma_{j,k} e^{-jn\omega_0\tau_{j,k}}, \tag{2.51}$$

where $n = 0, 1, ..., N-1$, $\hat{x}_j(n\omega_0)$ and $\hat{g}(n\omega_0)$ are Fourier transform of $x_j(t)$ and $g(t)$ evaluated at $\omega = n\omega_0$, respectively. In this way, the delay estimation has been transformed into a line spectral estimation problem which can be recovered using super-resolution techniques [51, 72].

Algorithm 2: Atomic Norm Based Localisation Method

Input: propagation speed s, parameters of the transmitted waveform θ, sampling rate f_s and threshold value ε.

Output: The locations of K targets.

1: **for** every transmitter/receiver
2: Extract the observed signal as in Eqn. (2.51).
3: Apply primal solver (2.19) to recover Toeplitz matrix Toep($\boldsymbol{y}_j$), for $j = 1, ..., J$.
4: Apply matrix pencil method (**Algorithm** 1) to retrieve JK exponentials from Toep($\boldsymbol{y}_j$), for $j = 1, ..., J$.
5: Obtain JK delays from the retrieved exponentials.
6: **end for**
7: Determine the locations of K targets based on the JK delays through trilateration.

2.3.3 Localisation Method

At each AP, a segment of observed signal can be obtained as in (2.51). It is the superposition of K exponentials and noise. To gain robustness against noise, the signal is reconstructed by atomic norm based minimisation (2.19). Matrix pencil method is then applied to the observed signal to retrieve K exponentials. With the obtained unique delays, the location of the targets can be calculated through trilateration. **Algorithm 2** summarizes the proposed atomic norm based localisation method.

2.3.4 Simulation Results

In this section, simulation results are presented. **Table** 2.1 summarizes the parameters setting for simulation. The transmitter, APs and the MDs are assumed to be in 2 dimensional space. The strength of the noise is measured using Signal-to-Noise Ratio (SNR) in dB, where $SNR = 10 \log_{10} \left(\frac{\text{Signal Power}}{\text{Noise Power}} \right)$. White Gaussian noise is added to the received time sequence samples $x_j[n]$ with $SNR = 0$ dB.

Figure 2.5 shows the estimation results by the super-resolved method and discrete CS method (orthogonal matching pursuit (OMP) [82]). The unknown targets are placed at intersection of two ellipses, which result in poor resolution in multi-stage estimation. The OMP approach suffers from performance degradation, since each sub-stage is unable

Number of transmitters/receivers	$N_T = 1/N_R = 3$
Number of MDs	$K = 6$
Number of samples	$N = 128$
Propagation speed	$c = 3 \times 10^5\,\mathrm{m/s}$
Transmitted waveform duration	$T_0 = 0.1\,\mathrm{s}$
Frequency of sinusoidal	$f_c = 1000\,\mathrm{Hz}$
Standard deviation of Gaussian	$\sigma_g = 0.001$
Sampling rate	$f_s = 10 \times 2^{10}\,\mathrm{Hz}$
Scale	$1 \times 10^4\,\mathrm{m}$

Table 2.1: Parameters setting for simulation.

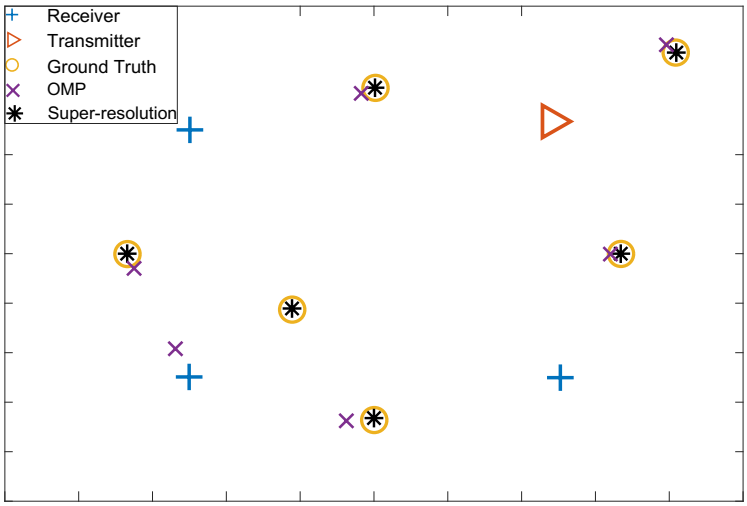


Figure 2.5: The simulation results for 2-dimensional localisation

to correctly detect or estimate the targets. Because the OMP approach assumes that the unknown MDs are located at predefined grids. When this assumption is satisfied, the OMP approach can provide a well estimation. However, the performance of OMP could be degraded when the true MDs are not falling on the discrete grid. A denser grid can help for improving performance by minimising the difference between true parameter and imaginary discrete grids. But, in practice, it is hard to determine the proper scale of grid points. Since the dictionary will become very coherent and computational cost is increased by increasing the number of grid points. In this simulation, the estimated results of the OMP show a low positioning accuracy and failed to estimate one target. Compared with the discrete approach, the super-resolved localisation is able to recover the unknown targets $\{\boldsymbol{m}_k\}_{k=1}^{K=6}$ with high precision. The super-resolved localisation scheme avoids drawbacks of discrete CS by working on continuous parameter space. Note that

this continuous parameter space will not result in a global coherence of the dictionary. As illustrated in Theorem 2.1, the performance of super-resolution technique is determined by the local coherence between the atoms composing the true signal.

Sample Number	Solver	Time [s]	Iteration
ANM(N= 64)	SDPT3 | SeDuMi	1.833 |1.623	14.12 | 23.45
ANM(N=128)	SDPT3 | SeDuMi	10.784 |8.234	16.63 | 26.21
ANM(N=256)	SDPT3 | SeDuMi	114.342 |90.778	19.43 |28.75
ANM(N=364)	SDPT3 | SeDuMi	376.166 |555.532	18.92 |39.55

Table 2.2: In case of SNR = 10 dB, the average running time of the ANM. The ANM is solved by Primal Solver in (2.19). In the Primal Solver, the SDP is solved by SDPT3 [83] and SeDuMi [84]. CPU: Intel Core i7-7700HQ, 2.80GHz.

The computational cost of ANM is measured by the average running time over 200 Monte Carlo trials. The corresponding SDP is solved by SDPT3 [83] and SeDuMi [84] respectively. Based on the Table 2.2, the SDPT3 performs less iterations compared with the SeDuMi. In case of $N = 64, N = 128, N = 256$, the SDPT3 spends more running time compared with the SeDuMi. However, in case of $N = 365$, the running time of SDPT3 is significantly decreased compared with the SeDuMi.

2.4 Conclusion

This Chapter has shown a super-resolved localisation scheme based on super-resolution technique. Firstly, we introduce background and motivation of super-resolution technique. It is designed to recover a superposition of point sources with high precision. Then we formulate the localisation problem as a convex optimisation problem and provide two solvers for the problem. Finally, the simulation result shows the superior performance of super-resolved localisation compared with discrete CS method. However, directly apply the super-resolution technique to localisation involving multi-stage process i.e., first estimating ToFs, then localising MD. The performance is degraded due to data association problem and error propagation. The task of associating intermedia parameters

(ToF, TDoF and DoA) to MDs or scatters is known as an NP-hard problem, leading to huge computational cost [85,86]. Moreover, the estimation of intermedia parameters may introduce additional noise and lead to error propagation.

In order to address these issues, we consider direct joint processing for localising MD. The estimation is directly obtained by processing all the measurement simultaneously. Then the error propagation and data association is avoided. The objective function is formulated as

$$\min_{\boldsymbol{m}} \sum_{j}^{J} \|\boldsymbol{r}_j - \boldsymbol{y}_j(\boldsymbol{m})\|_2^2 + \lambda \mathcal{G}([\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_J]) \,, \tag{2.52}$$

where $\lambda > 0$ is the regularization parameter that tunes a tradeoff between the fidelity to the observation and the size of the penalty function; $\mathcal{G}(\cdot)$ is the penalty function to promote sparse solution. The technical difficulties come from two aspects:

- How to design the sparse penalty function $\mathcal{G}(\cdot)$ to super-resolve the location of MD in various propagation environments by jointly processing all the received signals.
- The problem (2.52) can not be solved by SDP. To apply the SDP formulation for ANM, one must first define the atom properly. Precisely, it requires that atom should follow the simple exponential form as

$$\boldsymbol{a}(\tau) = \left[e^{-i2\pi\tau 0}, e^{-i2\pi\tau}, \cdots, e^{-i2\pi\tau(N-1)}\right]^{\mathrm{T}}. \tag{2.53}$$

  However, in the direct mapping formulation the atom is in a totally different form

$$\boldsymbol{a}(\boldsymbol{m}) = \left[e^{-i2\pi\tau(\boldsymbol{m})0}, e^{-i2\pi\tau(\boldsymbol{m})}, \cdots, e^{-i2\pi\tau(\boldsymbol{m})(N-1)}\right]^{\mathrm{T}}, \tag{2.54}$$

  where the atom $\boldsymbol{a}(\boldsymbol{m})$ is directly mapping from the location of MD $\boldsymbol{m} = [m^x, m^y]^{\mathrm{T}}$; $\tau(\boldsymbol{m})$ is the ToF from the MD to AP.

In the Chapter 3, 4, we propose two different penalty functions for different localisation problems. We give a novel method to solve the proposed optimisation problems. This

method iteratively adds candidate estimation to the support set at each iteration, then followed by pruning step to remove estimations when they are found not contributing to reducing loss function. Finally, the precision of estimations are further improved by descent method over continuous parameter space. Numerical simulation results verify its feasibility and the advantage over traditional methods.

Chapter 3

Direct Joint Localisation with Self-calibration

Under high angular resolution, the DoAs of LoS paths can provide precise estimation of MDs. In this Chapter, we consider directional error that degrade the performance of DoA based localisation scheme. Besides the NLoS bias, another directional error is called array directional error. The array directional error implies the deviation in the orientation of the antenna array placement which is rarely noticed. We present a novel super-resolved localisation method that estimates multiple MDs and performs self-calibration to correct array directional errors simultaneously. To achieve joint localisation, we directly map MD locations and array directional errors to received signals. Then the group sparsity based optimisation is proposed to only exploit the geometric consistency that received LoS paths at different APs are originating from common MDs. Thus, the proposed method also can be applied to separate LoS/NLoS paths. The rest of the Chapter is organized as follows. Section 3.1 introduces backgrounds of the problem. Section 3.2 describes the signal model. Section 3.3 explains the feasibility of self-calibration from the aspect of necessary conditions. Section 3.4 briefly reviews the basic concepts of group sparse recovery and ANM, followed by establishing an optimisation problem for direct passive joint emitter localization with selfcalibration. Section 3.5 presents the proposed novel

method to solve the formulated optimisation problem. Section 3.6 shows some numerical simulation results. The conclusion is drawn in Section 3.7.

3.1 Introduction

Due to the rapid development of MDs and wireless techniques, localisation with wireless signals is in strong demand of various applications. In general, these localisation methods can be divided into several categories by the types of signal measurements, such as DoA, ToF, TDoF, and RSS. This Chapter considers the general problem of localising multiple MDs based on measurements at APs of which the locations are given. The APs are equipped with antenna array and only DoA information can be obtained for localisation. DoA is referred as the angle at the AP of the received path. Localising MD based on DoAs of LoS paths have attracted considerable attention under high angular resolution of antenna arrays and independent of absolute time measurement (clock of sensors can be either synchronized or unsynchronized).

Under high angular resolution, the DoAs of LoS paths can provide precise estimation of MDs. However, the performance of this localisation scheme is sensitive with directional errors. The directional errors come from two aspects: NLoS bias, and array directional error. The NLoS path results in extra angular bias (NLoS bias) in DoA measurement. Therefore, only DoAs of LoS paths are used to determine MD locations. There are numerous methods designed for reducing effect of NLoS bias in the literature [10–13, 15]. The array directional error implies deviations in the orientation of the array placement which is rarely noticed. In this Chapter, we focus on self-calibrating the array directional errors and estimating multiple MDs simultaneously. Remark that our proposed method also can be used to reduce adverse effect of NLoS bias by separating LoS/NLoS path and discarding the NLoS paths. Since the idea behind our method is exploiting the angular geometry consistency of LoS paths between APs and MDs, but NLoS paths have arbitrary DoAs.

In the literature, various kinds of self-calibration methods have been proposed. These methods are following the principle of typical *two-step localisation* approaches that first measure immediate parameters (i.e., DoA), and then fuse them to achieve localisation. Traditional self-calibration techniques can be roughly divided into two categories. One calibrates errors in the first step to improve DoA measurements [87–89] and then fuse them for localisation. The other methods such as [90–92] calibrate errors in the data fusing step, using the uncalibrated measurements directly.

However, these self-calibration methods in the literature rely on data association [85], thus resulting in a big challenge. Precisely, in joint MD localisation, DoA of each MD DoA with respect to each AP is measured in advance. These DoAs measurements are required to be associated with certain MDs, and then fused together to achieve localisation of the MD. Nevertheless, the task of associating the DoAs of all the APs to multiple MDs is NP-hard [85], leading to huge computational complexity. On the other hand, *two-step localisation* may introduce additional noise and lead to error propagation.

To avoid data association, the concept of direct localisation has been proposed in [5,15,93]. By mapping the MD positions to the received signals, a direct localisation approach eliminates the intermediate measurements (DoAs) and estimate the MD positions from the received signals directly. To the best knowledge of authors, self-calibration on array directional errors in direct localisation has not been widely studied [94]. Existing works [95–97] studied this problem under the condition of positioning single MD with multiple snapshots, which is not suitable for multiple MDs localisation using a single snapshot as we consider. The research in [94] extends the localisation model to multiple MDs, however, it aims at self-calibration on sensor gain and phase errors, which are different from the error considered in this chapter. Therefore, to achieve direct passive joint multiple MDs localisation with NLoS biases reduction and array directional errors self-calibration, a novel approach is in demand.

To propose our approach, we refer to compressed sensing based ideas [61] and construct

a direct joint localisation formulation based on sparse recovery. Typical sparse recovery methods discretize the parameter space into finite grid points and assume that true parameters exactly lie on these grids [65, 67, 70]. However, MD locations are distributed in some continuous domain, and may be off these grids. The grid mismatch problem could result in serious degradation in performance [69]. To avoid this problem, a super-resolution technique called ANM has been developed [6, 70]. Instead of being divided into grids, parameters are estimated by identifying a small number of atoms parameterized over continuous parameter space. Particularly, when the atoms are Fourier bases with their line spectrum being the unknown parameters that are to estimate, ANM is equivalent to some SDP [98], and can be efficiently accomplished by some off-the-shelf convex toolboxes, such as CVX [73].

There are two challenges in the way of applying ANM to passive joint MD localisation with error self-calibration. On the one hand, most ANM based methods depend on the prior of accurate array manifold, which violates the fact that some APs may have inevitable unknown array directional errors. On the other hand, the corresponding atoms are not Fourier structured w.r.t. the MD locations. As a consequence, the efficient SDP solver in [51, 75] is not directly applicable.

To solve the above problems, we establish a non-convex optimisation problem based on group sparse recovery, which exploits the geometric consistency that received LoS paths are originating from a common MDs. On the one hand, the joint estimation of MD locations and array directional errors self-calibration are achieved simultaneously. On the other hand, this formulation is based on the received signals and not restricted to Fourier structure. To solve the proposed optimisation problem, we give a novel method named as group sparsity exploitation for self-calibration (GSE-SC). This method iteratively adds MDs estimation to the support set one by one, while locally improving the estimates of MD locations and array directional errors by a two-loop alternating gradient descent. This descent method guarantees the decrease of the cost function in each iteration and the two-loop structure helps avoid local minimums and self-calibrate array directional er-

rors. Numerical simulation results verify its feasibility and the advantage over traditional methods.

3.2 System Model

In this section, the signal model of joint MD localisation with array directional errors is given. We consider the scenario with J APs equipped with passive uniform linear arrays (ULA) and I MDs in two-dimensional (2D) plane. As for the j-th AP, it is composed of N_R antenna arrays with the first element located at $\boldsymbol{AP}_j = [AP_j^x, AP_j^y]^{\mathrm{T}}$, the interval between adjacent elements is denoted by d_j, and the default angle of antenna array versus the horizontal axis is denoted by α_j. The i-th MD location is denoted by $\boldsymbol{m}_i = [m_i^x, m_i^y]^{\mathrm{T}}$. The schematic diagram of the scenario is shown in Fig 3.1. At j-th AP, the angle of the i-th MD w.r.t. the normal direction of antenna array, i.e., the DoA, is denoted by $\theta_j(\boldsymbol{m}_i)$, expressed as

$$\theta_j(\boldsymbol{m}_i) = \arctan\frac{m_i^x - AP_j^x}{m_i^y - AP_j^y} + \alpha_j, \tag{3.1}$$

for $j = 1, \ldots, J$ and $i = 1, \ldots, I$.

In practical scenario, the perturbations in the antenna array deployment are inevitable, resulting in that the actual parameters α_j are different from their pre-assumed counterparts, denoted by α_j'. We denote array directional errors by $\delta_j = \alpha_j' - \alpha_j \in \Delta$, where Δ is a rough range of δ_j. The DoA measurements of LoS paths, which are crucial in joint MD localisation, are sensitive to array directional errors. Considering effect of array directional error, the DoA of the i-th MD w.r.t. the pre-assumed j-th AP, denoted by $\theta_j'(\boldsymbol{m}_i)$, is expressed as

$$\theta_j'(\boldsymbol{m}_i) = \arctan\frac{m_i^x - AP_j^x}{m_i^y - AP_j^y} + \alpha_j' = \theta_j(\boldsymbol{m}_i) + \delta_j, \tag{3.2}$$

for $j = 1, \ldots, J$ and $i = 1, \ldots, I$. Note that array directional error is highly structured, prevalent on mobile platforms, but not widely studied yet. What is the effect of the kind

of error and how to achieve self-calibration well are still open problems, which are our focus in this chapter. The geometry of the arrays and MDs is depicted in Fig 3.1.

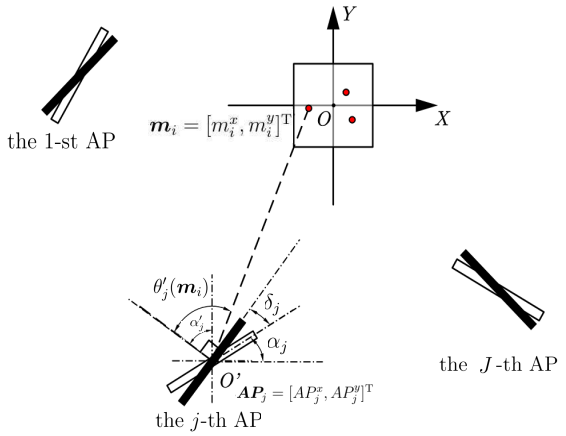


Figure 3.1: The geometry of the MDs and arrays in XOY plane. White blocks, black blocks and dots represent the actual, pre-assumed deployment of arrays and location of MDs, respectively.

Before introducing the signal model, there are some pre-conditions listed as follows: 1) The baseband signals of the transmitted signals from MDs are assumed to be narrow-band and wide-sense stationary. 2) The MDs are in the far field [99]. 3) In each array, one sample is taken at the same time as a single-snapshot. Then, the received signal at j-th AP is given by

$$\boldsymbol{r}_j = \sum_{i=1}^{I} \gamma_{i,j} \boldsymbol{a}_j(f_{i,j}) + \boldsymbol{v}_j, \tag{3.3}$$

where $\boldsymbol{r}_j \in \mathbb{C}^{N_R \times 1}$, $\gamma_{i,j}$ is a complex coefficient characterizing the unknown attenuation of the transmitted signals from the i-th MD to the j-th AP, $f_{i,j} = \frac{d_j}{\lambda} \sin\theta_j(\boldsymbol{m}_i)$ is often referred to the spatial frequency, $\boldsymbol{v}_j \in \mathbb{C}^{N_R \times 1}$ denotes the additive noise, and the steering vector $\boldsymbol{a}_j(\cdot)$ is denoted by

$$\boldsymbol{a}_j(f_{i,j}) = \left[e^{j2\pi f_{i,j}\cdot 0}, \ldots, e^{j2\pi f_{i,j}\cdot(N_R-1)}\right]^{\mathrm{T}}, \tag{3.4}$$

where $j = 1, \ldots, J$ and $i = 1, \ldots, I$.

By substituting (3.2) to (3.3), we recast the received signals as

$$\boldsymbol{r}_j = \sum_{i=1}^{I} \gamma_{i,j} \boldsymbol{a}_j(\boldsymbol{m}_i, \delta_j) + \boldsymbol{v}_j, \tag{3.5}$$

where the steering vector $\boldsymbol{a}_j(\cdot,\cdot)$ is denoted by

$$\boldsymbol{a}_j(\boldsymbol{m}, \delta_j) = \left[e^{j2\pi \frac{d_j}{\lambda} \sin\left(\theta'_j(\boldsymbol{m}) - \delta_j\right) \cdot 0}, \ldots, e^{j2\pi \frac{d_j}{\lambda} \sin\left(\theta'_j(\boldsymbol{m}) - \delta_j\right) \cdot (N_R - 1)} \right]^{\mathrm{T}}. \tag{3.6}$$

For ease of notation we use $\boldsymbol{\delta} = [\delta_1, \ldots, \delta_J]^T$, $\boldsymbol{M} = [\boldsymbol{m}_1, \ldots, \boldsymbol{m}_I]$, $\boldsymbol{\Gamma} = \{\gamma_{i,j}, \text{ for } i = 1, \ldots, I, \text{and } \ j = 1, \ldots, J\}$ and $\mathcal{R} = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_J\}$ to represent the directional errors vector, position matrix, complex coefficient set and measurement set, respectively. In (3.5), we note that $\mathcal{R}$, M, N_R and the function of $\theta'_j(\cdot)$ as (3.2) are known. Unknown parameters $\{\boldsymbol{M}, \boldsymbol{\Gamma}, \boldsymbol{\delta}\}$ and I are to be estimated.

Note that (3.5) is an *incoherent* model, which means that the coefficient $\gamma_{i,j}$ of the i-th MD varies in amplitude and phase w.r.t. different APs and the relationship between them is unknown. The challenge of recovering MD locations $\boldsymbol{m}$ from (3.5) lies in the highly non-convex structure for frequencies $f_{i,j}$ w.r.t. $\boldsymbol{m}_i$ and unknown array directional errors δ_j, $j = 1, \ldots, J$ and $i = 1, \ldots, I$.

3.3 Necessary Conditions

In this section, we explain the feasibility of self-calibrating array directional errors in joint MD localisation from the aspect of necessary conditions.

First, we give a necessary condition in our scenario by comparing the number of equations and unknowns. Particularly, for the equations, the given received signals set $\mathcal{R}$ embodies a series of equations as (3.3), whose number is IJ. For the unknowns, there are I MD positions in the 2D plane and J array directional errors to be estimated, which is $2I + J$ in total. A necessary condition to achieve the joint estimation of $\{\boldsymbol{M}, \boldsymbol{\delta}\}$ is to guarantee

that the number of equations is larger than the unknowns, i.e.,

$$IJ \geqslant 2I + J. \tag{3.7}$$

From (3.7), it is not hard to find that when $I \geqslant 2$, many enough APs J guarantee (3.7), yielding the success of self-calibrating array directional errors. However in practice, due to the impact of noises, sensor system settings, sensor and MD spatial geometry and so on, more APs than the threshold in (3.7) are generally in demand.

Especially, in case of $I = 1$, (3.7) does not hold no matter what J is, yielding the inevitable failure of the joint estimation of $\{\boldsymbol{m}, \boldsymbol{\delta}\}$. It implies that in passive joint MD localisation, we are not able to calibrate all the directional errors when $I = 1$. Here we further explore this phenomenon as follows: If $I = 1$, for any true values $\{\boldsymbol{m}_1, \boldsymbol{\delta}\}$ and a false MD position $\boldsymbol{m}_1'$, it is always available to estimate the directional errors as $\delta_j' = \theta_j'(\boldsymbol{m}_1') - \theta_j'(\boldsymbol{m}_1) + \delta_j$, such that the following equation is satisfied:

$$\sin\big(\theta_j'(\boldsymbol{m}_1) - \delta_j\big) = \sin\big(\theta_j'(\boldsymbol{m}_1') - \delta_j'\big), \tag{3.8}$$

where $j = 1, \ldots, J$. Therefore, $\{\boldsymbol{m}_1', \boldsymbol{\delta}'\}$ yields the same signals as $\{\boldsymbol{m}_1, \boldsymbol{\delta}\}$ according to (3.5) when $\boldsymbol{\Gamma}$ and noises are fixed. This indicates that $\{\boldsymbol{m}_1', \boldsymbol{\delta}'\}$ and $\{\boldsymbol{m}_1, \boldsymbol{\delta}\}$ can not be distinguished only from the received signals $\mathcal{R}$. Correspondingly, the joint estimation of $\{\boldsymbol{M}, \boldsymbol{\delta}\}$ can be either one of them and will fail eventually.

To avoid the estimation failure in case of $I = 1$, we assume that there are some precise antenna without array directional errors. Denote the number of APs with array directional errors by J_r, $1 < J_r < J$. A necessary condition for successful localisation in case of $I = 1$ is $J \geqslant 2 + J_r$. On the other hand, we consider another scenario assuming that the directional errors in multiple APs are the same. Similar assumptions on the errors can be seen in [94], where distributed sensor observations are obtained by sampling different moments of a single moving array with errors, and the position relationship of each array

at different samplings is assumed to be exactly known. Under this condition, we have 3 unknowns (2 for MD positions and 1 for directional error) and J equations, yielding the necessary condition as $J \geqslant 3$.

In summary, we explain the potential of self-calibrating the directional errors in terms of necessary conditions when $I \geqslant 2$. For the $I = 1$ case, where the necessary condition always does not hold, we also give the additional conditions for the self-calibration to be available. Note that in the sequel, our proposed method is suitable for both cases.

3.4 Objective Formulation

In this section, a group sparsity based optimisation is proposed for joint localising MDs and self-calibration. In subsection 3.4.1, we introduce the motivation of using group sparsity. In subsection 3.4.2, we review the super-resolution CS methods such as ANM and group ANM. Finally, in subsection 3.4.3, we establish a new group sparse recovery model that simultaneously accomplish calibration of array directional errors and estimation of MD positions.

3.4.1 Group Sparsity Exploitation

Considering the sparsity of MDs in the spatial space, it is a common belief of using sparse recovery methods for localisation. In the literature, the general sparse recovery methods incorporates multiple distributed APs and realises a direct inverse mapping from their measurements to MDs, which is referred to as direct localisation. The benefits of direct localisation technique are avoiding error propagation in multistage processing and data association. However, simple application of atomic norm $\|\cdot\|_{\mathcal{A}}$ for exploiting sparsity separately may result in possible spurious MDs estimations.

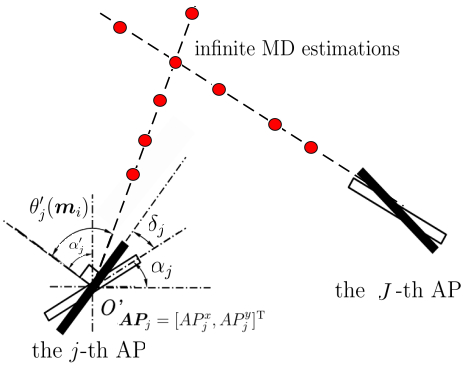
To address this issue, group sparsity [100], which enhances the mapping from common

MDs to the measurements for all the sensors, is generally used in distributed sensor networks and proven to achieve better performance [101–103]. Particularly in joint MD localisation with array directional errors considered, group sparsity exploitation also inhibits the bias of MD localisation caused by the sensor errors to some extent. Therefore, we exploit group sparsity for the data fusion of multiple APs with errors in this paper. The demonstration of showing difference between sparsity and group sparsity method is shown in Fig 3.2. As shown in Fig 3.2(a), exploiting sparsity in measurements of each AP separately lead to infinite estimations. Finding intersection of DoAs of two LoS paths will involve error propagation and data association. As shown in Fig 3.2(b), the geometric consistency indicates the truth that received LoS paths are originating from common MD, but DoA of NLoS path is arbitrary. It means that LoS paths admit the group sparsity. The precise estimation can be obtained by designing group sparsity over common MD. On the other hand, the group sparsity formulation also can be applied for separating LoS path from multipath, thus effect of NLoS bias is reduced.

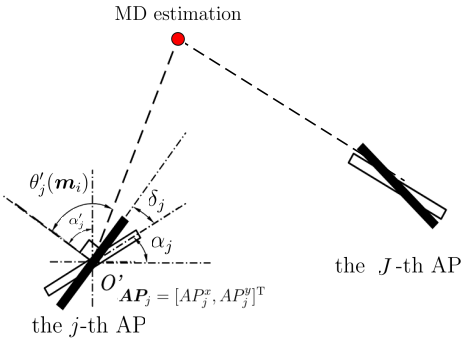
In the literature, methods with group sparsity are usually grid-based, i.e., they divide the parameter domain into many grids and assume that the true values are on the grids. However, this assumption is hardly satisfied in practice and grid-based sparse recovery inevitably encounters the grid mismatch problem, which degrades the performance [69]. To this end, super-resolution techniques such as the typical ANM methods are proposed to tackle the aforementioned challenge. These methods avoid dividing discrete grids and achieve sparse recovery on continuum. There are some existing works [98, 104] that extend ANM methods to exploit group sparsity, named as group ANM here. In the next subsection, for better clarity, we give a brief review of ANM and group ANM methods.

3.4.2 ANM and Group ANM

A brief review of ANM and group ANM is presented in subsection 3.4.2.1 and subsection 3.4.2.2, respectively.



(a) Sparsity Estimation



(b) Group Sparsity Estimation

Figure 3.2: Difference between sparsity and group sparsity method.

3.4.2.1 ANM

Typical ANM stems from the frequency estimation from a superposition of complex sinusoids, given by

$$\boldsymbol{x} = \sum_{i=1}^{I} \gamma_i \boldsymbol{a}(f_i, \phi_i) \in \mathbb{C}^{N\times 1}, \tag{3.9}$$

where $f_i \in [0, 1)$, $\gamma_i > 0$ and $\phi_i \in [0, 2\pi)$ are unknown parameters to estimate, denoting frequency, amplitude and phase of the i-th sinusoid, respectively, $\boldsymbol{a}(f, \phi) \in \mathbb{C}^{N\times 1}$ with entry $[\boldsymbol{a}(f, \phi)]_n = e^{j(2\pi f(n-1)+\phi)}$ is defined as a linear spectral atom, $n = 1, \ldots, N$, and I denotes the number of sinusoids. Denote $\mathcal{A} = \{\boldsymbol{a}(f, \phi) : f \in [0, 1), \phi \in [0, 2\pi)\}$ as the set of atoms.

The atomic norm $\|\cdot\|_{\mathcal{A}}$ is defined by identifying its unit ball with the convex hull of $\mathcal{A}$, $\text{conv}(\mathcal{A})$, given by

$$\begin{aligned} \|\boldsymbol{x}\|_{\mathcal{A}} &= \inf\{t > 0 : \boldsymbol{x} \in t\text{conv}(\mathcal{A})\} \\ &= \inf_{\substack{\gamma_i > 0, \phi_i \in [0,2\pi), \\ f_i \in [0,1)}} \left\{ \sum_{i}^{I} \gamma_i : \boldsymbol{r} = \sum_{i}^{I} \gamma_i \boldsymbol{a}(f_i, \phi_i) \right\}. \end{aligned} \tag{3.10}$$

The atomic norm in (3.10) can be viewed as the continuous counterpart of the L_1 norm in grid-based CS methods. ANM is to solve a convex optimisation problem minimising (3.10) under certain constraints. Especially, ANM problems can be equivalently solved by SDP with performance guarantee when the frequencies $\{f_i\}_{i=1}^{I}$ are well separated [70].

3.4.2.2 Group ANM

Based on ANM, we consider that there are J measurement vectors of I complex sinusoids, with the j-th vector given by

$$\boldsymbol{x}_j = \sum_{i}^{I} \gamma_{i,j} \boldsymbol{a}(f_i, \phi_{i,j}) \in \mathbb{C}^{N_R}, \tag{3.11}$$

where $f_i \in [0,1)$, $\gamma_{i,j} > 0$ and $\phi_{i,j} \in [0, 2\pi)$ denote the frequency, coefficient and phase of j-th measurement w.r.t. the i-th sinusoid, respectively. These coefficients and phases varying with j indicate the incoherence of the signal models. In the (3.11), we could also choose to absorb the phase $\phi_{i,j}$ into the coefficient $\gamma_{i,j}$ as we did in (2.1).

Unlike the ANM promoting sparsity for each measurement separately, group ANM provides features across all the measurements with joint sparsity. In our framework, the group ANM encourages that received LoS paths are originating from the common MDs. The group atomic norm, denoted by $\|\cdot\|_{\mathcal{G}}$, can be expressed as

$$\|\mathcal{X}\|_{\mathcal{G}} = \inf_{\substack{\gamma_{i,j}>0,\\ \phi_{i,j}\in[0,2\pi),\\ f_i\in[0,1)}} \left\{ \sum_i^I \sqrt{\sum_{j=1}^J \gamma_{i,j}^2} : \boldsymbol{x}_j = \sum_i \gamma_{i,j} \boldsymbol{a}\left(f_i, \phi_{i,j}\right), j = 1, \ldots, J \right\}, \tag{3.12}$$

where $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_J\}$ denotes the set of multiple measurements. The group atomic norm in (3.12) is often regarded as the continuous counterpart of the $L_{2,1}$ norm in the MMV model of grid-based CS methods [105]. Group ANM is to solve a convex optimisation problem minimizing (3.12) under certain constraints. Similar with ANM, group ANM also can be transformed into SDP which is a convex optimisation with finite constrains [98].

Here we explain that the group ANM model and corresponding SDP can not accurately address joint MD localisation with AP self-calibration. At a first glance, the MMV model (3.11) is similar with the received signals (3.5) using multiple sensors, indicating the potential of using group ANM for our scenario considered. However, there are at least two significant differences between (3.11) and (3.5): 1) There exists unknown directional errors $\boldsymbol{\delta}$ in (3.5), which is not present in (3.11); 2) Atoms in (3.11) and (3.5) are different, because f_i in (3.11) are frequencies of line spectrum, while $\boldsymbol{m}_i$ in (3.5) are not, due to the non-linearity of the map $\boldsymbol{m}_i \mapsto f_{i,j}$. These differences hold back applying group ANM directly in our scenario, while motivating us to incorporate self-calibration in the group

ANM model, which we will introduce below.

3.4.3 Group ANM Model Incorporating Self-calibration

In this subsection, based on group ANM, we propose a new sparse recovery framework that jointly address group sparsity exploitation and AP self-calibration.

In order to self-calibrate the directional error, we heuristically formulate directional errors $\boldsymbol{\delta}$ into the group atomic norm and then propose a new sparsity representation operator for joint MD localisation, denoted by $\|\cdot\|_{\mathcal{S}}$,

$$\|\boldsymbol{\mathcal{X}}\|_{\mathcal{S}} = \inf_{\boldsymbol{m},\boldsymbol{\Gamma},\boldsymbol{\delta}} \left\{ \sum_{i}^{I} \sqrt{\sum_{j=1}^{J} |\gamma_{i,j}|^2} : \boldsymbol{x}_j = \sum_{i} \gamma_{i,j} \boldsymbol{a}_j \left(\boldsymbol{m}_i, \delta_j\right), j = 1, \ldots, J \right\}, \tag{3.13}$$

where $\boldsymbol{\mathcal{X}} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_J]$ is the set of noise-free measurements in joint MD localisation with sensor self-calibration; the atoms $\boldsymbol{a}_j\left(\boldsymbol{m}_i, \delta_j\right)$ follows the definition in (3.6). Based on the proposed operator (3.13), we can formulate a denoising lasso problem to jointly estimate $\{\boldsymbol{M}, \boldsymbol{\Gamma}, \boldsymbol{\delta}\}$:

$$\min_{\boldsymbol{m},\boldsymbol{\Gamma},\boldsymbol{\delta}} \frac{1}{2} \sum_{j=1}^{J} \|\boldsymbol{r}_j - \boldsymbol{x}_j\|_2^2 + \mu \|\boldsymbol{\mathcal{X}}\|_{\mathcal{S}}, \tag{3.14}$$

where $\mu > 0$ is the regularization parameter characterizing group sparsity and the bias between received measurements $\boldsymbol{r}_j$ and noise-free measurements $\boldsymbol{x}_j$.

Different from the group ANM model, (3.14) properly addresses antenna array directional errors in passive joint MD localisation. However, difficulty emerges since (3.14) is a non-convex optimisation problem with multiple unknowns as $\{\boldsymbol{M}, \boldsymbol{\Gamma}, \boldsymbol{\delta}\}$. This is a consequence of the fact that the proposed operator in (3.13) is not corresponding to convex hulls as atomic norms due to the existence of $\boldsymbol{\delta}$. To solve the above problem, we propose a novel method, detailed in Section 3.5.

3.5 Solving Optimisation Problems

In this section, we propose GSE-SC method to solve (3.14), yielding estimates of unknown parameters $\{\boldsymbol{M}, \boldsymbol{\Gamma}, \boldsymbol{\delta}\}$. In each step, we add MD estimation to the support, while implementing local improvement on the unknown variables. Here, local improvement is based on a two-loop alternating gradient descent to guarantee the decline of the cost function and avoid possible local minima. Meanwhile, the support set is pruned by removing weak elements when they are found not contributing to reducing the loss function (3.14). To clearly present the proposed GSE-SC, the framework is introduced in subsection 3.5.1, and local improvement is detailed in subsection 3.5.2.

3.5.1 Framework

Our proposed method, GSE-SC, is based on such a scheme: the MD estimates are added to the support set one by one, then all the unknown variables are improved locally by alternating gradient descent and support set prune. Before introducing this algorithm framework, we denote the counter of steps as t. Assume that after $t-1$ steps, we obtain the position candidate matrix $\widehat{\boldsymbol{M}}^{t-1}$ that contains $t-1$ MD position estimates, i.e., $\widehat{\boldsymbol{M}}^{t-1} = [\hat{\boldsymbol{m}}_1, \ldots, \hat{\boldsymbol{m}}_{t-1}]$, the corresponding coefficient matrix $\widehat{\boldsymbol{\Gamma}}^{t-1}$ with the (i,j)-th entry $\hat{\gamma}_{i,j}^{t-1}$, as well as the directional errors $\hat{\boldsymbol{\delta}}^{t-1} = [\hat{\delta}_1^{t-1}, \ldots, \hat{\delta}_J^{t-1}]^{\mathrm{T}}$.

It is well-known that initialisation plays a crucial role in the iterative methods. In the t-th step, proposed GSE-SC method initialises the MD location through using grid-based CS methods for accuracy, particularly by solving the following $L_{2,1}$ norm based lasso problem as

$$\widehat{\boldsymbol{S}} = \underset{\boldsymbol{S}}{\operatorname{argmin}} \ \frac{1}{2} \sum_{j=1}^{J} \left\| \boldsymbol{e}_j^{t-1} - \boldsymbol{A}_j \boldsymbol{S}_{1:G,j} \right\|_2^2 + \mu_{cs} \|\boldsymbol{S}\|_{2,1}, \tag{3.15}$$

where $\mu_{cs} > 0$ is the regularization parameter, the $\|\cdot\|_{2,1}$ is $L_{2,1}$ norm which is defined as

$$\|\boldsymbol{S}\|_{2,1} = \sum_{g=1}^{G} \sqrt{\sum_{j=1}^{J} |\boldsymbol{S}_{g,j}|^2}, \tag{3.16}$$

and $\boldsymbol{A}_j$ is the dictionary matrix constructed by uniformly dividing the range of MD position $\mathcal{U}$ into $G = g_x \times g_y$ grids as $\bar{\boldsymbol{m}}_g, g = 1, \ldots, G$, and concatenating the corresponding atoms, i.e., $\boldsymbol{A}_j = [\cdots, a_j(\bar{\boldsymbol{m}}_g, 0), \cdots]$. Here, $\boldsymbol{e}_j^{t-1}$ is the residual signal of the j-th sensor in the $(t-1)$-th step, given by

$$\boldsymbol{e}_j^{t-1} = \boldsymbol{r}_j - \sum_{i=1}^{t-1} \hat{\gamma}_{i,j}^{t-1} \boldsymbol{a}_j(\hat{\boldsymbol{m}}_i, 0). \tag{3.17}$$

Then once the convex optimisation problem (3.15) is solved. The t-th initialisation of MD position, $\hat{\boldsymbol{m}}_t$, is then derived from the optimal solution $\hat{\boldsymbol{S}}$ by selecting the grid with the largest intensity as

$$\hat{\boldsymbol{m}}_t = \bar{\boldsymbol{m}}_{\hat{g}}, \ \hat{g} = \arg\max_{g} \|\hat{\boldsymbol{S}}_{g,1:J}^{\mathrm{T}}\|_2. \tag{3.18}$$

The t-th corresponding coefficient matrix $\boldsymbol{\Gamma}^t$ is then initialised by minimising the temp residual as

$$\min_{\boldsymbol{\Gamma}_j} \left\| \boldsymbol{r}_j - \sum_{i=1}^{t} \gamma_{i,j} \boldsymbol{a}_j(\hat{\boldsymbol{m}}_i, 0) \right\|_2^2 \tag{3.19}$$

for $j = 1, \ldots, J$, which can be solved via least squares, yielding the closed form solutions as

$$\widehat{\boldsymbol{\Gamma}}_j^t = (\boldsymbol{B}_j^H \boldsymbol{B}_j)^{-1} \boldsymbol{B}_j \boldsymbol{r}_j, \tag{3.20}$$

where $\boldsymbol{B}_j = [\boldsymbol{a}_j(\hat{\boldsymbol{m}}_1, 0), \ldots, \boldsymbol{a}_j(\hat{\boldsymbol{m}}_t, 0)]$. Without prior knowledge, directional errors $\hat{\boldsymbol{\delta}}^t$ is thus initialized to be $\mathbf{0}$.

The initialized $\{\boldsymbol{M}, \boldsymbol{\Gamma}, \boldsymbol{\delta}\}$ is then improved locally by an alternative gradient descent method to guarantee the decline of the cost function in (3.14) and avoid possible local minimums, which is shown in the next subsection. Consequently, we will obtain $\{\widehat{\boldsymbol{m}}^t, \widehat{\boldsymbol{\Gamma}}^t, \hat{\boldsymbol{\delta}}^t\}$, which implies the residuals $\{\boldsymbol{e}_j^t\}_{j=1}^J$. Then we repeat the procedures from

(3.15) to (3.20) to continue the refinement of $\{\boldsymbol{M}, \boldsymbol{\Gamma}, \boldsymbol{\delta}\}$ and this iteration terminates when t reaches the maximum number of steps, $t_{\max}$. Generally, $t_{\max}$ is set to be larger than the MD number I.

Lastly, for the identification of the MD number I, we assume that the true I is pre-known here. Therefore, the estimation of $\{\boldsymbol{m}, \boldsymbol{\Gamma}, \boldsymbol{\delta}\}$ is actually given by $\left\{\widehat{\boldsymbol{m}}^I, \widehat{\boldsymbol{\Gamma}}^I, \hat{\boldsymbol{\delta}}^I\right\}$. Remark that there are specific studies on this issue and some existing principles such as AIC [106] and BIC [107] have been widely used. Our main contribution is not in this and previous related works are available for reference. In Algorithm 3, we retain the framework with unknown I for further extension on the identification of I.

3.5.2 Joint Estimation of $\{\boldsymbol{M}, \boldsymbol{\Gamma}, \boldsymbol{\delta}\}$

In this subsection, we detail the local improvement procedure, which aims to minimise the cost function in (3.14) by an alternating gradient descent method.

The difficulties of solving the optimisation problem (3.14) reflect on two aspects. On the one hand, (3.14) is non-convex, most well known optimisation methods do not guarantee a global optimal solution. On the other hand, array directional errors $\boldsymbol{\delta}$ are key unknowns to the non-convexity and complexity of (3.14). Therefore, how to self-calibrate $\boldsymbol{\delta}$ by a precise descent style is important and desired.

In response to the challenges above, our strategy has a two-loop structure called outer loop and inner loop, respectively. Particularly, we simultaneously improve the joint estimation of $\{\boldsymbol{M}, \boldsymbol{\Gamma}, \boldsymbol{\delta}\}$ and $\{\boldsymbol{M}, \boldsymbol{\Gamma}\}$ by gradient descent in the outer and inner loop, respectively. In between, we prune the support set of $\boldsymbol{M}$ by removing weak elements as support$(\cdot)$, a typical procedure commonly seen in ANM methods such as ADCG [108] and greedy CoGEnT [109]. In summary, the role of this structure is reflected in the following three points:

A1) The outer loop with gradient descent on $\{\boldsymbol{M}, \boldsymbol{\Gamma}, \boldsymbol{\delta}\}$ guarantees to bring the cost

function in (3.14) down.

A2) The inner loop inside an outer loop is used to refine $\{\boldsymbol{M}, \boldsymbol{\Gamma}\}$ beforehand. In this way, we avoid many possible local minimums in the outer loop iteration due to the imprecise initialisation of $\{\boldsymbol{M}, \boldsymbol{\Gamma}\}$.

A3) The two-loop structure is designed for self-calibrating the directional errors precisely. This is reflected in that after the inner loop, $\{\boldsymbol{M}, \boldsymbol{\Gamma}\}$ reach their local minimums for a certain $\boldsymbol{\delta}$. Then, the iteration direction and step size of gradient descent w.r.t. $\{\boldsymbol{M}, \boldsymbol{\Gamma}, \boldsymbol{\delta}\}$ in the outer loop will mainly depend on the errors $\boldsymbol{\delta}$, which is beneficial for precise and fast estimation of $\boldsymbol{\delta}$.

Therefore, this two-loop structure well achieves our purpose.

Particularly, we introduce the details of the two-loop structure here. As a preliminary, we use k as the iteration counter of the outer loop, denote intermediate estimate by $\hat{\cdot}^{k-1}$ and initial $\hat{\cdot}^{k=0}$, where $\cdot$ belongs to the set $\{\boldsymbol{M}, \boldsymbol{\Gamma}, \boldsymbol{\delta}\}$ or $\{\boldsymbol{m}_i, \gamma_{i,j}, \delta_j\}$. Also denote by $C(\boldsymbol{M}, \boldsymbol{\Gamma}, \boldsymbol{\delta})$ the cost function in (3.14) and by $\nabla_{\circ} C(\widehat{\boldsymbol{M}}^{k-1}, \widehat{\boldsymbol{\Gamma}}^{k-1}, \hat{\boldsymbol{\delta}}^{k-1})$ the partial derivative of the cost function w.r.t. $\circ$ at $(\widehat{\boldsymbol{M}}^{k-1}, \widehat{\boldsymbol{\Gamma}}^{k-1}, \hat{\boldsymbol{\delta}}^{k-1})$, where $\circ$ belongs to the set $\{\boldsymbol{m}_i, \gamma_{i,j}, \delta_j\}$. For the convenience of presentation, we remain the calculation of these derivatives in Appendix A.

In the k-th outer loop iteration, the inner loop is first carried out, i.e., $\{\hat{\boldsymbol{m}}_i^{k-1}, \hat{\gamma}_{i,j}^{k-1}\}$ is renewed by gradient descent repeatedly until the maximum number of repetitions $l_{\max}$, which is large enough to guarantee the convergence. Note that we use this convergence criterion for brief expression and other criteria such as stopping the iteration when the gradients are small enough are also available. Here, let l denote the index of the inner loop repetitions. Then, we perform

$$\begin{cases} \hat{\boldsymbol{m}}_i^l = \hat{\boldsymbol{m}}_i^{l-1} - \kappa^l \nabla_{\boldsymbol{m}_i} C(\widehat{\boldsymbol{M}}^{l-1}, \widehat{\boldsymbol{\Gamma}}^{l-1}, \hat{\boldsymbol{\delta}}^{k-1}), \\ \hat{\gamma}_{i,j}^l = \hat{\gamma}_{i,j}^{l-1} - \kappa^l \nabla_{\gamma_{i,j}} C(\widehat{\boldsymbol{M}}^{l-1}, \widehat{\boldsymbol{\Gamma}}^{l-1}, \hat{\boldsymbol{\delta}}^{k-1}), \end{cases} \tag{3.21}$$

where κ^l denotes the step size in the l-th repetition and is determined via Armijo line search [110]. When the inner loop ends, we achieve the updated parameters denoted by $\{\widetilde{\boldsymbol{M}}^k, \widetilde{\boldsymbol{\Gamma}}^k\}$.

The aim of pruning support step is to produce sparse solution. Given the updated parameters $\{\widetilde{\boldsymbol{M}}^k, \widetilde{\boldsymbol{\Gamma}}^k\}$ of the inner loop, the support is pruned by

$$\widehat{\boldsymbol{M}}^t = \text{support}(\widetilde{\boldsymbol{\Gamma}}^t), \tag{3.22}$$

where

$$\widetilde{\boldsymbol{\Gamma}}_j^t = (\widetilde{\boldsymbol{B}}_j^H \widetilde{\boldsymbol{B}}_j)^{-1} \widetilde{\boldsymbol{B}}_j \boldsymbol{r}_j, \tag{3.23}$$

and

$$\widetilde{\boldsymbol{B}}_j = [\boldsymbol{a}_j(\tilde{\boldsymbol{m}}_1, 0), \ldots, \boldsymbol{a}_j(\tilde{\boldsymbol{m}}_t, 0)]. \tag{3.24}$$

By pruning support, the source locations are removed from candidate set when they are found not contributing to reducing the loss function. This is motivated by the drawback of greed step: the source location added at previous iteration may not be helpful comparing with other later added sources.

Array directional error $\hat{\boldsymbol{\delta}}^k$ is then updated by gradient descent together with the immediate parameters $\{\widehat{\boldsymbol{M}}^k, \widehat{\boldsymbol{\Gamma}}^k\}$ as

$$\begin{cases} \hat{\boldsymbol{m}}_i^k = \hat{\boldsymbol{m}}_i^k - \kappa^k \nabla_{\boldsymbol{m}_i} C(\widehat{\boldsymbol{M}}^k, \widehat{\boldsymbol{\Gamma}}^k, \hat{\boldsymbol{\delta}}^{k-1}), \\ \hat{\gamma}_{i,j}^k = \hat{\gamma}_{i,j}^k - \kappa^k \nabla_{\gamma_{i,j}} C(\widehat{\boldsymbol{M}}^k, \widehat{\boldsymbol{\Gamma}}^k, \hat{\boldsymbol{\delta}}^{k-1}), \\ \hat{\delta}_j^k = \hat{\delta}_j^{k-1} - \kappa^k \nabla_{\delta_j} C(\widehat{\boldsymbol{M}}^k, \widehat{\boldsymbol{\Gamma}}^k, \hat{\boldsymbol{\delta}}^{k-1}), \end{cases} \tag{3.25}$$

where the k-th step size κ^k is also determined via Armijo line search. The alternating iterations, (3.21) and (3.25), continue to be carried out until k reaches the maximum number of steps $k_{\max}$, yielding the end of the outer loop. The specific GSE-SC method is

summarized in Algorithm 3.

Algorithm 3: GSE-SC

Input: Signal $\mathcal{R}$ and parameters μ, κ^0, $t_{\max}$, $k_{\max}$, $i_{\max}$, $\widehat{\boldsymbol{M}}^0 = \widehat{\boldsymbol{\Gamma}}^0 = \varnothing$, $\hat{\boldsymbol{\delta}}^0 = \mathbf{0}$, $\boldsymbol{e}_j^0 = \boldsymbol{r}_j$

Output: $\widehat{\boldsymbol{M}}^{t_{\max}}$, $\widehat{\boldsymbol{\Gamma}}^{t_{\max}}$ and $\hat{\boldsymbol{\delta}}^{t_{\max}}$

for $t = 1, ..., t_{\max}$ **do**

(1) Localise the next emitter with (3.18), yielding $\hat{\boldsymbol{m}}_t$.

(2) Update the supports as $\widehat{\boldsymbol{M}}^t = [\widehat{\boldsymbol{M}}^{t-1}, \hat{\boldsymbol{m}}_t]$, the corresponding intensities $\boldsymbol{\Gamma}$ as (3.20) and $\hat{\boldsymbol{\delta}}^t = \mathbf{0}$.

(3) Alternating gradient descent:

for $k = 1, ..., k_{\max}$ **do**

1) Refine $\{\widehat{\boldsymbol{M}}, \widehat{\boldsymbol{\Gamma}}\}$ using (3.21) while $l \leqslant l_{\max}$ (inner loop).

2) Prune support by (3.22).

3) Locally improve $\{\widehat{\boldsymbol{M}}, \widehat{\boldsymbol{\Gamma}}, \hat{\boldsymbol{\delta}}\}$ together using (3.25).

end for

end for

Note that Algorithm 3 is proposed for the $I \geqslant 2$ case, but also available for the particular $I = 1$ case mentioned in Section 3.3 by just viewing δ_j for $j = 1, \ldots, J$ as a single unknown parameter and iteratively update it similarly as (3.25).

3.6 Numerical Simulation

In this section, we perform numerical simulations to compare our proposed method, GSE-SC with existing methods including matched filtering (MF) [111], grid-based group CS (GCS) [100, 112] and ADCG [108] , as well as Cramér-Rao bound (CRB) [16, 17]. For these methods, we examine the influence of noises, the number of sensors J and the number of sensors with sensor errors J_r on the recovery of MD positions $\boldsymbol{m}$. Meanwhile, the estimation of the directional errors $\boldsymbol{\delta}$ and the MD number I using GSE-SC method is also presented. Finally, in the GSE-SC method, we consider the necessary conditions for precise localisation w.r.t. J and J_r.

3.6.1 Simulation Setting

We consider using J sensors to passively receive signals from $I = 2$ MDs. Each sensor is equipped with a ULA array and the number of array elements is $N_R = 12, j = 1, 2, \ldots, J$. The intervals of the adjacent elements are set as half the wavelength, i.e., $d_j = \lambda/2$. The positions of these sensors are set randomly and uniformly in a circle with center $(0, 0)$ m and radius 120 m. The pre-assumed angles, α'_j, are set as 0. Among these J sensors, J_r sensors have unknown location errors, denoted imprecise sensors, and the locations of the rest $J - J_r$ sensors are exactly known. It is assumed known that which sensor has errors and which is exactly located. In the simulations, the entries in $\boldsymbol{\delta}$ are set randomly in $(0, \pi/30]$. MD positions are set randomly and uniformly distributed in a circle with center $(0, 0)$ m and radius 40 m. The amplitude matrix $\boldsymbol{\Gamma}$ is set as a standard complex Gaussian random matrix. We assume that the additive noise $\boldsymbol{v}_j$ is i.i.d. white Gaussian with zero mean and variance σ_v^2.

We compare our approach GSE-SC with MF, GCS, ADCG methods and CRB. Particularly, MF method is to solve the following optimisation problem (3.26) w.r.t. $\boldsymbol{m} \in \mathcal{U}$, given by

$$\hat{\boldsymbol{m}} = \arg\max_{\boldsymbol{m}\in\mathcal{U}} \sum_{j=1}^{J} \frac{\|\boldsymbol{a}_j^H(\boldsymbol{m}, 0)\boldsymbol{r}_j\|_2^2}{N_R}. \tag{3.26}$$

Note that in this way, only one MD can be estimated using the MF method. GCS method solves the lasso problem (3.15). The position estimates of the I MDs correspond to the I largest values in the set $\left\{\|\boldsymbol{S}_{i,1:J}^{\mathrm{T}}\|_2, i = 1, 2, \ldots, g_x \times g_y\right\}$. For ADCG method, it is detailed in [108]. Based on the gradients in Appendix A, the CRBs are calculated as in [113] .

The parameter settings of these methods are shown as follows: The region of interest is set as $\mathcal{P} = [-50, 50] \times [-50, 50]$ m. The regularization parameter μ is set by borrowing the corresponding idea in [75]. In GCS method, we set the grid number as $g_x = g_y = 21$. In GSE-SC method, we set the total iteration times $t_{\max} = 4$, the outer loop iteration times $k_{\max} = 200$ and the inner loop iteration times $i_{\max} = 50$.

We use the indicator root mean square error (RMSE) in $\log_{10}$ scale to measure the recovery performance of $\{\boldsymbol{m}, \boldsymbol{\delta}\}$. Particularly, we carry out T_r Monte Carlo trials and denote RMSE of $\{\boldsymbol{m}, \boldsymbol{\delta}\}$ by

$$\begin{cases} \text{RMSE of } \boldsymbol{M} : \log\left(\sqrt{\frac{\sum_{t=1}^{T_r}\sum_{i=1}^{I}\|\hat{\boldsymbol{m}}_i^t - \tilde{\boldsymbol{m}}_i^t\|_2^2}{IT_r}}\right), \\ \text{RMSE of } \boldsymbol{\delta} : \log\left(\sqrt{\frac{\sum_{t=1}^{T_r}\|\hat{\boldsymbol{\delta}}^t - \tilde{\boldsymbol{\delta}}^t\|_2^2}{J_r T_r}}\right), \end{cases} \tag{3.27}$$

where $\tilde{\cdot}$ and $\hat{\cdot}$ represent the true values and estimation of $\cdot$.

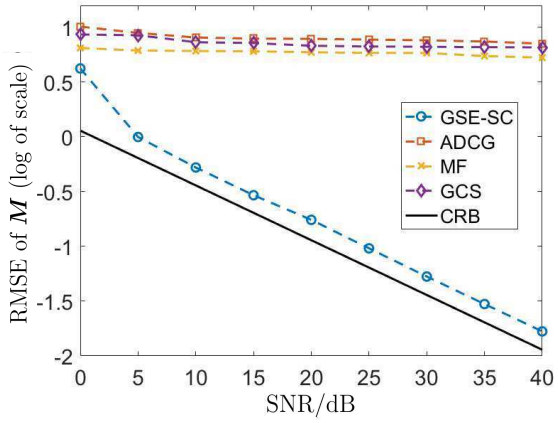
To further study the feasibility of GSE-SC method in certain scenarios such as $I = 1$, we also use hit rate as the evaluation index. Particularly, it is recognized as a successful hit if the RMSE of MD positions $\boldsymbol{m}$ is less than a specific threshold, denoted by T_p. Then hit rate is defined as the probability of the successful hit.

3.6.2 RMSE versus SNR and Computational Cost

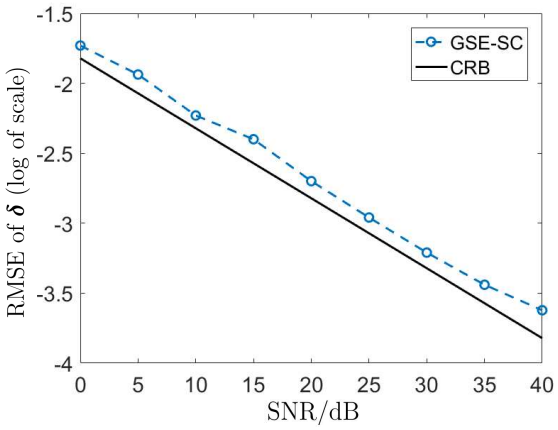
In this subsection, we compare the performance of MF, GCS, ADCG and GSE-SC methods with the CRB in terms of RMSE under different levels of noises.

Particularly, we set the total sensor number $J = 8$ and the imprecise sensor number $J_r = 4$. We change SNR from 0dB to 40dB and perform $T_r = 100$ Monte Carlo trials for each SNR and method. The CRBs of $\boldsymbol{M}$ and $\boldsymbol{\delta}$ are calculated w.r.t. SNR. To calculate RMSE as (3.27), we assume that $I = 2$ is pre-known. Fig 3.3(a) and Fig 3.3(b) show RMSEs of $\boldsymbol{M}$ by tested methods and RMSEs of $\boldsymbol{\delta}$ using GSE-SC method, respectively.

From Fig 3.3(a), we find that under the same SNR, RMSEs of $\boldsymbol{M}$ using GSE-SC method are significantly lower than the counterparts of MF, GCS and ADCG methods, indicating that GSE-SC outperforms the other methods in the estimation accuracy of MD positions. And the RMSE of $\boldsymbol{M}$ decreases much faster along with the increase of SNR by GSE-SC method than MF, GCS and ADCG methods. This happens because array directional errors play a more important role than noises here, and GSE-SC efficiently alleviates the



(a) RMSE of $\boldsymbol{M}$



(b) RMSE of $\boldsymbol{\delta}$

Figure 3.3: RMSEs of $\boldsymbol{M}$ and $\boldsymbol{\delta}$ versus SNR.

errors, while MF, GCS and ADCG methods do not. Meanwhile, the RMSE of $\boldsymbol{M}$ using GSE-SC method is close to CRB when SNR is large enough. Fig 3.3(b) demonstrates that RMSE of $\boldsymbol{\delta}$ using GSE-SC method decreases along with the increase of SNR and is also close to CRB.

3.6.3 Computational Cost

Estimator	Time [s]
GSE-SC	137.46
ADCG	82.18
MF (g_x=g_y=21)	0.121
GCS (g_x=g_y=21)	15.19
MF (g_x=g_y=64)	0.52
GCS (g_x=g_y=64)	400.05
MF (g_x=g_y=128)	1.91
GCS (g_x=g_y=128)	4543.5

Table 3.1: In case of SNR = 10 dB, the average running time of the considered estimators. CPU: Intel Core i7-7700HQ CPU 2.80GHz.

Following the simulation setup in section 3.6.2, the computational cost of estimator is measured by the average running time. The Both GSE-SC and ADCG methods are working on continuous parameters space, and both MF and GCS methods are two discrete based methods. As shown in Table 3.1, the GSE-SC has a longer running time than ADCG. This is due to the ADCG focuses on estimating less unknown parameters. As for discrete methods, the running time is increased by increasing the number of grid points g_x, g_y.

3.6.4 Advantages of Group Atomic Norm based Method

In order to demonstrate the superior performance of group atomic based method, i.e., GSE-SC comparing with the corresponding discrete method, i.e., GCS, we change the imprecise sensor number J_r into zero. Thus, the GCS also can be used for localising

MDs. In this case, the GSE-SC can be viewed as continuous counterpart of the GCS. In the discrete method, the ground truth is assumed on the grid points. However, the discrete method may lead to the leakage effect when the ground truth is off-the-grid, and deteriorate the estimation performance. On the other hand, a finer grid may mitigate the off-the-grid leakage but may result in large computational cost. The average running time and RMSE are shown in Table 3.2. Comparing with GCS, the GSE-SC can achieve a higher precision with less running time.

Estimator	Time [s]	RMSE [m]
GSE-SC	139.11	2.02
GCS (g_x=g_y=64)	398.68	33.89

Table 3.2: In case of SNR = 10 dB and the imprecise sensor number $J_r = 0$, the average running time and average RMSE of GSE-SC and GCS. CPU: Intel Core i7-7700HQ, 2.80GHz.

3.6.5 Convergence of GSE-SC

Following the simulation setup in section 3.6.2, we present results with a single trial simulation, to give some intuition on the convergence of GSE-SC. In the Fig 3.4, we show the trajectory of objective function of local improvement step along with the iterative times. The trajectory terminate close to the corresponding ground truth, which demonstrates that GSE-SC obtains accurate estimation.

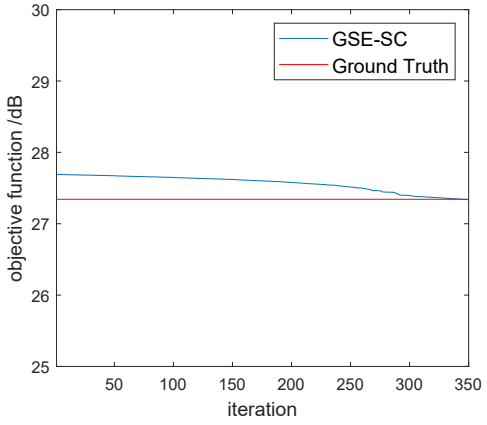


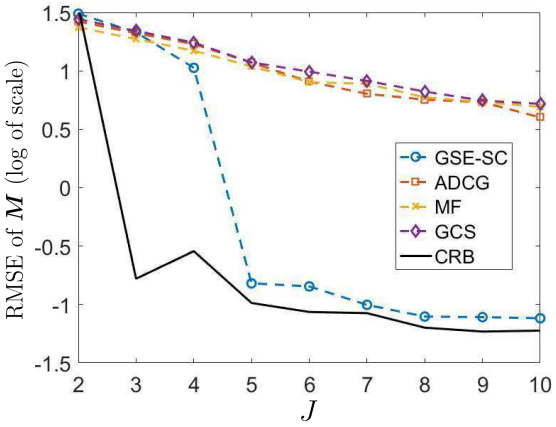
Figure 3.4: Convergence of the proposed GSE-SC

Theoretically, our proposed GSE-SC is globally convergence. The proposed GSE-SC combines convex and non-convex optimisation techniques. Firstly, the convex optimisation guarantees a global convergence. In the final step, the non-convex optimisation is a subroutine that takes estimations from the convex optimisation and attempts to use gradient information to reduce the value of objective function. Therefore, this non-convex optimisation does not change the convergence guarantees. The non-convex step provides a significantly sparser solutions by allowing the unknown parameters to move continuously within the parameter space.

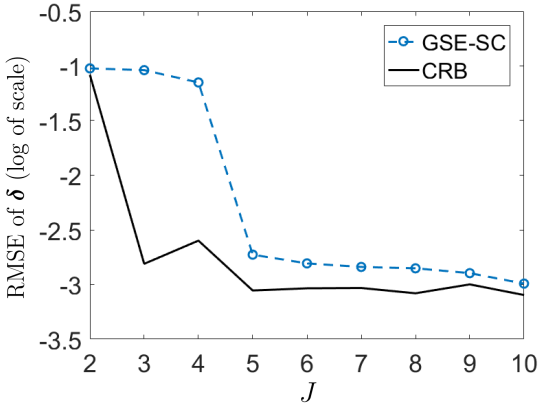
3.6.6 RMSE versus J

In this subsection, we investigate the impact of the number of sensors J on the estimation performance of $\{\boldsymbol{M}, \boldsymbol{\delta}\}$. Here, SNR is fixed as 25 dB, J is varied from 2 to 10 and J_r is set as $\lfloor J/2 \rfloor$. We then carry out $T_r = 100$ Monte Carlo trials in each case of J. The simulation results are given in Fig 3.5.

We observe from Fig 3.5 that RMSEs of $\boldsymbol{M}$ and $\boldsymbol{\delta}$ using GSE-SC method first decrease along with the increase of J and then tend to level off. Remarkably, there is a sharp drop from $J = 4$ to $J = 5$, implying a necessary condition for GSE-SC method in this case is $J \geqslant 5$. As for the simulated methods except the GSE-SC, these methods are failing to estimate $\boldsymbol{M}$ without calibrating the array directional errors. In Fig 3.5(a), we find that the RMSE of GSE-SC method is much smaller than the other tested methods and close to CRB when J is larger enough, $J \geqslant 5$. As J increases furthermore, the RMSE of GSE-SC and CRB both decline slowly which means the estimated performance cannot be improved significantly after the necessary condition. This phenomenon helps in the effective use of multiple arrays in practice. Fig 3.5(b) shows that the curve of the RMSE of $\boldsymbol{\delta}$ is similar with the RMSE of $\boldsymbol{m}$. This is because that more sensors are beneficial to the self-calibration on sensor errors, which further improves the localisation accuracy.



(a) RMSE of $\boldsymbol{M}$

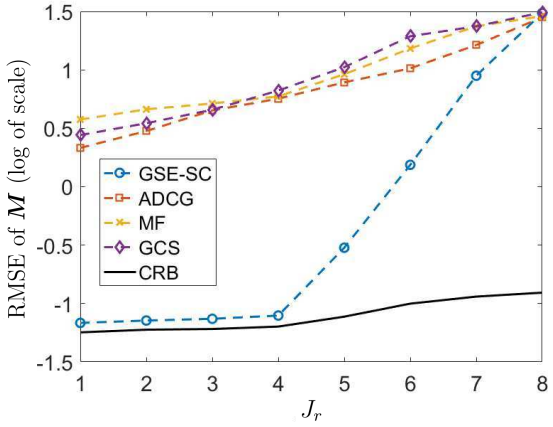


(b) RMSE of $\boldsymbol{\delta}$

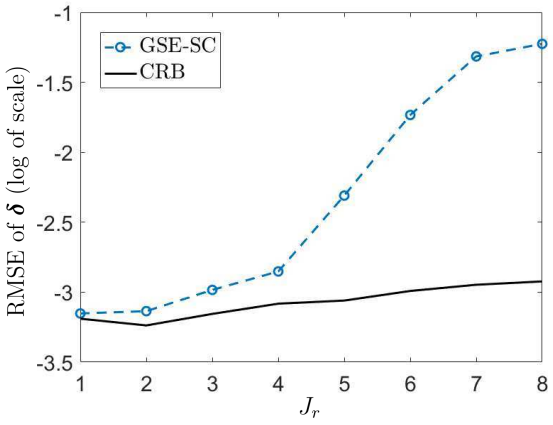
Figure 3.5: RMSEs of $\boldsymbol{M}$ and $\boldsymbol{\delta}$ versus J.

3.6.7 RMSE versus J_r

In this subsection, when the total sensor number J and SNR are fixed, we analyze the effect of the number of sensors with errors J_r on the localisation performance. This indeed reflects the performance bound of the multiple-array system in self-calibrating directional errors. Particularly, we set $J = 8$, SNR as 25dB and then vary J_r from 1 to 8. The simulation results are shown as Fig 3.6.



(a) RMSE of $\boldsymbol{M}$



(b) RMSE of $\boldsymbol{\delta}$

Figure 3.6: RMSEs of $\boldsymbol{M}$ and $\boldsymbol{\delta}$ versus J_r.

From Fig 3.6, we find that the RMSEs of $\boldsymbol{M}$ and $\boldsymbol{\delta}$ using GSE-SC method are close to CRB when $J_r \leqslant 4$ and increase apparently when $J_r \geqslant 5$. This reflects a performance bound of GSE-SC method in self-calibrating directional errors, and the limit is $J_r = 4$ when $J = 8$ in this scenario. However, in most cases, GSE-SC method still performs

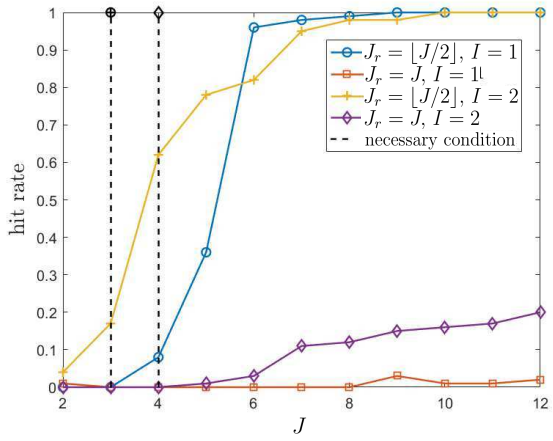
better than other tested methods.

3.6.8 Hit Rate versus J

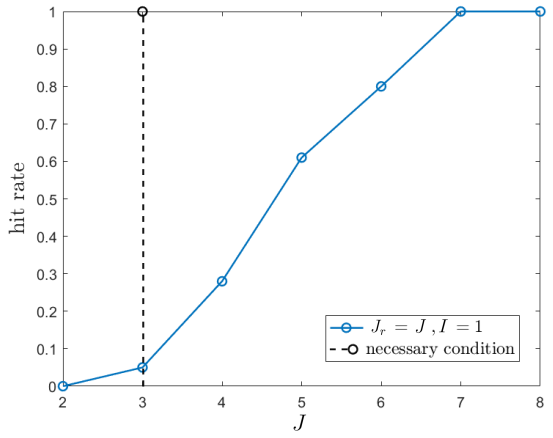
In this subsection, we study that by using GSE-SC method, how many sensors are required to achieve precise localisation in the cases with $I = 1, 2$, respectively. The corresponding results are compared with the necessary conditions.

Here we fix the SNR as 25dB. Based on the simulation results in Fig 3.3 and Fig 3.5, the threshold of hit rate is empirically set as $T_p = 0.25$. Particularly, we focus on the scenarios with $J_r = \lfloor J/2 \rfloor$ or $J_r = J$ and $I = 1, 2$. We perform $T_r = 100$ Monte Carlo trials in each case and the hit rate results, as well as the corresponding necessary conditions (distinguished by markers), are shown in Fig 3.7(a). Then, as assuming the same δ_j for $j = 1, \ldots, J$ for $I = 1$ case in Section 3.3, we also carry out $T_r = 100$ trials and the hit rate result is shown in Fig 3.7(b).

From Fig 3.7(a), we find that when $J_r = \lfloor J/2 \rfloor$, hit rates with $I = 1, 2$ tend to be 1 as J increases. And the least numbers of sensors for hit rate larger than 0.9 are 5 and 7 for $I = 1, 2$, respectively, which are larger than the necessary conditions. When $J_r = J$ and $I = 2$, hit rates grow slowly as J increases, since more unknowns are to be estimated compared with the $J_r = \lfloor J/2 \rfloor$ case. When $J_r = J$ and $I = 1$, the failure of the joint estimation of $\{\boldsymbol{m}, \boldsymbol{\delta}\}$ is unavoidable as discussed in Section 3.3, and hence the hit rate is close to 0. Therefore, we do not mark its necessary condition. For the $I = 1$ case with assuming the same δ_j, Fig 3.7(b) indicates that large enough $J = 7$ guarantees the hit rate close to 1.



(a) Different δ_j.



(b) The same δ_j.

Figure 3.7: Hit rate w.r.t. J with the same or different δ_j for $j = 1, \ldots, J$.

3.7 Conclusion

A novel super-resolved localisation method has been derived in presence of array directional errors. In order to achieve estimation and self-calibration simultaneously, we first formulate a direct localisation model consist of location of MDs and errors. Then the group sparsity is proposed to exploit the geometric consistency. Finally, we propose a method called GSE-SC to solve this optimisation problem. Simulation results demonstrate that GSE-SC method outperforms existing methods including MF, GCS and ADCG methods.

Chapter 4

MuG : A Multipath-Exploited and Grid-free Localisation in Multipath Environments

Localisation is challenging in presence of NLoS paths. Typical methods focus on reducing adverse impact due to the NLoS bias by separating LoS/NLoS path or NLoS mitigation. However, the complete reducing of the impact may not always succeed which degrades the localisation accuracy. Instead of dealing with NLoS bias, this Chapter presents a novel localisation method that exploits both LoS and NLoS paths in a much more general setting. Another unique feature is avoiding hard decisions on separating LoS/NLoS path and hence relevant possible error. A grid-free sparse inverse problem is formulated for localisation which avoids error propagation between multiple stages and handles multipath in a unified way. Extensive localisation experiments on different propagation environments and localisation systems are presented to illustrate the high performance of the proposed algorithm comparing with theoretical analysis. Especially, in one of the case studies, single antenna AP can locate a single antenna MD even when all paths are NLoS, which according to the authors knowledge is the first time in the literature. The rest of the Chapter is organized as follows. Section 4.1 introduces backgrounds of the problem.

Section 4.2 describes the signal model. We adopt a commonly assumed setup in the literature for the purpose of highlighting the idea without being drown into great technical details. Section 4.3 presents the proposed method to solve the localisation problem in multipath environments. Firstly, we show the feasible and geometric insight behind the proposed method. Then, a unified formulation is proposed for LoS path and NLoS path. Finally, we present a group sparse recovery based optimisation problem. Section 4.4 shows some numerical simulation results. The conclusion is drawn in Section 4.5.

4.1 Introduction

This chapter considers the general problem of localising a transmitting MD based on measurements at APs of which the locations are given. We follow the standard setup of this problem [1–4, 10–12, 14, 15, 27–31] by assuming a cooperative scenario where the MD and APs are synchronised, the transmitted signals from the MD are known at APs, and the signals from different MDs are orthogonal so that without loss of generality it suffices to consider only one MD. This chapter focuses on localisation in a 2D plane for compositional clarity though the extension to 3D space is straightforward.

A major challenge in localisation is the efficiency and accuracy of the estimation in presence of NLoS paths. Since without prior knowledge, the NLoS paths carry no information for localising MD's location. This claim is supported by the analysis in [16, 17], which showed that the CRB of the source location estimate depends only on the LoS paths. Different from the LoS path, NLoS paths occur when there is a scatter between MD and AP which results in different signal model and commonly modeled as LoS path with additive NLoS bias. In the literature, the NLoS bias is the majority of localisation error, and it is defined as a large and always positive error which results in additional propagation delay, attenuation, and angular bias. Typical environments including such as residential, office, and urban area have a high occurrence of NLoS situations. In such an environment, the use of the GPS might be impractical. Therefore, it is critical to understand the impact

of NLoS conditions on localisation systems and to develop techniques that exploit them for localisation.

The optimal strategy is to reduce effect of NLoS paths as much as possible, and exploit information carried by LoS paths to infer the location of MD. Many methods have been investigated to deal with NLoS paths which fall in two categories: *LoS/NLoS separation* and *NLoS removal*.

1. The *LoS/NLoS separation* focus on separating LoS/NLoS paths, and estimating the location of MD by using LoS paths. After separation, the NLoS paths are either discarded [10–14, 18] or used for restricting the feasible region [15, 19, 20].

2. The *NLoS mitigation* is typically adopted to reduce of the adverse effect NLoS paths, assuming that NLoS path has been separated or statistical information is available. In [36], the impact of NLoS path is removed with proper weighting or scaling.

In both two categories, the NLoS separation play a curial role and it can be achieved by by statistical approach or geometrical approach. 1) Under the statistical methods, the NLoS bias is considered as a random variable that can be constant [10], a random process of a Gaussian [11, 13, 14], or uniform distribution [18] with given parameters. In [12], the LoS/NLoS path can be separated by comparing the estimated variance of measurement with the prior historical information of LoS/NLoS path. To build an accurate statistical model, these methods require prior information about the distribution of NLoS paths or historical information. 2) In the geometrical method, the geometric relationship between the MD and AP is exploited to separate LoS paths based on measured ToF or DoA. In [47], the method exploits the fact that LoS paths typically arrive with a shorter ToF than NLoS paths. In [15, 47], a CS based framework is proposed for separating LoS path by exploiting the fact that LoS paths must originate from the same MD.

Discarding the NLoS paths may not be a viable option, as the number of available LoS paths may be limited. There are few works exploiting NLoS path for localisation in

NLoS and OLoS environment, but the methods are designed for some special settings. In [1–4,114], methods are designed for multi-antenna system where extra angle information is used for localisation. Typically, the MD is estimated in a two-step approach, first the intermediate parameters such as DoA, DoD [1–4, 114], then fusing them for localisation. It is also clear that the accuracy of localisation highly depends on the path parameter estimation. Error propagation is inherent since the accuracy of localisation highly depends on the intermediate parameter estimation. Another technical difficulty comes from data association between intermediate parameters such as ToF/DoA/DoD information and locations of scatters. As the same scatter creates different paths and gives rise to different intermediate parameters for different APs. When there are multiple scatters in the scene, the task of associating the intermediate variables estimates to scatters is an NP-hard problem for which the solution cannot be solved efficiently.

In this Chapter, we first study the geometry of localisation in multipath environment. We identify and analyse scenarios that localisation requires exploitation of NLoS paths and incorporating scatters' unknown locations. Motivated by the geometric insight, we then propose the usage of virtual scatters so that hard-decision, separation, and removal of NLoS paths are avoided. An optimisation formulation is then developed where a grid-less approach eliminates the tradeoff of discrete grids between localisation precision and computational complexity. In the numerical demonstrations, the performance improvement of our approach has been illustrated. In one of the numerical case studies, we show that localisation is possible even when all propagation paths are NLoS and there is no antenna array in the system to provide direction information, which according to the authors' knowledge is the first time in the literature.

To summarise, we propose an end-to-end optimisation framework which can be directly applied to different system configurations and various propagation environments. The technical novelties behind are virtual scatter and direct map, sparse recovery formulation and a grid-free formulation as introduced in Section 4.3.2, 4.3.3 and 4.3.4. As shown in Table 4.1, MuG is a universal scheme that can work for all three propagation envi-

ronments. Note that the algorithm proposed in [12, 36] also can be extended for other propagation environment. But the method proposed in [12] relies on additional laborious experimental campaigns to build up a database of NLoS error. The performance of [36] depends on how much a priori statistical information is available for localisation.

	LoS environment	OLoS environment	NLoS environment
ToF based techniques	MuG, [10–12,14,20, 27, 28, 36, 115]	MuG, [12]	MuG, [10–12,14,20, 27, 28, 36]
Hybrid ToF-DoA/DoD based techniques APs	MuG, [1–4, 15, 114, 116]	MuG, [1–5, 114]	MuG, [1–4, 15, 114]

Table 4.1: Without any prior information, MuG is the only technique can be used for all system setups and propagation conditions.

4.2 System Model

We focus on the following localisation problem by adopting a commonly assumed setup in the literature [1–4, 10–12, 14, 15, 27–31]. Consider a wireless communication system where J cooperative APs jointly estimate the location of one MD. We include two types of localisation systems where all APs can be equipped with single antenna and ULA with $N_R \geqslant 0$ antenna elements. It is typically assumed that via control signalling, the MD and the involved APs are synchronised [1, 28] and the transmitted waveform from the MD is known to APs. In the signal model, we only consider either LoS or NLoS with single-bounced signals from the MD to the APs. This is motivated by the fact that signals scattered twice or more times typically suffer from great propagation losses and are thus less perceptible [4, 5]. It is noteworthy that the above setup is a simplification of actual systems. For example, the assumption of synchronisation and the complete discard of multiple-bounced signals may be problematic in practice. Nevertheless, the above setup is widely adopted in the literature [1–4, 10–12, 14, 15, 27–31] for the purpose of highlighting the approach/idea without being drowned into great technical details.

The MD is assumed to transmit a signal $h(t) \in \mathbb{C}$ with a single antenna. The DoD information is assumed not available for localisation. In practice, the observation at the j-th AP is uniform N time-samples of the received signal, which can be expressed as

$$\boldsymbol{Y}_j = \boldsymbol{Y}_j^{\mathrm{LoS}}(\boldsymbol{m}) + \boldsymbol{Y}_j^{\mathrm{NLoS}}(\boldsymbol{m}, \boldsymbol{s}) + \boldsymbol{V}_j, \tag{4.1}$$

where $\boldsymbol{Y}_j \in \mathbb{C}^{N \times N_R}$ denotes the general form of received signal; The term $\boldsymbol{Y}_j^{\mathrm{LoS}}$ represents the received LoS path; $\boldsymbol{Y}_j^{\mathrm{NLoS}}$ represents the received NLoS path; For the LoS environment, the observation is $\boldsymbol{Y}_j = \boldsymbol{Y}_j^{\mathrm{LoS}}(\boldsymbol{m}) + \boldsymbol{V}_j$; For the OLoS environment, the observation is $\boldsymbol{Y}_j = \boldsymbol{Y}_j^{\mathrm{NLoS}}(\boldsymbol{m}, \boldsymbol{s}) + \boldsymbol{V}_j$; For the NLoS environment, the observation is $\boldsymbol{Y}_j = \boldsymbol{Y}_j^{\mathrm{LoS}}(\boldsymbol{m}) + \boldsymbol{Y}_j^{\mathrm{NLoS}}(\boldsymbol{m}, \boldsymbol{s}) + \boldsymbol{V}_j$; $\boldsymbol{m} = [m^x, m^y]^{\mathrm{T}}$ denotes the unknown location of the MD, and m^x and m^y are the horizontal and vertical coordinates, respectively; without loss of generality, the position of scatter is $\boldsymbol{s} = [s^x, s^y]^{\mathrm{T}}$ and $\boldsymbol{V}_j$ is the additive noise. More in details,

$$\boldsymbol{Y}_j^{\mathrm{LoS}}(\boldsymbol{m}) = \gamma_j \boldsymbol{h}\left(\tau_j^{\mathrm{LoS}}(\boldsymbol{m})\right) \boldsymbol{a}\left(\theta_j^{\mathrm{LoS}}(\boldsymbol{m})\right)^{\mathrm{T}}, \tag{4.2}$$

$$\boldsymbol{Y}_j^{\mathrm{NLoS}}(\boldsymbol{m}, \boldsymbol{s}) = \sum_{k=1}^{K-1} \gamma_{j,k} \boldsymbol{h}\left(\tau_{j,k}^{\mathrm{NLoS}}(\boldsymbol{m}, \boldsymbol{s})\right) \boldsymbol{a}\left(\theta_{j,k}^{\mathrm{NLoS}}(\boldsymbol{m}, \boldsymbol{s})\right)^{\mathrm{T}}, \tag{4.3}$$

where $\gamma \in \mathbb{C}$ is unknown coefficient modelling the signal attenuation along the path; K is the number of received paths at the AP; the position of the AP is $\boldsymbol{AP}_j = [{AP_j}^x, {AP_j}^y]^{\mathrm{T}}$; $\boldsymbol{h}(\tau_j)$ is the vector of N time-samples at the j-th AP with dealy τ_j :

$$\boldsymbol{h}(\tau_j) = \left[h(0 - \tau_j), h(\frac{1}{f_s} - \tau_j), \cdots, h(\frac{N-1}{f_s} - \tau_j)\right]^{\mathrm{T}}, \tag{4.4}$$

where f_s is the sampling frequency; $\boldsymbol{a}(\theta) \in \mathbb{C}^{N_R}$ reflects the phase differences of received signals due to DoA and it takes the form of

$$\boldsymbol{a}(\theta) = \left[1, e^{i\frac{2\pi}{\lambda}L\sin(\theta)}, \cdots, e^{i\frac{2\pi}{\lambda}L\sin(\theta)(N_R-1)}\right]^{\mathrm{T}}, \tag{4.5}$$

where λ is the wavelength of the carrier, and L is the distance between adjacent array

elements; $N_R \geqslant 0$ is the number of antenna elements at AP; $\boldsymbol{V}_j$ is the additive noise.

We would like to emphasise the difference between LoS path and NLoS from the computation of ToF and DoA. The LoS path is directly originating from MD $\boldsymbol{m}$. Therefore, the unknown parameters of ToF $\tau_j^{\mathrm{LoS}}(\boldsymbol{m})$ and DoA $\theta_j^{\mathrm{LoS}}(\boldsymbol{m})$ are only related to location of MD $\boldsymbol{m}$:

$$\tau^{\mathrm{LoS}}(\boldsymbol{m}) = \|\boldsymbol{m} - \boldsymbol{AP}\|_2/c, \tag{4.6}$$

$$\theta^{\mathrm{LoS}}(\boldsymbol{m}) = \mathrm{atan}(m^x - AP^x)/(m^y - AP^y). \tag{4.7}$$

Different from the LoS path, the NLoS path is originating from MD and scattering on scatter. The NLoS parameters $\tau_j^{NLoS}(\boldsymbol{m}, \boldsymbol{s})$ and $\theta_j^{NLoS}(\boldsymbol{m}, \boldsymbol{s})$ are specific to the location of MD $\boldsymbol{m}$ and scatter $\boldsymbol{s}$, which can be written in the following form:

$$\tau^{\mathrm{NLoS}}(\boldsymbol{m}, \boldsymbol{s}) = (\|\boldsymbol{m} - \boldsymbol{s}\|_2 + \|\boldsymbol{s} - \boldsymbol{AP}\|_2)/c, \tag{4.8}$$

$$\theta^{\mathrm{NLoS}}(\boldsymbol{m}, \boldsymbol{s}) = \mathrm{atan}(s^x - AP^x)/(s^y - AP^y). \tag{4.9}$$

Note that our aim is providing a unified method to tackle localisation problem for different environments and localisation systems. The propagation environments can be LoS environment, OLoS environment and NLoS environment. Different from the literature [10–15, 18–20, 36], we assume that the estimator does not know the information about the propagation environment and statistical distribution of NLoS path. Thus, it is very difficult to separate LoS/NLoS paths. Additionally, there is no guarantee that a specific path exists from the MD to a particular AP, e.g. the OLoS environment. The minimum delay path may correspond to a NLoS path. Therefore, most existing methods might be infeasible to be implemented here. As for the localisation system, we assume the number of antenna elements $N_R \geqslant 0$ is the number of ; $N_R = 1$ represents the single

antenna AP that only ToF information is available; $N_R > 1$ represents multi-antenna AP that both ToF and DoA are available for localisation.

4.3 Proposed MuG method

4.3.1 Feasibility and Geometry

Without prior knowledge, it was shown in [16, 17] that the CRB of the MD location estimate depends only on the LoS paths, and NLoS paths cannot used for localisation. Under this condition, typical methods suggest that the NLoS paths are separated and discarded, and the MD location is estimated by using only the LoS paths. However, there is no guarantee that the mistaken sapration can be avoided completely and the LoS paths always exist from the MD to APs, which degrades the performance of these localisation algorithms. Towards localisation with both LoS paths and NLoS paths, the underlying geometry becomes much more complicated and much less discussed. In the following, we study the geometry behind localisation with multipath to demonstrate that the proposed MuG method can exploit both LoS paths and NLoS paths for localisation. In order to show it is a unified method, we start from the most generic single antenna APs system where only ToF information is available. Then we demonstrate that the same localisation principle can be applied to other system configuration such as: APs equipped with antenna arrays where both ToF and DoA information can be extracted.

In this subsection, we assume that LoS path and NLoS path is separated and data association between intermediate parameters and scatters is solved. In the literature, they are two crucial problems to affect performance of traditional localisation methods. The assumptions are made for following reasons: 1) We focus on identifying and analysing geometry of scenarios that incorporate both LoS and NLoS path for localisation. 2) Motivated by the geometric insight, we propose the usage of virtual scatters so that hard-decision on LoS/NLoS path separation is avoided completely. Moreover, the direct

map from location of MD and scatters is proposed so that the notorious data association problem is avoided.

4.3.1.1 Single Antenna APs

Typical localisation methods in the literature rely on trilateration algorithm which is the most basic and intuitive method to determine the location of the MD. The basic principle of this algorithm is to estimate the location of the MD (in 2D plane) by requiring at least three APs with known locations and their distances from the MD to be localised. Given the distance from MD to the AP, it is known that the feasible location of MD (FLMD) must be along the circumference of a circle centered at AP and a radius equal to the distance from MD to AP. The intersection of these three circles is the location of the unknown MD. The principle of this method is illustrated in Fig 1.3.

Trilateration localisation only focuses on estimating location of MD and it is effective when only LoS paths are presented. However, in case of multipath, there is much more uncertainty in localisation. For LoS path, it is directly originating from MD, which has the shortest propagation distance between MD and AP. For a given AP, the FLMD of LoS path is along the circumference of a circle, as illustrated in Fig 4.1(a). The corresponding radius equals the measured ToF times speed of light. Different from LoS path, the NLoS path is originating from MD and scattering on scatter. Therefore, for a given AP, the FLMD of NLoS path is restricted to a disk region, as depicted in Fig 4.1 (b). The corresponding radius is summation of the distance between MD and scatter and the distance between scatter and AP.

With increase in the number of multipath, the FLMD could be further reduced. In the case of the localisation system with multiple APs, we assume intermedia parameters and associated scatters are known. This assumption is reasonable, since the locations of MD and scatters are directly estimated by proposed sparse inverse problem which avoids estimating of intermedia parameters and data association problem. For a pair of APs, the

FLMD obtained by two LoS paths contains two estimations which are located at intersections of two circles, as depicted in Fig 4.2(a). In the case of two NLoS paths, the FLMD is intersection of two disk region and it still contains infinite number of estimations, as depicted in Fig 4.2(b). For a group of three APs, FLMD obtained by three LoS paths contains one estimation which is uniquely located at intersection of three circles, as depicted in Fig 4.3(a). In presence of three NLoS paths, directly applying trilateration principle results in erroneous estimation since the biased propagation distance, as illustrated in Fig 4.3(b).

Theorem 4.1. *In $D = 2, 3$ dimensional space, given D LoS paths originating from the same MD, the estimated location of MD obtained by trilateration localisation method is not unique.*

Proof. In D dimensional space, let γ^{L} denote FLMD obtained by LoS path, d is the propagation distance from MD $\boldsymbol{m} \in \mathbb{R}^D$ to AP $\boldsymbol{AP} \in \mathbb{R}^D$. The corresponding FLMD γ^{L} is formulated as

$$\gamma^{\mathrm{L}} = \left\{ \boldsymbol{m} \in \mathbb{R}^D | \|\boldsymbol{m} - \boldsymbol{AP}\|_2 = d \right\}. \tag{4.10}$$

The FLMD obtained by LoS path γ^{L} is along the circumference of a circle centered at AP $\boldsymbol{AP}$, and the radius d is equal to propagation distance of LoS path.

As for LoS paths, consider a set $\mathcal{C}^{\mathrm{LoS}} = \left\{ \gamma_1^{\mathrm{L}}, \gamma_2^{\mathrm{L}}, \cdots, \gamma_J^{\mathrm{L}} \right\}$ of J circles, whose centers $\{\boldsymbol{AP}\}_{j=1}^{J}$ and radius $\{d\}_{j=1}^{J}$ are known. The intersection of J circles in $\mathcal{C}^{\mathrm{LoS}}$ is denoted by

$$\mathcal{I}^{\mathrm{LoS}(J)} = \cap \gamma_j^{\mathrm{L}}, \text{ for } j = 1, ..., J. \tag{4.11}$$

More precisely, it can be expressed as

$$\mathcal{I}^{\mathrm{LoS}(J)} = \{\boldsymbol{m} \in \mathbb{R}^D | \|\boldsymbol{m} - \boldsymbol{AP}_j\|_2 = d_j, \\ \text{for } j = 1, ..., J.\}. \tag{4.12}$$

In case of $D = 2$ space, the number of LoS paths of localisation is then $J = 2$. The intersection of $J = 2$ circles $\mathcal{I}^{\mathrm{LoS}(2)}$ is determined by solving following equations

$$\|\boldsymbol{m} - \boldsymbol{AP}_1\|_2 = d_1 \tag{4.13}$$

$$\|\boldsymbol{m} - \boldsymbol{AP}_2\|_2 = d_2 \tag{4.14}$$

Without loss of generality, we can rotate and translate any two APs as $\boldsymbol{AP}_1 = [0, 0]^{\mathrm{T}}$, $\boldsymbol{AP}_2 = [AP_2^x, 0]^{\mathrm{T}}$. By squaring (4.13) from (4.14) and then eliminating m^y, we have

$$(m^x - AP_2^x)^2 + (d_1)^2 - (m^x)^2 = (d_2)^2. \tag{4.15}$$

Expand and simplify (4.15)

$$-2m^x(AP_2^x)^2 + (AP_2^x)^2 + (d_1)^2 = (d_2)^2. \tag{4.16}$$

Then m^x of $\boldsymbol{m}$ is

$$m^x = \frac{(d_1)^2 - (d_2)^2 + (AP_2^x)^2}{2AP_2^x}. \tag{4.17}$$

Substituting (4.17) into (4.13), the m^y of $\boldsymbol{m}$ is

$$m^y = \pm\sqrt{(d_1)^2 - (m^x)^2}. \tag{4.18}$$

Based on (4.12) with $D = 2$, we have $(m^x)^2 \leq (d_1)^2$. Therefore, (4.18) indicates the trilateration solution determined by $D = 2$ LoS paths is not uniquely. Especially, in case

of $m^x = d_1$, mathematically speaking, it is referred as a special case that the two possible solutions are overlapped.

In case of $D = 3$ space, the number of LoS paths of localisation is then $N_L = 3$. The intersection of $N_L = 3$ circles $\mathcal{I}^{\text{LoS}(3)}$ is determined by solving following equations

$$\|\boldsymbol{m} - \boldsymbol{AP}_1\|_2 = d_1, \tag{4.19}$$

$$\|\boldsymbol{m} - \boldsymbol{AP}_2\|_2 = d_2, \tag{4.20}$$

$$\|\boldsymbol{m} - \boldsymbol{AP}_3\|_2 = d_3. \tag{4.21}$$

Without loss of generality, we can rotate and translate any three APs as $\boldsymbol{AP}_1 = [0, 0, 0]^{\text{T}}$, $\boldsymbol{AP}_2 = [AP_2^x, 0, 0]^{\text{T}}$, $\boldsymbol{AP}_3 = [AP_3^x, AP_3^y, AP_3^y]^{\text{T}}$. By subtracting (4.19) and (4.20) to solve for m^x, we have

$$m^x = \frac{(d_1)^2 - (d_2)^2 + (AP_2^x)^2}{2AP_2^x}. \tag{4.22}$$

Then we subtract (4.19) and (4.21) to solve for m^y:

$$m^y = \frac{(d_1)^2 - (d_3)^2 + (AP_3^x)^2 + (AP_3^y)^2 - 2AP_3^x m^x}{2AP_3^y} \tag{4.23}$$

Based on the expression of m^x and m^y, we can subtract it into (4.19) to solve for m^z:

$$m^z = \pm\sqrt{(d_1)^2 - (m^x)^2 - (m^y)^2}, \tag{4.24}$$

Based on the (4.12) with $D = 3$, we have $(m^x)^2 + (m^y)^2 \leq (d_1)^2$. Therefore, (4.24) indicates the trilateration solution determined by $D = 3$ LoS paths is not unique. Especially, in case of $(m^x)^2 + (m^y)^2 = (d_1)^2$, mathematically speaking, it is referred as a special case that the two possible solutions are overlapped. ■

Theorem 4.2. *In $D = 2, 3$ dimensional space, if all paths are NLoS, the estimated location of MD obtained by the trilateration localisation method is not unique.*

Proof. In D dimensional space, let γ^{N} denote FLMD obtained by NLoS path, d is the propagation distance from MD $\boldsymbol{m} \in \mathbb{R}^D$ to AP $\boldsymbol{AP} \in \mathbb{R}^D$. The corresponding FLMD γ^{N} is formulated as

$$\gamma^{\mathrm{N}} = \left\{ \boldsymbol{m} \in \mathbb{R}^D | \|\boldsymbol{m} - \boldsymbol{AP}\|_2 + b = d \right\}, \tag{4.25}$$

where $0 < b < d$ is the NLoS bias to model the extra propagation distance due to scattering. Therefore, as illustrated in Fig 4.1 (b), the FLMD γ^{N} obtained by NLoS path is a disk region centered at AP $\boldsymbol{AP}$, and the radius d is equal to the propagation distance of NLoS path.

As for NLoS paths, consider a set $\mathcal{C}^{\mathrm{NLoS}} = \left\{ \ldots, \gamma^{\mathrm{N}}_{j,g}, \cdots, \gamma^{\mathrm{N}}_{J,G} \right\}$ of JG circles, whose centers $\{\boldsymbol{AP}_j\}_{j=1}^{J}$ and radius $\{d_{j,g}\}_{j=1,g=1}^{J,G}$ are known; J and G are number of APs and scatters, respectively. The intersection of JG circles in $\mathcal{C}^{\mathrm{NLoS}}$ is denoted by

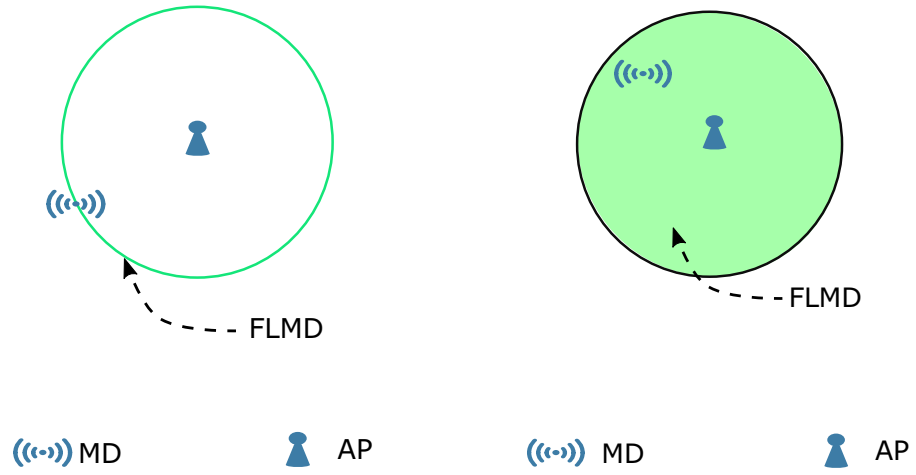
$$\mathcal{I}^{\mathrm{NLoS}(JG)} = \cap \gamma^{\mathrm{N}}_{j,g}, \text{for } j = 1, ..., J, g = 1, ..., G. \tag{4.26}$$

More precisely, the intersection of JG circles can be expressed as

$$\begin{aligned} \mathcal{I}^{\mathrm{NLoS}(JG)} = \{ \boldsymbol{m} \in \mathbb{R}^D | \|\boldsymbol{m} - \boldsymbol{AP}_j\|_2 + b_{j,g} = d_{j,g}, \\ \text{for } j = 1, ..., J, g = 1, ..., G \}. \end{aligned} \tag{4.27}$$

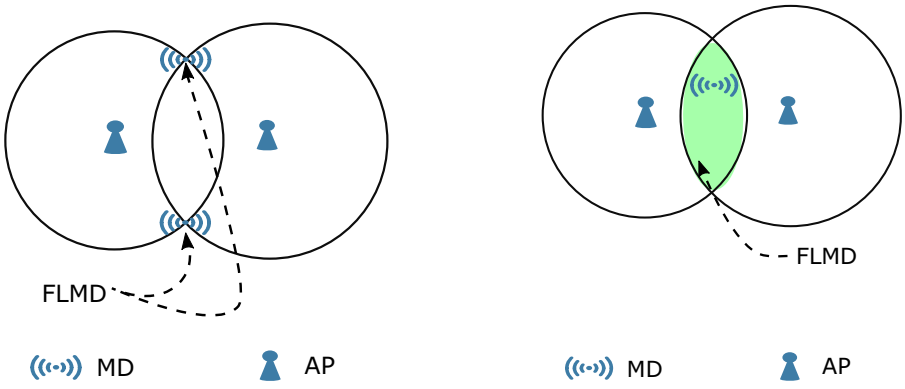
Note that the number of unknowns $D+JG$ are always larger than the number of equations JG for solving. The (4.27) is an underdetermined problem whose solution cannot be uniquely determined. ■

Theorem 4.2 indicates that increasing the number of NLoS paths, the FLMD obtained by trilateration localisation contains infinite number of estimations. In NLoS environment, ToF-based range estimates are positively biased with high probability, since the first arriving NLoS path travels a distance that is in excess of the true distance between MD and



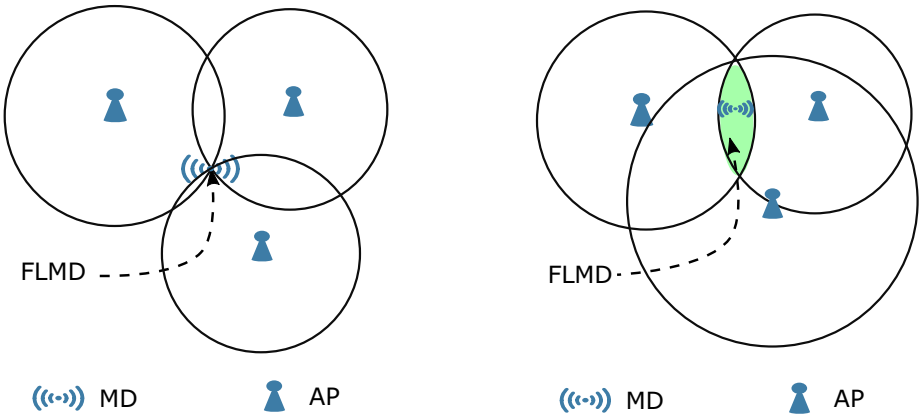
(a) (b)

Figure 4.1: (a) For a given AP, FLMD obtained by LoS path is along the circumference of a circle. (b) For a given AP, FLMD obtained by NLoS path is restricted to a disk region.



(a) (b)

Figure 4.2: (a) For a given pair of APs, FLMD obtained by two LoS path contains two intersections of two circles. (b) For a given pair of APs, FLMD obtained by NLoS path is intersection of two disk region.



(a) (b)

Figure 4.3: (a) For given group of three APs, FLMD obtained by LoS paths contains one integration of three circles. (b) For given group of three APs, FLMD obtained by NLoS paths is intersection of three disk region.

AP. These effects result in FLMD that involves much more uncertainty estimation compared with LoS paths. The accuracy of estimated MD can be adversely affected. Recall the estimator (1.45) derived for NLoS environment is a constrained nonlinear underdetermined problem. There might be an infinite number of solutions. Therefore, localisation methods in the literature suggest to use LoS paths for localisation. The technical difficulty comes from that the multipath signal is mixture of LoS paths and NLoS paths. Separating LoS path from multipath requires prior information and mistaken separation can not be avoided completely.

Theorem 4.3. *Except for the non-general setting that $D+1$ APs are placed in alignment. In $D = 2, 3$ dimensional space, the location of scatter can be uniquely estimated, given $D + 1$ NLoS paths scattered from the same scatter.*

Proof. In D dimensional space, let γ^{s} denote feasible location of scatter (FLS) obtained by NLoS paths, d is the propagation distance from MD $\boldsymbol{m} \in \mathbb{R}^D$ to AP $\boldsymbol{AP} \in \mathbb{R}^D$. The corresponding FLS γ^s is formulated as

$$\gamma^s_{j,i} = \left\{ \boldsymbol{s} \in \mathbb{R}^D | \|\boldsymbol{s} - \boldsymbol{AP}_j\|_2 - \|\boldsymbol{s} - \boldsymbol{AP}_i\|_2 = d_j - d_i, j \neq i \right\}. \tag{4.28}$$

Based on time different of flight (TDoF), the FLMD obtained by two NLoS paths γ^s is a half-hyperbola with two focus at $\boldsymbol{AP}_j$ and $\boldsymbol{AP}_i$, respectively.

As for NLoS paths, consider a set $\mathcal{C}^{\mathrm{s}} = \left\{ \gamma^{\mathrm{s}}_{1,2}, \gamma^{\mathrm{s}}_{1,3}, \cdots, \gamma^{\mathrm{s}}_{J-1,J} \right\}$ of $\frac{J!}{2(J-2)!}$ half-hyperbolas whose focuses $\{\boldsymbol{AP}_j\}_{j=1}^J$ are known. The intersection of $N_H = \frac{J!}{2(J-2)!}$ half-hyperbolas in $\mathcal{C}^{\mathrm{s}}$ is denoted by

$$\mathcal{I}^{\mathrm{S}(N_H)} = \cap \gamma^{\mathrm{s}}_{i,j}, \text{ for } j \neq i, \text{and } j, i = 1, ..., J. \tag{4.29}$$

More precisely, it can be expressed as

$$\mathcal{I}^{\mathrm{S}(N_H)} = \{\boldsymbol{s} \in \mathbb{R}^D | \|\boldsymbol{s} - \boldsymbol{AP}_j\|_2 - \|\boldsymbol{s} - \boldsymbol{AP}_i\|_2 = d_j - d_i, \\ \text{for } j \neq i, \text{and } j, i = 1, ..., J.\}. \tag{4.30}$$

In case of $D = 2$ dimensional space, the number of NLoS paths of localisation is then $N_H = 3$. The intersection of hyperbolas $\mathcal{I}^{\mathrm{S}(3)}$ is determined by solving following equations

$$\|\boldsymbol{s} - \boldsymbol{AP}_1\|_2 - \|\boldsymbol{s} - \boldsymbol{AP}_2\|_2 = d_1 - d_2, \tag{4.31}$$

$$\|\boldsymbol{s} - \boldsymbol{AP}_1\|_2 - \|\boldsymbol{s} - \boldsymbol{AP}_3\|_2 = d_1 - d_3, \tag{4.32}$$

without loss of generality, we can rotate and translate any three APs as $\boldsymbol{AP}_1 = [0, 0]^{\mathrm{T}}$, $\boldsymbol{AP}_2 = [AP_2^x, 0]^{\mathrm{T}}$, $\boldsymbol{AP}_3 = [AP_3^x, AP_3^y]^{\mathrm{T}}$. The presented solution is essentially transforming the hyperbolic equation into linear form associated with $\boldsymbol{s} = [s^x, s^y]^{\mathrm{T}}$. We rewrite (4.31) as a difference of squares:

$$\|\boldsymbol{s} - \boldsymbol{AP}_1\|_2^2 - \|\boldsymbol{s} - \boldsymbol{AP}_2\|_2^2 = (d_1)^2 - (d_2)^2. \tag{4.33}$$

Let $\boldsymbol{AP}_1 = [0, 0]^{\mathrm{T}}$, $\boldsymbol{AP}_2 = [AP_2^x, 0]^{\mathrm{T}}$. We have

$$s^x = \frac{(d_1)^2 - (d_2)^2 + (AP_2^x)^2}{2AP_2^x}. \tag{4.34}$$

By similar analysis of the propagation difference between d_1 and d_3, (4.32) can be written as:

$$s^y = \frac{(d_1)^2 - (d_3)^2 + (AP_3^x)^2 + (AP_3^y)^2 - 2AP_3^x s^x}{2AP_3^y}. \tag{4.35}$$

When $AP_3^y = 0$, it indicates the non-general setting that 3 APs are placed in an aligned line. From mathematical expression of s^x (4.34) and s^y (4.35), in $D = 2$ dimensional space, the unique solution of scatter requires at least $D = 3$ NLoS paths.

In case of $D = 3$ dimensional space, the number of NLoS paths of localisation is then $N_H = 4$. The intersection of hyperbolas $\mathcal{I}^{\mathrm{S}(4)}$ is determined by solving following equations

$$\|\boldsymbol{s} - \boldsymbol{AP}_1\|_2 - \|\boldsymbol{s} - \boldsymbol{AP}_2\|_2 = d_1 - d_2, \tag{4.36}$$

$$\|\boldsymbol{s} - \boldsymbol{AP}_1\|_2 - \|\boldsymbol{s} - \boldsymbol{AP}_3\|_2 = d_1 - d_3, \tag{4.37}$$

$$\|\boldsymbol{s} - \boldsymbol{AP}_1\|_2 - \|\boldsymbol{s} - \boldsymbol{AP}_4\|_2 = d_1 - d_4, \tag{4.38}$$

Without loss of generality, we can rotate and translate any four APs as $\boldsymbol{AP}_1 = [0, 0, 0]^{\mathrm{T}}$, $\boldsymbol{AP}_2 = [AP_2^x, 0, 0]^{\mathrm{T}}$, $\boldsymbol{AP}_3 = [AP_3^x, AP_3^y, AP_3^z]^{\mathrm{T}}$, $\boldsymbol{AP}_4 = [AP_4^x, AP_4^y, AP_4^z]^{\mathrm{T}}$. By similar analysis as in $D = 2$ dimensional space, we can rewrite (4.36) as a difference of squares:

$$\|\boldsymbol{s} - \boldsymbol{AP}_1\|_2^2 - \|\boldsymbol{s} - \boldsymbol{AP}_2\|_2^2 = (d_1)^2 - (d_2)^2 \tag{4.39}$$

Let $\boldsymbol{AP}_1 = [0, 0, 0]^{\mathrm{T}}$, $\boldsymbol{AP}_2 = [AP_2^x, 0, 0]^{\mathrm{T}}$. We have

$$s^x = \frac{(d_1)^2 - (d_2)^2 + (AP_2^x)^2}{2AP_2^x}. \tag{4.40}$$

By similar analysis of the propagation distance between d_1 and d_3, (4.37) can be written as:

$$2AP_3^y s^y + 2AP_3^z s^z = E_1, \tag{4.41}$$

where

$$E_1 = (d_1)^2 - (d_3)^2 + (AP_3^x)^2 + (AP_3^y)^2 + (AP_3^z)^2 - 2AP_3^x s^x. \tag{4.42}$$

By similar analysis of the propagation distance between d_1 and d_4, (4.38) can be written as:

$$2AP_4^y s^y + 2AP_4^z s^z = E_2, \tag{4.43}$$

where

$$E_2 = (d_1)^2 - (d_4)^2 + (AP_4^x)^2 + (AP_4^y)^2 + (AP_4^z)^2 - 2AP_4^x s^x. \tag{4.44}$$

By substituting (4.41) and (4.42) into (4.43) and (4.44), we have

$$s^y = \frac{E_1 - 2AP_3^z s^z}{2AP_3^y}, \tag{4.45}$$

$$s^z = \frac{AP_3^y E_2 - AP_4^y E_1}{2AP_4^y AP_3^z - 2AP_3^y AP_4^z}. \tag{4.46}$$

When $AP_3^y = 0$, $AP_4^y = 0$, it indicates one of non-general settings that 4 APs are placed in aligned plane. From mathematical expression of s^x (4.40), s^y (4.45) and s^z (4.46), in $D = 3$ dimensional space, the unique solution of scatter requires at least $D = 4$ NLoS paths.

■

Typical methods in the literature only consider location of MD, although the locations of scatters can help for localising the MD. The proposed MuG method concerns locations of MD and scatters, which is able to exploit the information carried by LoS paths and NLoS paths. Given a group of three APs, one may uniquely determine the location of scatter by time different of flight (TDoF) based technique. Remark that the estimated scatter is overlapped with MD in case of LoS paths. More specifically, for single TDoF measurement between the j-th AP $\boldsymbol{AP}_j$ and the l-th AP $\boldsymbol{AP}_l$, the feasible locations of scatter correspond to a hyperbola. As a result, it can be shown in Fig 4.4 that the scatter is located at intersection of at least three hyperbolas. The FLMD further generates a circle which is centred at the estimated scatter. The corresponding radius is the propagation distance between MD and scatter which can be computed based on estimated scatter and ToF of NLoS path. In Fig 4.4, the demonstration of FLMD is illustrated as green solid line. Once more than three scatters have been determined. Then following the trilateration method, the location of MD is located at the intersection of these FLMDs. See Fig 4.5

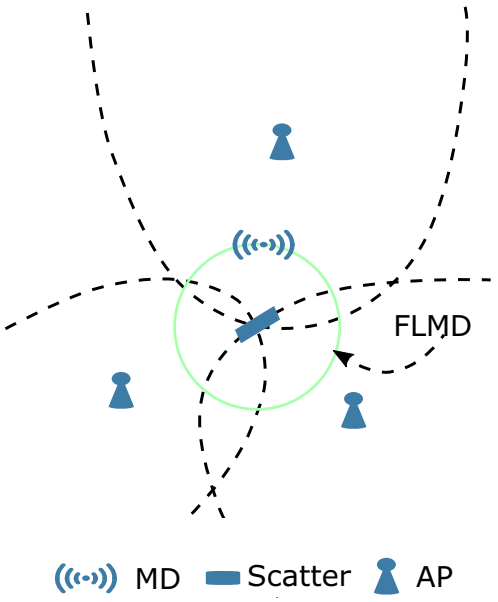


Figure 4.4: TDoF based localisation method for localising scatter. The FLMD further generates a circle in green which is centred at the estimated scatter.

for an illustration. When the multipaths are the LoS paths, the radius of FLMD is zero which indicates the location of MD overlaps with estimated scatter. Therefore, in OLoS environment, the unique FLMD is obtained by acquiring at least three APs and three scatters. In the case of LoS environment, the location of MD can be determined by three APs which coincides with traditional trilateration localisation method presented in Fig 4.3(a).

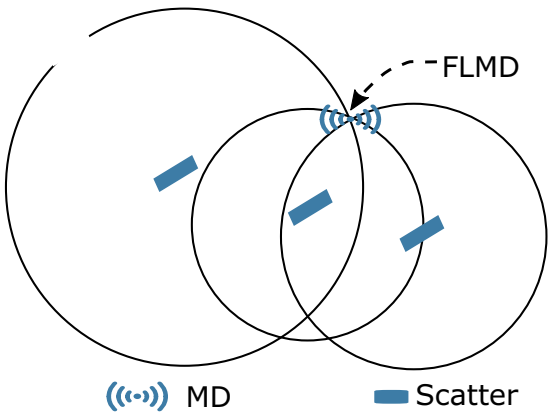


Figure 4.5: In case of NLoS paths, the trilateration localisation method can be applied to find location of MD when there are three estimated scatters.

4.3.1.2 APs with Antenna Arrays

The proposed localisation principle concerns locations of MD and scatters, and it can be extended for different system configuration. We take the multi-antenna APs as an example, in which both ToF and DoA measurements are available. As for a LoS path, the DoA of LoS path indicates the propagation direction from MD. Thus, for a given AP, the FLMD obtained by LoS path contains unique estimated MD based on known ToF and DoA (see Fig 4.6(a) for illustration). Unlike the LoS path, the DoA of NLoS path indicates the direction from scatter. For a given AP, the location of scatter can be anywhere along the angle direction line (see Fig 4.6(b) plotted with 'red solid line'). However, the FLMD of NLoS path is involved much more uncertainty, since it depends on location of scatter. Two possible scatters obtained by NLoS paht are depicted in Fig 4.6(b), the corresponding FLMD is along the circumference of circles centered at scatters which lead to infinite feasible locations of MD.

Unlike the LoS path, the NLoS path cannot be directly used for localising MD. For a given pair of APs, the FLMD of LoS paths is uniquely obtained by finding intersection of angle direction lines from different APs (see Fig 4.7(a) for illustration). However, this idea can not be used for finding FLMD of NLoS paths. Different from FLMD obtained from two LoS paths, the FLMD obtained by two NLoS path has much more uncertainty. Because the DoA indicates the direction from scatter which is irrelevant with location of MD. As shown in Fig 4.7(b), the corresponding FLMD is along the circumference of a circle centered at the scatter.

In order to incorporate NLoS paths for localisation, the localisation principle is also relying on locations of scatters. Given a pair of two APs and received NLoS paths, one may uniquely determined is location of scatter by finding the intersection of DoA lines. Since the NLoS paths are scattering from the same scatter. As shown in Fig 4.7(b), the corresponding FLMD is along the circumference of circle centered at scatter. Recalling the idea presented in Fig 4.5, the process of estimating location of MD is the same as single

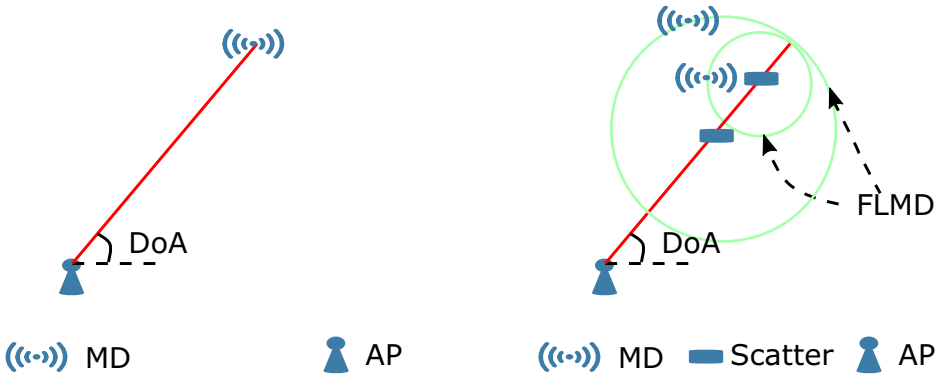


Figure 4.6: (a) As for LoS path, the FLMD is uniquely determined based on ToF and DoA information. (b) As for NLoS path, the FLMD contains infinite feasible solutions which is along the circumference of circles centered at scatters.

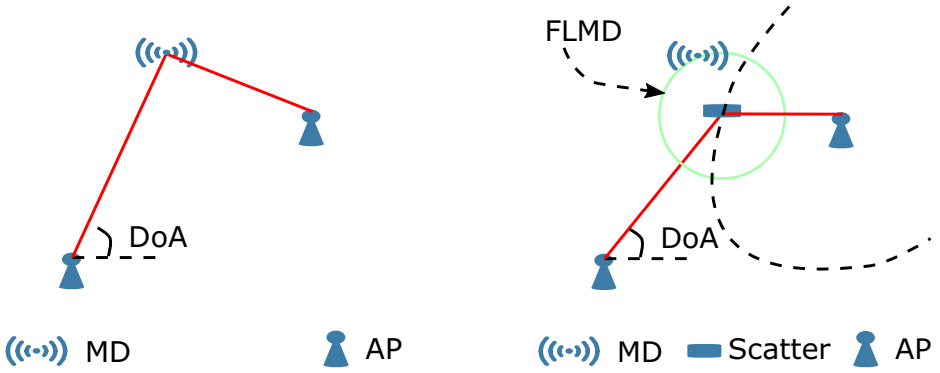


Figure 4.7: (a) As for LoS paths, the FLMD is located at intersection of two angle direction lines from different APs. (b) As for NLoS paths, the FLMD is along the circumference of circle centered at the scatter.

antenna case. The unique FLMD is possible by applying the trilateration localisation method when the localisation system contains more than three scatters. Therefore, in OLoS environment, the unique FLMD is obtained by acquiring at least two APs and three scatters. In the case of LoS environment, the location of MD can be determined by at least two APs. Note that above processes are implicitly embedded into our end-to-end framework.

4.3.2 Virtual Scatter and Direct Map

While much efforts in the literature focus on separating and removing NLoS paths [10–15], our approach avoids hard decision on separating propagation path and accommodates

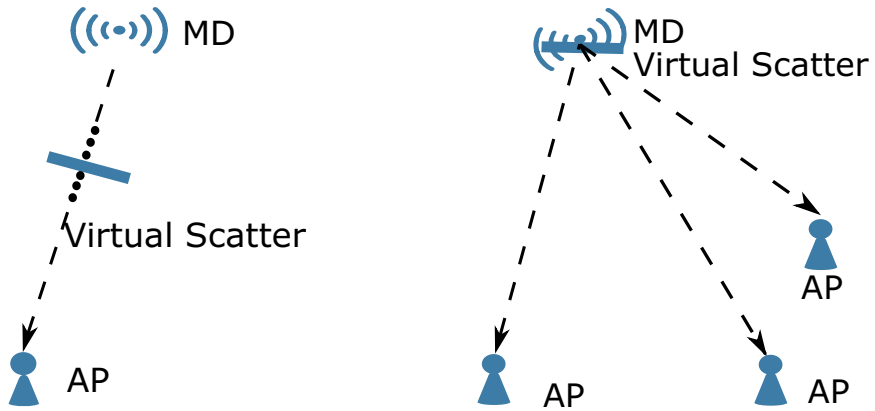


Figure 4.8: (a) When there is only one LoS path existing in the environment, the location of the virtual scatter can be anywhere on the straight line between AP and MD. This arbitrariness will not affect the localisation performance, as the ultimate goal is to locate the MD. (b) When there are multiple LoS paths existing in the environment, a sparsity constraint in (4.61) will lead to a solution that all virtual scatters merge to one being located at the MD position. Technical novelties behind are the virtual scatter, direct map and a grid-free formulation as introduced in Section 4.3.2, 4.3.3 and 4.3.4.

both LoS and NLoS paths in a unified way. We simplify the two-component model $\boldsymbol{Y}_j = \boldsymbol{Y}_j^{\text{LoS}} + \boldsymbol{Y}_j^{\text{NLoS}}$ (see Section 4.2 for details) into one-component model $\boldsymbol{Y}_j = \overline{\boldsymbol{Y}}_j^{\text{NLoS}}$ (see below for details). This simplification allows LoS paths to be treated and exploited in the same way as NLoS paths.

The key to achieve this is to introduce virtual scatter for LoS paths: a physical LoS path can be viewed as a NLoS path where an artificial scatter is added on the straight line segment between MD and AP. In proposed MuG method, the LoS path is alternatively viewed as a special case of NLoS path which originates from MD $\boldsymbol{m}$ and scattered on virtual scatter $\overline{\boldsymbol{s}} = [\overline{s}^x, \overline{s}^y]^{\text{T}}$. Also note that the similar concept was very briefly mentioned in [15]. The method proposed in [15] only studies localisation problem under OLoS environment, the DoA information of LoS path is omitted since the virtual scatter is forced to locate where AP is, which leads to a quite different from ours. Compared with the method in literature, our novel formulation brings two benefits: 1) The LoS/NLoS path separation is avoided completely, and hence, the prior statistical information of multipath is not required for separating LoS/NLoS paths. Conventional approaches suggest separating LoS/NLoS based on statistical information, and only using LoS paths for localisation.

Note that the mistaken separation cannot be avoided completely. If the actual path is a scattered path but it is mistakenly regarded as a direct path, then localisation fails, and vice versa. 2) The MuG method relies on NLoS paths for localisation which can be used for tackling more general localisation problem. For example, in urban or indoor scenarios, the number of LoS paths is typically limited, but there are sufficient NLoS paths.

The technical difficulty of using a virtual scatter is that the locations of virtual scatters are not uniquely. As illustrated in Fig 4.8 (a), in case of LoS path, it is clear that the virtual scatter can be any point along the path between MD and AP. Thus, directly applying virtual scatter will lead to infinite number of solutions. In order to solve this issue, the proposed objective function (4.61) in Section 4.3.4 is imposed sparsity constraints to exploit the fact that multipath actually originates from the minimal common MD and scattered on minimal common (real or virtual) scatters. From geometrical point of view, as illustrated in Fig 4.8 (b), when there exists multiple LoS paths from MD to APs, the common virtual scatter of LoS paths is exactly located at the position of MD.

Based on the virtual scatter $\overline{\boldsymbol{s}}$, the received LoS path can be reformulated as a special NLoS path

$$\boldsymbol{Y}_j^{\mathrm{LoS}} = \boldsymbol{Y}_j^{\mathrm{NLoS}}(\boldsymbol{m}, \overline{\boldsymbol{s}}). \tag{4.47}$$

The signal model introduced in (4.1) is replaced with an unified forward model associated with both location of MD $\boldsymbol{m}$, virtual scatter $\overline{\boldsymbol{s}}$ and scatters $\{\boldsymbol{s}\}_{k=1}^{K-1}$, which is the sum of NLoS paths

$$\begin{aligned} \boldsymbol{Y}_j &= \overline{\boldsymbol{Y}}_j^{\mathrm{NLoS}} + \boldsymbol{V}_j \\ &= \boldsymbol{Y}_j^{\mathrm{NLoS}}(\boldsymbol{m}, \overline{\boldsymbol{s}}) + \boldsymbol{Y}_j^{\mathrm{NLoS}}(\boldsymbol{m}, \boldsymbol{s}) + \boldsymbol{V}_j. \end{aligned} \tag{4.48}$$

In order to simplify later notation, in the rest of the chapter, we denote both real and

virtual scatter by $\{\boldsymbol{s}\}_{k=1}^{K}$. The received signal in (4.48) is written as

$$\boldsymbol{Y}_j = \sum_{k=1}^{K} \gamma_{j,k} \boldsymbol{h}\left(\tau_{j,k}^{\mathrm{NLoS}}\left(\boldsymbol{m},\boldsymbol{s}\right)\right) \boldsymbol{a}\left(\theta_{j,k}^{\mathrm{NLoS}}\left(\boldsymbol{m},\boldsymbol{s}\right)\right)^{\mathrm{T}} + \boldsymbol{V}_j. \tag{4.49}$$

The technical development of forward model can be simplified as follows. We work on the DFT coefficients $\boldsymbol{r}_j = \mathrm{DFT}(\mathrm{vec}(\boldsymbol{Y}_j)) \in \mathbb{C}^{NN_R}$ of the received signals $\boldsymbol{Y}_j$, where the subscript j denotes the signals at the j-th AP, vec() is vectorised operator, N is the number of time instances for taking samples, N_R is the number of antenna elements, and DFT transform is taken on $\mathrm{vec}(\boldsymbol{Y}_j)$. Then the signal model $\boldsymbol{r}_j \in \mathbb{C}^{NN_R}$ is given by

$$\boldsymbol{r}_j = \sum_{k=1}^{K} \gamma_{j,k} \boldsymbol{b}_j(\boldsymbol{m},\boldsymbol{s}_k) + \boldsymbol{v}_j \tag{4.50}$$

where $\boldsymbol{b}_j(\boldsymbol{m},\boldsymbol{s})$ represents the direct map from locations of the MD $\boldsymbol{m}$ and a scatter $\boldsymbol{s}$ to the received signal at the j-th AP:

$$(\boldsymbol{m},\boldsymbol{s}) \in \mathbb{R}^2 \times \mathbb{R}^2 \mapsto \boldsymbol{b}_j\left(\boldsymbol{m},\boldsymbol{s}\right) \in \mathbb{C}^{NN_R}, \tag{4.51}$$

$$\boldsymbol{b}_j(\boldsymbol{m},\boldsymbol{s}) = \mathrm{vec}\left(\hat{\boldsymbol{h}}_F\left(\tau_j\left(\boldsymbol{m},\boldsymbol{s}\right)\right) \boldsymbol{a}\left(\theta_j(\boldsymbol{m},\boldsymbol{s})\right)^{\mathrm{T}}\right). \tag{4.52}$$

The $\hat{\boldsymbol{h}}_F(\tau_j\left(\boldsymbol{m},\boldsymbol{s}\right)) \in \mathbb{C}^N$ is expressed as

$$\hat{\boldsymbol{h}}_F(\tau_j\left(\boldsymbol{m},\boldsymbol{s}\right)) = \left[\cdots, h_F(n) e^{-i2\pi f_s n \tau_j(\boldsymbol{m},\boldsymbol{s})/N}, \cdots\right]^{\mathrm{T}}, \tag{4.53}$$

where h_F is transmitted signal h in frequency domain; computation of $\tau_j\left(\boldsymbol{m},\boldsymbol{s}\right)$ follows the form of (4.8); $\boldsymbol{a}\left(\theta_j(\boldsymbol{m},\boldsymbol{s})\right) \in \mathbb{C}^{N_R}$ is the steering vector of DoA obtained by substituting (4.9) into (4.5); $\boldsymbol{v}_j$ is the additive noise, i.e., which follows an additive white Gaussian distribution with zero mean and variance σ^2.

Different from the indirect method which relies on estimating intermediate parameters such as ToF and DoA [1–4], our approach is based on a direct mapping from locations to the received signals. The indirect methods are suboptimal because they estimate the

intermediate parameters of each AP independently. In our proposed MuG method, it takes all received signal into account that both MD and scatters are simultaneously estimated. Intermediate parameters are not explicitly used and the notorious data association problem is avoided.

4.3.3 Sparse Recovery Formulation

In this subsection, the non-linear inverse localisation problem is reformulated as a sparse recovery problem, which has been studied extensively in the literature. Given the measurements at j-th AP $\boldsymbol{r}_j = \boldsymbol{\Psi}_j \boldsymbol{\gamma}_j + \boldsymbol{v}_j$, where $\boldsymbol{\Psi}_j$ is an $NN_R \times N_g^2$ pre-defined measurement matrix with $NN_R \ll N_g^2$ of K sparse solution $\boldsymbol{\gamma}_j$. More specifically, $\forall j \in [J]$,

$$\boldsymbol{r}_j = \underbrace{\left[\boldsymbol{b}_j(\boldsymbol{m}_1, \boldsymbol{s}_1), \cdots, \boldsymbol{b}_j(\boldsymbol{m}_{N_g}, \boldsymbol{s}_{N_g})\right]}_{\boldsymbol{\Psi}_j} \underbrace{\begin{bmatrix} \gamma_{j,1} \\ \vdots \\ \gamma_{j,N_g N_g} \end{bmatrix}}_{\boldsymbol{\gamma}_j} + \boldsymbol{v}_j, \tag{4.54}$$

where $\boldsymbol{\gamma}_j = [\gamma_{j,1}, \cdots, \gamma_{j,N_g N_g}]^{\mathrm{T}}$ is the coefficients at the j-th AP. Without loss of generality, we assume there are multiple MDs $\{\boldsymbol{m}\}_1^{N_g}$ in the environment. For every potential location $(\boldsymbol{m}, \boldsymbol{s})$, a vector $\boldsymbol{b}_j(\boldsymbol{m}, \boldsymbol{s})$ can be calculated according to (4.52) and it becomes a column of $\boldsymbol{\Psi}_j$ in (4.54). The unknown variables are $\boldsymbol{\gamma}_j : \gamma_{j,n_g} \neq 0, \text{for } n_g = 1, ..., N_g$, if and only if the received signal at the j-th AP contains a path originated from $\boldsymbol{m}_{n_g}$ and scattered at $\boldsymbol{s}_{n_g}$; the value γ_{j,n_g} depends on path attenuation, and the signal strength loss due to scattering.

Localisation problem is equivalent to recovery $\boldsymbol{\gamma}_j$ from these linear measurements. Typically, the number of observation NN_R is much smaller than N_g^2, in which the linear inverse problem (4.54) is underdetermined. To solve the problem, it is reasonable to assume that MD and scatters are not everywhere, equivalently, there are only a few non-zero entries in vector $\boldsymbol{\gamma}_j$. A canonical technique for sparse recovery is solved by compressive sens-

ing (CS) [60, 72, 108, 117]. In this case, $\boldsymbol{\gamma}_j$ is the solution of the following constrained optimisation problem:

$$\min_{\boldsymbol{\gamma}_j} \sum_{j}^{J} \| \left(\boldsymbol{r}_j - \boldsymbol{\Psi}_j \boldsymbol{\gamma}_j\right) \|_2^2 + \mathcal{G}(\boldsymbol{\Gamma}) \,, \tag{4.55}$$

where $\mathcal{G}(\boldsymbol{\Gamma})$ is the penalty function to promote sparse solution, $\boldsymbol{\Gamma} = [\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_J]$. The technical difficulties come from several aspects:

- How to design the measurement matrix $\boldsymbol{\Psi}_j$. In conventional CS techniques, the $\boldsymbol{\Psi}_j$ is obtained by discretising parameter space into grid points. However, discretisation leads to the leakage effect when the ground-truth is off-the-grid. The signal generated from an off-the-grid $(\boldsymbol{m}, \boldsymbol{s})$ typically cannot be well approximated by signals from a small number of grid points. It has been observed that off-the-grid points typically result in approximation error that is proportional to signals themselves, and deteriorate the estimation performance. On the other hand, a finer grid may mitigate the off-the-grid leakage but may result in large computational cost. Each tuple $(\boldsymbol{m}, \boldsymbol{s})$ has four variables m^x, m^y, s^x, s^y. An N_g-point grid of location of MD $\boldsymbol{m}$ and scatter $\boldsymbol{s}$ results in N_g^2 many grid points.

- How to design the sparse penalty function $\mathcal{G}(\boldsymbol{\Gamma})$ to exploit the truth that multipath is originating from one MD and scattered from a few scatters. In the literature, the sparse penalty functions are designed only for separating LoS paths from NLoS paths and exploiting information carried by LoS paths for localisation, such as L_1 norm [47] and $L_{2,1}$ norm [15].

4.3.4 A Grid-free Formulation

In order to tackle the first difficulty proposed in Section 4.3.3, we cast sparse inverse problem as a grid-free formulation where the parameters $(\boldsymbol{m}, \boldsymbol{s})$ are defined on continuous space. This is significantly different from [15, 47, 114] where a discrete grid is used as the parameter space. Our grid-free formulation avoids large computational complexity of

ultra-fine discrete grids, or self-interference caused by mismatch between discrete grid points and actual locations.

The following concepts are needed for the grid-free formulation. We consider a more general problem setting that an environment may contain multiple MDs and scatters. Denote the number of MD by I, and the number of scatters by G. Then the number of possible multipath in the environment is JIG. For the j-th AP, the location of the i-th MD and the g-th scatter are encoded using an atom $\boldsymbol{b}_j(\boldsymbol{m}_i, \boldsymbol{s}_g)$, where $(\boldsymbol{m}_i, \boldsymbol{s}_g) \in \mathbb{R}^2 \times \mathbb{R}^2$. The corresponding coefficient is denoted as $\gamma_{j,i,g}$. Based on atom and coefficient, the noiseless measurement $\boldsymbol{x}_j \in \mathbb{C}^{NN_R}$ at the j-th AP is formulated as:

$$\boldsymbol{x}_j = \sum_{i=1}^{I} \sum_{g=1}^{G} \gamma_{j,i,g} \boldsymbol{b}_j(\boldsymbol{m}_i, \boldsymbol{s}_g). \tag{4.56}$$

In order to design a sparse penalty function $\mathcal{G}(\cdot)$ to exploit multipath for localisation, there are two aspects that need to be exploited. The first aspect is based on the view of *sparse recovery formulation*, and the second aspect is based on the view of *propagation environment*.

In the view of *sparse recovery*, the signal $\boldsymbol{x}_j$ is assumed to consist of a few weighted atoms $\{\boldsymbol{b}_j(\boldsymbol{m}_i, \boldsymbol{s}_g)\}_{i=1,g=1}^{I,G}$. Therefore, it is natural to impose sparsity constraint to seek a parsimonious decomposition of the signal that contains the smallest number of atoms. Then the atomic norm $\|\cdot\|_{\mathcal{A}}$ is used to promote an overall sparsity of the composing atoms. The atomic norm is defined to promote sparsity of number of multipath for each AP [51, 72]:

$$\|\boldsymbol{\mathcal{X}}\|_{\mathcal{A}} = \inf_{\boldsymbol{M},\boldsymbol{\Gamma},\boldsymbol{S}} \left\{ \sum_{j=1}^{J} \sum_{i=1}^{I} \sum_{g=1}^{G} |\gamma_{j,i,g}| : \boldsymbol{x}_j = \sum_{i} \sum_{g} \gamma_{j,i,g} \boldsymbol{b}_j\left(\boldsymbol{m}_i, \boldsymbol{s}_g\right), j = 1, \ldots, J \right\}, \tag{4.57}$$

where $\boldsymbol{\mathcal{X}} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_J]$ is the matrix of noiseless measurements; $\boldsymbol{M} = [\boldsymbol{m}_1, \cdots, \boldsymbol{m}_I]$ is the matrix of MDs, $\boldsymbol{\Gamma} = [\boldsymbol{\gamma}_j, \cdots, \boldsymbol{\gamma}_J]$ is the matrix of coefficients; $\boldsymbol{\gamma}_j = [\gamma_{j,1,1}, \cdots, \gamma_{j,i,g}, \cdots, \gamma_{j,I,G}]^{\mathrm{T}}$ is coefficients at the j-th AP ; $\boldsymbol{S} = [\boldsymbol{s}_1, \cdots, \boldsymbol{s}_G]$ is the matrix of scatters.

Mathematically speaking, all the true atoms $\boldsymbol{b}_j(\boldsymbol{m}, \boldsymbol{s}), j = 1, \cdots, J$ are defined on the same support $\{\boldsymbol{m}_i, \boldsymbol{s}_g\}_{i=1,g=1}^{I,G}$, which is referred as common MD and scatters. For each AP, the received multipath is originating from common MDs $\{\boldsymbol{m}_i\}_{i=1}^{I}$ and scattered on scatters $\{\boldsymbol{s}_g\}_{g=1}^{G}$ but with different attenuation factors. The group atomic norm encourages the group sparsity of coefficients. We use $\|\cdot\|_{\mathcal{MS}}$ in our formulation to promote this share of locations among APs:

$$\|\boldsymbol{\mathcal{X}}\|_{\mathcal{MS}} = \inf_{\boldsymbol{M},\boldsymbol{\Gamma},\boldsymbol{S}} \left\{ \sum_{i=1}^{I} \sum_{g=1}^{G} \sqrt{\sum_{j=1}^{J} |\gamma_{j,i,g}|^2} : \boldsymbol{x}_j = \sum_i \sum_g \gamma_{j,i,g} \boldsymbol{b}_j \left(\boldsymbol{m}_i, \boldsymbol{s}_g\right), j = 1, \ldots, J \right\}. \tag{4.58}$$

In the view of *propagation environment*, it is reasonable to assume that propagation environment is consisted of minimal number of MDs and scatters. Moreover, finding a sparse solution of MDs and scatters is well-motivated since the the true solution is sparse and unique, see section 4.3.1 for details. To promote sparsity of MDs, the group atomic norm $\|\cdot\|_{\mathcal{M}}$ on MDs is introduced:

$$\|\boldsymbol{\mathcal{X}}\|_{\mathcal{M}} = \inf_{\boldsymbol{M},\boldsymbol{\Gamma},\boldsymbol{S}} \left\{ \sum_{i=1}^{I} \sqrt{\sum_{g=1}^{G} \sum_{j=1}^{J} |\gamma_{j,i,g}|^2} : \boldsymbol{x}_j = \sum_i \sum_g \gamma_{j,i,g} \boldsymbol{b}_j \left(\boldsymbol{m}_i, \boldsymbol{s}_g\right), j = 1, \ldots, J \right\}. \tag{4.59}$$

To promote sparsity of scatters, the group atomic norm $\|\cdot\|_{\mathcal{S}}$ on scatters is introduced:

$$\|\boldsymbol{\mathcal{X}}\|_{\mathcal{S}} = \inf_{\boldsymbol{M},\boldsymbol{\Gamma},\boldsymbol{S}} \left\{ \sum_{g=1}^{G} \sqrt{\sum_{i=1}^{I} \sum_{j=1}^{J} |\gamma_{j,i,g}|^2} : \boldsymbol{x}_j = \sum_i \sum_g \gamma_{j,i,g} \boldsymbol{b}_j \left(\boldsymbol{m}_i, \boldsymbol{s}_g\right), j = 1, \ldots, J \right\}. \tag{4.60}$$

With definitions of atomic norm and group atomic norms, the localisation problem can

be then formulated as a LASSO-type optimisation programming:

$$\min_{\boldsymbol{M},\boldsymbol{\Gamma},\boldsymbol{S}} \sum_{j}^{J} \| \left(\boldsymbol{r}_j - \boldsymbol{x}_j\right) \|_2^2 + \mathcal{G}(\boldsymbol{\mathcal{X}}) \,, \tag{4.61}$$

$$\begin{aligned} \text{s.t. } \mathcal{G}(\boldsymbol{\mathcal{X}}) &= \lambda_1 \|\boldsymbol{\mathcal{X}}\|_{\mathcal{A}} + \lambda_2 \|\boldsymbol{\mathcal{X}}\|_{\mathcal{MS}} \\ &\quad + \lambda_3 \|\boldsymbol{\mathcal{X}}\|_{\mathcal{M}} + \lambda_4 \|\boldsymbol{\mathcal{X}}\|_{\mathcal{S}} \,, \\ I &= 1, \end{aligned}$$

where $\lambda_1, \lambda_2, \lambda_3, \lambda_4 > 0$ are regularisation constants; $[J] := \{1, \cdots, J\}$; the number of MD is set $I = 1$ because the signals from different MDs are orthogonal so that without loss of generality it suffices to consider only one MD. The cost function (4.61) favours solution that has the sparse number of common MD and scatters.

4.3.5 Solving the Super-resolution Problem

In this section we develop a variant of the alternating descent conditional gradient (ADCG) [108] called MuG to solve the problem in (4.61). Remark that MuG is not simple application of ADCG, since the unknown parameters are associated with different number of parameter $\{\boldsymbol{m}_i, \boldsymbol{s}_g\}_{i=1,g=1}^{I,G}$ for $I \neq G$. It is non-trivial to solve the problem (4.61), since the unknown parameters are defined over infinite parameter space. In order to jointly estimate location of MD and scatters simultaneously, we use the gradient descent approach and detailed computation is presented accordingly. Similarly with the ADCG, each iteration of MuG method has the three main steps to estimate location of MD and scatters : conditional gradient step, the pruning support step and local descent step. The aim of conditional gradient is finding possible estimation of the problem. The pruning support step is used to produce sparse solution. The final step is improving convergence speed and sparsity level of solution. We make further description in following about detailed implementations of Algorithm 4.

The conditional gradient step is solving an approximation to (4.61) based on lineari-

Algorithm 4: MuG: Multipath-exploited and Grid-free Method

Input: $\mathcal{M}_0 = \emptyset, \mathcal{S}_0 = \emptyset$, $d = 1$, ϵ, $\boldsymbol{r}_j$, $j = 1, ..., J$

Output: Locations of MD : $\mathcal{M}^d$ and scatter : $\mathcal{S}^d$.

repeat

(1) Compute gradient of loss :

$$\boldsymbol{g}_j^{d-1} = \nabla_{\boldsymbol{r}_{\mathrm{res}j}^{d-1}} \left\| \boldsymbol{r}_{\mathrm{res}j}^{d-1} \right\|_2^2, \tag{4.62}$$

where

$$\boldsymbol{r}_{\mathrm{res}j}^{d-1} = \boldsymbol{r}_j - \sum_i \sum_g \boldsymbol{b}_j(\boldsymbol{m}_i, \boldsymbol{s}_g)\gamma_{j,i,g}, \tag{4.63}$$
$$\forall \boldsymbol{m}_i \in \mathcal{M}^{d-1}, \forall \boldsymbol{s}_g \in \mathcal{S}^{d-1};$$

(2) Conditional gradient:

$$\left(\boldsymbol{m}^d, \boldsymbol{s}^d\right) = \arg\max_{\boldsymbol{m}^d, \boldsymbol{s}^d} \sum_{j=1} \left| \left\langle \boldsymbol{b}_j \left(\boldsymbol{m}^d, \boldsymbol{s}^d\right), \boldsymbol{g}_j^{d-1} \right\rangle \right|; \tag{4.64}$$

(3) Update the candidate set :

$$\mathcal{M}^d = \mathcal{M}^{d-1} \cup \boldsymbol{m}^d, \quad \mathcal{S}^d = \mathcal{S}^{d-1} \cup \boldsymbol{s}^d; \tag{4.65}$$

for t **do**

1) Compute weights:

$$\begin{aligned} \boldsymbol{\gamma}^d = \arg\min_{\boldsymbol{\gamma}} & \sum_j \left\| \boldsymbol{r}_j - \sum_i \sum_g \boldsymbol{b}_j(\boldsymbol{m}_i, \boldsymbol{s}_g)\gamma_{j,i,g} \right\|_2^2 \\ & + \lambda_1 \sum_j \sum_i \sum_g |\gamma_{j,i,g}| + \lambda_2 \sum_i \sum_g \|\boldsymbol{\gamma}_{:,i,g}\|_2 \\ & + \lambda_3 \sum_i \|\boldsymbol{\gamma}_{:,i,:}\|_F + \lambda_4 \sum_g \|\boldsymbol{\gamma}_{:,:,g}\|_F, \\ & \forall \boldsymbol{m}_i \in \mathcal{M}^d, \forall \boldsymbol{s}_g \in \mathcal{S}^d; \end{aligned} \tag{4.66}$$

2) Prune support:

$$\mathcal{M}^d = \boldsymbol{m}_i, \text{for } i = \arg\max \|\boldsymbol{\gamma}_{:,i,:}^d\|_F, \tag{4.67}$$

$$\mathcal{S}^d = \operatorname*{supp}_{\boldsymbol{s}_g \in \mathcal{S}^d} \left(\boldsymbol{\gamma}_{:,i,g}^d\right), \text{for } i = \arg\max \|\boldsymbol{\gamma}_{:,i,:}^d\|_F. \tag{4.68}$$

3) Locally improve support :

$$\begin{aligned} \mathcal{M}^d, \mathcal{S}^d = & \text{gradient\_descent}\left(\left(\boldsymbol{m}_i, \boldsymbol{s}_g\right), \boldsymbol{\gamma}\left(\{\boldsymbol{m}_i, \boldsymbol{s}_g\}\right)\right) \\ & ,\forall \boldsymbol{m}_i \in \mathcal{M}^d, \forall \boldsymbol{s}_g \in \mathcal{S}^d. \end{aligned} \tag{4.69}$$

end for

(4) Update parameters: $r_{\mathrm{cur}} = \sum_j^J \left\| \boldsymbol{r}_{\mathrm{res}j}^d \right\|_2^2$, $r_{\mathrm{pre}} = \sum_j^J \left\| \boldsymbol{r}_{\mathrm{res}j}^{d-1} \right\|_2^2$, $d = d + 1$.

until $r_{\mathrm{cur}}/r_{\mathrm{pre}} < \epsilon$

sation around the current iteration [108]. As an extension of ADCG to estimate the unknowns with common sparsity, at each iteration, MuG adds single source locations $(\boldsymbol{m}^d, \boldsymbol{s}^d) \in [0,1]^4$ to the corresponding approximated candidate set $\mathcal{M}^d$, $\mathcal{S}^d$ which maximize the summation of the inner product of gradient of loss $\boldsymbol{g}_j^d$ with its corresponding atom $\boldsymbol{b}_j(\boldsymbol{m}^d, \boldsymbol{s}^d)$ at different APs, and detailed formulation is presented in (4.64). Motivated by distributed compressive sensing framework as presented in [118, 119], (4.64) is equivalent to select the source that maximizes the value of the sum of the magnitudes of the projections of the residual $\boldsymbol{r}_{\text{res}j}^{d-1}$. Specifically, by substituting (4.62), (4.63) into (4.64), it is easy to show that the conditional step (4.64) solves the following problem:

$$\begin{aligned}\left(\boldsymbol{m}^d, \boldsymbol{s}^d\right) &= \underset{(\boldsymbol{m},\boldsymbol{s})\in[0,1]^4}{\arg\max} \sum_{j=1} \left| \left\langle \boldsymbol{b}_j\left(\boldsymbol{m}, \boldsymbol{s}\right), \nabla_{\boldsymbol{r}_{\text{res}j}^{d-1}} \left\| \boldsymbol{r}_{\text{res}j}^{d-1} \right\|_2^2 \right\rangle \right| &(4.70)\\ &= \underset{(\boldsymbol{m},\boldsymbol{s})\in[0,1]^4}{\arg\max} \sum_{j=1} \left| \left\langle \boldsymbol{b}_j\left(\boldsymbol{m}, \boldsymbol{s}\right), 2\boldsymbol{r}_{\text{res}j}^{d-1} \right\rangle \right|, &(4.71)\end{aligned}$$

where $\boldsymbol{r}_{\text{res}j}^{d-1}$ is the residual of the j-th AP at the d-th iteration; and $\nabla_{\boldsymbol{r}_{\text{res}j}^{d-1}} \left\| \boldsymbol{r}_{\text{res}j}^{d-1} \right\|_2^2$ denotes derivative of the function $\left\| \boldsymbol{r}_{\text{res}j}^{d-1} \right\|_2^2$ at $\boldsymbol{r}_{\text{res}j}^{d-1}$. In the rest of chapter, we denote $\nabla_{\circ} C(\cdot)$ as the partial derivative of the cost function $C(\cdot)$ w.r.t. $\circ$. Note that (4.71) is non-convex optimisation problem and it has no closed-form solution. In the implementation, the new sources are obtained by the gradient descent method over 4 dimensional continuous parameter space. The corresponding gradient is then

$$\nabla_{(\boldsymbol{m},\boldsymbol{s})} \sum_{j=1} \left| \left\langle \boldsymbol{b}_j\left(\boldsymbol{m}, \boldsymbol{s}\right), 2\boldsymbol{r}_{\text{res}j}^{d-1} \right\rangle \right| = \sum_{j=1} \frac{\partial}{\partial(\boldsymbol{m}, \boldsymbol{s})} \left| \left\langle \frac{\boldsymbol{b}_j\left(\boldsymbol{m}, \boldsymbol{s}\right)}{(\boldsymbol{m}, \boldsymbol{s})}, 2\boldsymbol{r}_{\text{res}j}^{d-1} \right\rangle \right|. \tag{4.72}$$

For ease of notation, we denote $v_j = \left\langle \boldsymbol{b}_j\left(\boldsymbol{m},\boldsymbol{s}\right), 2\boldsymbol{r}_{\mathrm{res}j}^{d-1}\right\rangle$, and the gradient becomes

$$\begin{aligned}\nabla_{(\boldsymbol{m},\boldsymbol{s})}\sum_{j=1}\left|\left\langle \boldsymbol{b}_j\left(\boldsymbol{m},\boldsymbol{s}\right), 2\boldsymbol{r}_{\mathrm{res}j}^{d-1}\right\rangle\right| &= \sum_j \frac{\partial\left|v_j\right|}{\partial(\boldsymbol{m},\boldsymbol{s})} \qquad (4.73)\\ &= \sum_j \frac{\partial\left(v_j^{\mathrm{H}}v_j\right)^{\frac{1}{2}}}{\partial(\boldsymbol{m},\boldsymbol{s})}\\ &= \sum_j \left(v_j^{\mathrm{H}}v_j\right)^{-\frac{1}{2}}\mathrm{Re}\left(\frac{\partial v_j}{\partial(\boldsymbol{m},\boldsymbol{s})}v_j^{\mathrm{H}}\right), \qquad (4.74)\end{aligned}$$

where

$$\begin{aligned}\frac{\partial v_j}{\partial(\boldsymbol{m},\boldsymbol{s})} &= \left(2\boldsymbol{r}_{\mathrm{res}j}^{d-1}\right)^{\mathrm{H}}\frac{\partial \boldsymbol{b}_j\left(\boldsymbol{m},\boldsymbol{s}\right)}{\partial(\boldsymbol{m},\boldsymbol{s})} \qquad (4.75)\\ &= \left(2\boldsymbol{r}_{\mathrm{res}j}^{d-1}\right)^{\mathrm{H}}\frac{\partial\tau_j\left(\boldsymbol{m},\boldsymbol{s}\right)}{\partial(\boldsymbol{m},\boldsymbol{s})}\boldsymbol{D}_1\boldsymbol{b}_j\left(\boldsymbol{m},\boldsymbol{s}\right) + \left(2\boldsymbol{r}_{\mathrm{res}j}^{d-1}\right)^{\mathrm{H}}\frac{\partial\theta_j\left(\boldsymbol{s}\right)}{\partial(\boldsymbol{m},\boldsymbol{s})}\boldsymbol{b}_j\left(\boldsymbol{m},\boldsymbol{s}\right)\boldsymbol{D}_2,\end{aligned}$$

and $\boldsymbol{D}_1, \boldsymbol{D}_2$ are two diagonal matrices as

$$\begin{aligned}\boldsymbol{D}_1 =& \mathrm{diag}\left(\left[-i2\pi f, \cdots, -i2\pi fn, \cdots, -i2\pi fN\right]\right), \qquad (4.76)\\ \boldsymbol{D}_2 =& \mathrm{diag}\left(\left[0, \cdots, i\frac{2\pi}{\lambda}L\cos(\theta_j)(n_R-1), \cdots, i\frac{2\pi}{\lambda}L\cos(\theta_j)(N_R-1)\right]\right). \qquad (4.77)\end{aligned}$$

The derivate of ToF $\tau_j(\boldsymbol{m},\boldsymbol{s})$ with respect to location of MD $\boldsymbol{m}$ is

$$\frac{\partial\tau_j(\boldsymbol{m},\boldsymbol{s})}{\partial\boldsymbol{m}} = \left[\frac{\partial\tau_j(\boldsymbol{m},\boldsymbol{s})}{\partial m^x} \quad \frac{\partial\tau_j(\boldsymbol{m},\boldsymbol{s})}{\partial m^y}\right]^{\mathrm{T}}, \qquad (4.78)$$

where

$$\frac{\partial\tau_j(\boldsymbol{m},\boldsymbol{s})}{\partial m^x} = \frac{1}{c\|\boldsymbol{m}-\boldsymbol{s}\|_2}(m^x - s^x), \qquad (4.79)$$

and

$$\frac{\partial\tau_j(\boldsymbol{m},\boldsymbol{s})}{\partial m^y} = \frac{1}{c\|\boldsymbol{m}-\boldsymbol{s}\|_2}(m^y - s^y). \qquad (4.80)$$

The derivate of ToF $\tau_j(\boldsymbol{m}, \boldsymbol{s})$ with respect to location of scatter $\boldsymbol{s}$ is

$$\frac{\partial \tau_j(\boldsymbol{m}, \boldsymbol{s})}{\partial \boldsymbol{s}} = \left[\frac{\partial \tau_j(\boldsymbol{m}, \boldsymbol{s})}{\partial s^x} \quad \frac{\partial \tau(\boldsymbol{m}, \boldsymbol{s})}{\partial s^y}\right]^{\mathrm{T}}, \tag{4.81}$$

where

$$\frac{\partial \tau_j(\boldsymbol{m}, \boldsymbol{s})}{\partial s^x} = \frac{1}{c\|\boldsymbol{m} - \boldsymbol{s}\|_2}(s^x - m^x) + \frac{1}{c\|\boldsymbol{s} - \boldsymbol{A}\boldsymbol{P}_j\|_2}(s^x - AP^x) \tag{4.82}$$

and

$$\frac{\partial \tau_j(\boldsymbol{m}, \boldsymbol{s})}{\partial s^y} = \frac{1}{c\|\boldsymbol{m} - \boldsymbol{s}\|_2}(s^y - m^y) + \frac{1}{c\|\boldsymbol{s} - \boldsymbol{A}\boldsymbol{P}_j\|_2}(s^y - AP^y) \tag{4.83}$$

The derivate of DoA $\theta_j(\boldsymbol{s})$ with respect to location of scatter $\boldsymbol{s}$ is

$$\frac{\partial \theta_j(\boldsymbol{s})}{\partial \boldsymbol{s}} = \left[\frac{\partial \theta_j(\boldsymbol{s})}{\partial s^x} \quad \frac{\partial \theta_j(\boldsymbol{s})}{\partial s^y}\right]^{\mathrm{T}}, \tag{4.84}$$

where

$$\frac{\partial \theta_j(\boldsymbol{s})}{\partial s^x} = \frac{1}{1 + (\tan(\theta_j))^2}\left(s^y - AP_j^y\right)^{-1} \tag{4.85}$$

and

$$\frac{\partial \theta_j(\boldsymbol{s})}{\partial s^y} = \frac{1}{1 + (\tan(\theta_j))^2}(AP^x - s^x)\left(s^y - AP_j^y\right)^{-2}. \tag{4.86}$$

The pruning support step aims to produce sparse solution. As shown in (4.67), the source locations are removed from candidate set when they are found not contributing to reducing the loss function (4.66). The coefficients $\boldsymbol{\gamma}^d$ is obtained by solving group lasso problem (4.66), which is a finite-dimensional convex optimization problem that can be solved with an off-the-shelf algorithm. This is motivated by the drawback of greed step (4.64): the source location added at previous iteration may not be helpful as compared with other

later added sources.

The aim of final step is improving convergence speed and sparsity level of solution by smoothly moving estimated set $\mathcal{M}^d$, $\mathcal{S}^d$ within continuous parameter space $[0,1]^{4d}$. By fixing the attenuation coefficients $\boldsymbol{\gamma}^d$, the local descent step is implemented with minimising following optimisation problem

$$\arg\min_{\boldsymbol{m}_i,\boldsymbol{s}_g} \sum_j^J \left\| \boldsymbol{r}_j - \sum_i \sum_g \boldsymbol{b}_j(\boldsymbol{m}_i,\boldsymbol{s}_g)\gamma_{j,i,g} \right\|_2^2, \tag{4.87}$$

$$\forall \boldsymbol{m}_i \in \mathcal{M}^d, \forall \boldsymbol{s}_g \in \mathcal{S}^d. \tag{4.88}$$

In the most cases, the gradient descent method is an efficient approach to reduce the function. The corresponding gradient is

$$\begin{aligned} &\nabla_{(\boldsymbol{m}_i\in\mathcal{M}^d,\boldsymbol{s}_g\in\mathcal{S}^d)} \sum_j^J \left\| \boldsymbol{r}_j - \sum_i \sum_g \boldsymbol{b}_j(\boldsymbol{m}_i,\boldsymbol{s}_g)\gamma_{j,i,g} \right\|_2^2 \\ &= -2\sum_j \sum_i \sum_g \mathrm{Re}\left(\frac{\partial \boldsymbol{b}_j^{\mathrm{T}}(\boldsymbol{m}_i,\boldsymbol{s}_g)}{\partial(\boldsymbol{m}_i,\boldsymbol{s}_g)}\gamma_{j,i,g} \right) \mathrm{Re}\left(\boldsymbol{r}_j - \sum_i \sum_g \boldsymbol{b}_j(\boldsymbol{m}_i,\boldsymbol{s}_g)\gamma_{j,i,g} \right), \end{aligned} \tag{4.89}$$

based on the equation (4.75), we have

$$\frac{\partial \boldsymbol{b}_j(\boldsymbol{m}_i,\boldsymbol{s}_g)}{\partial(\boldsymbol{m}_i,\boldsymbol{s}_g)} = \frac{\partial \tau_j(\boldsymbol{m}_i,\boldsymbol{s}_g)}{\partial(\boldsymbol{m}_i,\boldsymbol{s}_g)} \boldsymbol{D}_1 \boldsymbol{b}_j(\boldsymbol{m}_i,\boldsymbol{s}_g) + \frac{\partial \theta_j(\boldsymbol{m}_i,\boldsymbol{s}_g)}{\partial(\boldsymbol{m}_i,\boldsymbol{s}_g)} \boldsymbol{b}_j(\boldsymbol{m}_i,\boldsymbol{s}_g) \boldsymbol{D}_2,$$

The detailed computations of $\frac{\partial \tau_j(\boldsymbol{m}_i,\boldsymbol{s}_g)}{\partial(\boldsymbol{m}_i,\boldsymbol{s}_g)}$ and $\frac{\partial \theta_j(\boldsymbol{m}_i,\boldsymbol{s}_g)}{\partial(\boldsymbol{m}_i,\boldsymbol{s}_g)}$ are shown in (4.78) - (4.86).

4.3.6 Cramér-Rao Bound (CRB)

The CRB is the lower bound on the variance of the unbiased estimatator for unknown parameters [113]. Define all the unknown variables as

$$\boldsymbol{p} = \left[\boldsymbol{\alpha}^{\mathrm{T}}, \mathrm{Re}([..., \boldsymbol{\gamma}_j{}^{\mathrm{T}}, ...]), \mathrm{Im}([..., \boldsymbol{\gamma}_j{}^{\mathrm{T}}, ...) \right]^{\mathrm{T}}, \tag{4.90}$$

where $\boldsymbol{\alpha} = [\boldsymbol{m}^{\mathrm{T}}, \boldsymbol{s}_1{}^{\mathrm{T}}, \cdots, \boldsymbol{s}_k{}^{\mathrm{T}}, \cdots, \boldsymbol{s}_K{}^{\mathrm{T}},]$ contains unknown locations of MD and scatters, and $\boldsymbol{\gamma}_j = [\gamma_{j,1}, ..., \gamma_{j,k}, ...\gamma_{j,K}]^{\mathrm{T}}$ is the coefficient at the j-th AP. The CRB is a lower bound of the covariance of estimation error as

$$\mathrm{CRB}(\boldsymbol{p}) \leqslant \mathrm{E}\left\{[\boldsymbol{p} - \tilde{\boldsymbol{p}}]\,[\boldsymbol{p} - \tilde{\boldsymbol{p}}]^{T}\right\}, \tag{4.91}$$

where $\tilde{\boldsymbol{p}} = \left[\tilde{\boldsymbol{\alpha}}^{\mathrm{T}}, \mathrm{Re}([..., \tilde{\boldsymbol{\gamma}}_j{}^{\mathrm{T}}, ...]), \mathrm{Im}([..., \tilde{\boldsymbol{\gamma}}_j{}^{\mathrm{T}}, ...)\right]^{\mathrm{T}}$ is the estimated parameter.

For simplicity of notation, we use $f_j(\boldsymbol{p}, n, n_r)$ denoting the n-th noise free observation, at the n_r-th antenna array element and the j-th AP, which is

$$f_j(\boldsymbol{p}, n, n_r) = \sum_{k=1}^{K} \gamma_{j,k} b_j(\boldsymbol{m}, \boldsymbol{s}_k, n, n_r), \tag{4.92}$$

Then the noisy sample $r_j(n, n_r)$ corrupted by white Gaussian noise is expressed as

$$r_j(n, n_r) = f_j(\boldsymbol{p}, n, n_r) + v_j(n, n_r), \tag{4.93}$$

where $v_j(n, n_r)$ is additive noise i.i.d. white Gaussian with zero mean and variance σ^2.

The computation of CRB requires channel statistics, propagation path type, the number of MD and scatters, and noise variance as prior information. In the simulation, the prior information is usually unknown to MuG technique. The CRB is the inverse of the Fisher information matrix (FIM) $\boldsymbol{F}(\boldsymbol{p}) \in \mathbb{R}^{(2K+2KJ+2)\times(2K+2KJ+2)}$ as shown below [113]:

$$\mathrm{CRB}(\boldsymbol{p}) = \boldsymbol{F}^{-1}(\boldsymbol{p}). \tag{4.94}$$

The FIM is defined as

$$\boldsymbol{F}(\boldsymbol{p}) = \mathrm{E}\left(\frac{\partial \ln L(\boldsymbol{R}; \boldsymbol{p})}{\partial \boldsymbol{p}} \frac{\partial \ln L^{\mathrm{H}}(\boldsymbol{R}; \boldsymbol{p})}{\partial \boldsymbol{p}}\right), \tag{4.95}$$

where $L(\boldsymbol{R}; \boldsymbol{p})$ is the likelihood function of observation $\boldsymbol{R} = [\boldsymbol{r}_1, \cdots, \boldsymbol{r}_j, \cdots, \boldsymbol{r}_J]$ at all

APs conditioned with unknown parameters $\boldsymbol{p}$ and it can be expressed as

$$L(\boldsymbol{R};\boldsymbol{p}) = \prod_{j=1}^{J}\prod_{n=1}^{N}\prod_{n_r=1}^{N_R} \left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2\sigma^2}\left(r_j(n,n_r) - f_j(\boldsymbol{p},n,n_r)\right)^2\right]\right]. \tag{4.96}$$

Then the estimated location error $\sigma_{MD} = \sqrt{E(m^x - \tilde{m}^x)^2 + E(m^y - \tilde{m}^y)^2}$ of any unbiased localisation estimator has the lower bound at

$$\sigma_{MD} \geqslant \sqrt{\mathrm{CRB}(\boldsymbol{p})_{1,1} + \mathrm{CRB}(\boldsymbol{p})_{2,2}}, \tag{4.97}$$

where $\mathrm{CRB}(\boldsymbol{p})_{q,w}$ is the $[q,w]$-th entry of the CRB matrix. The entries of $\mathrm{CRB}(\boldsymbol{p})_{q,w}$ are derived in Appendix B.1.

4.4 Numerical Case Studies

4.4.1 Simulation Setup

The setup of case studies is as follows. Consider the problem of localisation a single antenna MD within 1km × 1km area under cooperative APs and three scatters. The locations of MD and three scatters are randomly generated as point sources [1–4, 15]. The transmitted signal can be chosen for different purpose. We take the orthogonal frequency division multiplexing (OFDM) as an example which contains 16 blocks of data. Each data block has $M = 32$ sub-carrier frequencies, the sub-carrier frequency spacing is $\triangle f = 10kHz$, the speed of light is $c = 3 \times 10^8$m/s, carrier frequency $f_c = 2GHz$. The path loss of LoS path follows an uniform distribution in the interval $[0, 0.01]$. The average reflection loss for the NLoS path is set to -10dB [3]. Noise follows an additive White Gaussian distribution with zero mean and variance σ^2. The threshold to terminate algorithm $\epsilon = 0.9999$. The performance of localisation technique is evaluated by the

RMSE, defined as spatial distance between true and estimated MD.

4.4.2 Performance of MuG and CRB

Two case studies are presented here, which includes performance of MuG method under different system configurations and different propagation environments. Motivated by practical scenarios that the number of LoS paths is typically limited, and the general scenario where the signal from the scatter may not be received by all APs. In simulation, we also include the performance of MuG with limited LoS and NLoS paths. It is also noteworthy that though our algorithm is propagation environment blind, the computation of CRB is not. The computation of CRB requires knowing which path is LoS and which is not, and the exact locations of scatters, both of which are absent in practice.

The first one concerns single antenna APs where there is no DoA information at all. We show it's possible to locate MD even when all the paths are NLoS, i.e., only exploiting NLoS paths without mitigating NLoS effect. According to our knowledge, this is the first time in the literature. As illustrated in Section 4.3.1, the success of localising MD requires at least 3 single antenna APs and 3 scatters, where cooperative APs are horizontally located at $(0, 0)$km, $(1, 0)$km and $(1, 1)$km. The numerical performance is presented in Fig 4.9 (a). The performance of MuG is shown in solid lines while CRB is shown as dashed lines. In LoS environment, the number of LoS paths is $L_p = 3$. In OLoS environment, the number of NLoS paths is $N_p = 9$ because there are three scatters; in NLoS environment, the multipath contains $L_p = 3$ LoS paths , $N_p = 9$ NLoS paths. All APs are assumed to be equipped with single antenna so that there is no DoA information to extract at APs. This case is particularly challenging because 1) throughout this chapter we don't assume any a priori information and historical data, so techniques based on statistical modelling [10–14] cannot be applied; 2) techniques separating LoS and NLoS paths based on DoA information [15] cannot be applied here either; 3) the paths with smallest ToFs can be NLoS ones. Note that although we show our results in different propagation

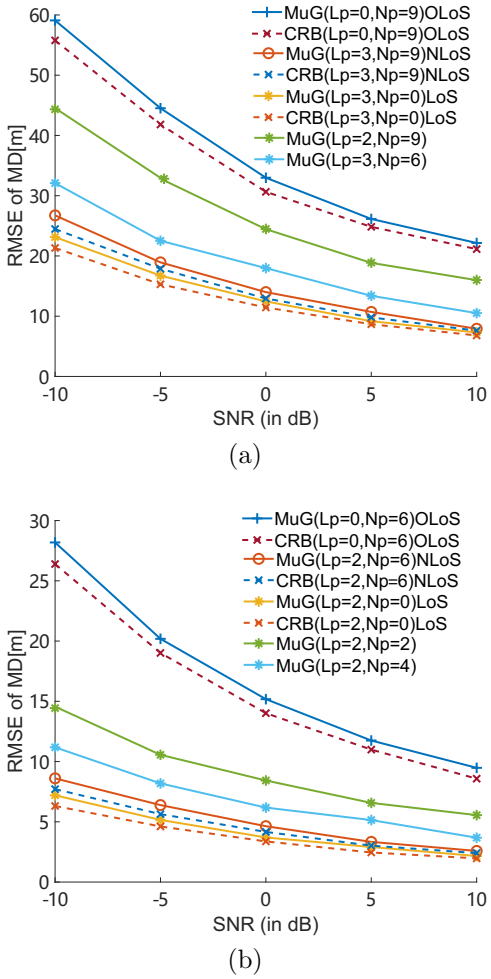


Figure 4.9: Simulation results of RMSE versus the signal-to-noise ratio (SNR) under different system configuration and environment. The L_p and N_p are total number of LoS paths and NLoS paths in the environment, respectively. The CRB represents the theoretical bounds under different propagation environment. (a) In the first case studies, 3 single antenna APs ($N_R = 1$) are employed for providing ToF information. (b) In the second case studies, 2 multi-antenna APs ($N_R = 12$) are employed for providing ToF and DoA information.

environments (LoS, NLoS and OLoS), the propagation environment is unknown to our algorithm.

The second case studies involves multi-antenna APs, which are located at $(0, 0)$km and $(1, 1)$km. As shown in Fig 4.9(b), the simulation result demonstrates the 'plug-and-play' nature of our method by directly applying it to different system configurations in various propagation environments. With extra DoA information, the minimum number of APs for localising MD is reduced to two. The number of multipath in LoS environment is $L_p = 2$, $N_p = 0$; in OLoS environment is $L_p = 0$, $N_p = 6$; in NLoS environment is $L_p = 2$,

$N_p = 6$. To tackle practical localisation problem with limited multipath, we include the performance of MuG under different number of LoS and NLoS paths and corresponding results are marked by a circle. It is observed that estimation performance of MuG under $L_p = 2$, $N_p = 2$ is improved by adding NLoS paths comparing with $L_p = 2$, $N_p = 4$ and $L_p = 2$, $N_p = 6$. In addition, increasing the number of LoS paths also can help improve the performance of MuG technique by comparing the performance of MuG under $L_p = 0$, $N_p = 6$ with $L_p = 2, N_p = 6$.

Environment	System Configuration	Time [min]
LoS	single antenna system	4.21
OLoS	single antenna system	8.02
NLoS	single antenna system	10.73
LoS	multi-antenna system	7.45
OLoS	multi-antenna system	13.78
NLoS	multi-antenna system	17.23

Table 4.2: In case of SNR = 10 dB, the average running time of the MuG in different environments and system configurations. CPU: Intel Core i7-7700HQ, 2.80GHz.

The computational cost of the MuG is measured by the average running time. In the numerical case studies, we assume that the exact propagation environment is unknown to the APs and there is no information for identifying or separating LoS/NLoS paths. Therefore, the existing methods are infeasible to be applied here. Table 4.2 shows the computational cost of the MuG in different environments and system configurations. The MuG method will stop when the improvement drops below a threshold $\epsilon = 0.9999$. In the LoS environment, the MuG has a lower running time than other environments. Under the same system configuration, the running time is increased by increasing the number of propagation paths. Based on the running time of LoS and OLoS environments, the LoS path can provide a high precision estimation with less computational cost.

In order to compare the performance of MuG with existing method in single antenna system, we assume that LoS/NLoS path has been separated perfectly. Therefore, the maximum likelihood (ML) estimator in [36] can works in LoS environment and NLoS environment. The simulation results are shown in Fig. 4.10(a). In case of single antenna

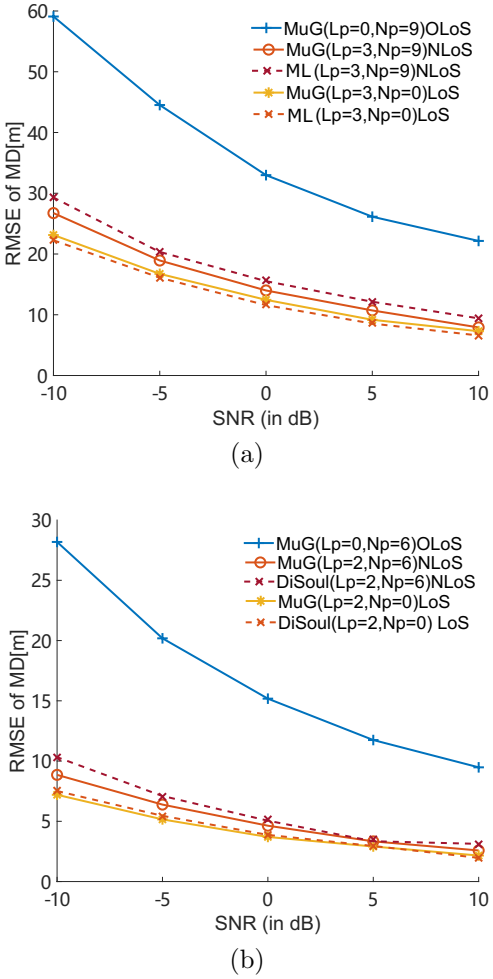


Figure 4.10: Simulation results of RMSE versus the signal-to-noise ratio (SNR) under different system configuration and environment. (a) In the first case studies, 3 single antenna APs ($N_R = 1$) are employed for providing ToF information. (b) In the second case studies, 2 multi-antenna APs ($N_R = 12$) are employed for providing ToF and DoA information.

system, the ML estimator takes a set of ranges (related to the ToFs and the speed of light) as inputs and outputs the estimation. The ToFs are obtained by the ANM (2.17) in Chapter 2. Note that the ML estimator requires LoS path for localisation. Therefore, it only works in LoS environment and NLoS environment. Based on the Fig. 4.10(a) the ML estimator works well in LoS environment. In NLoS environment, the performance of ML estimator is slightly degraded since complete mitigation of the NLoS bias may be impractical.

In case of multi-antenna system, the performance of MuG is compared with direct localisation method DiSouL [15]. The DiSouL takes the DoA information to separate LoS paths

from multipath and then localise MD by ToFs of LoS paths. Therefore, it only works in LoS environment and NLoS environment. Based on the Fig. 4.10(b), the DiSouL works well in LoS environment. In NLoS environment, the performance of DiSouL is slightly degraded since the mistaken LoS/NLoS separation cannot be avoided completely.

4.5 Conclusion

This chapter has shown a multipath-expolited and grid-free (MuG) localisation scheme. Concretely, it is an end-to-end localisation scheme to exploit information carried by multipath. We address the localisation problem by three key elements. First, we introduce the virtual scatter to LoS path so that there is no need to separate LoS/NLoS path, and both LoS and NLoS paths are uniformly formulated into a forward model. Then we formulate the localisation problem as a general sparse recovery problem which is different from previous works. Finally, the modern grid-free technique is adapted to solve the sparse inverse problem. The simulation results show that the proposed MuG method is capable to handle different system configurations and all three propagation environments. We also demonstrate superior performance of multipath exploitation strategy in case of NLoS paths dominating in the environment. Our approach addresses long-standing issues not completely solved in the literature, achieves good localisation accuracy and guarantees a global convergence.

The proposed MuG method combines convex and non-convex optimisation techniques. The convex optimisation guarantees a global convergence. In the final step, the non-convex optimisation is a subroutine that takes estimations from the convex optimisation and attempts to use gradient information to reduce the residual function. Therefore, this non-convex optimisation does not change the convergence guarantees. On the other hand, it provides a significantly sparser solutions by allowing the unknown parameters to move continuously within the parameter space. Our numerical case studies demonstrates that proposed MuG method achieves state-of-the-art results comparing with theoretical

CRB. In the numerical case studies, we generate synthetic data by adopting a commonly assumed setup in the literature [1–4, 10–12, 14, 15, 27–31]. It is noteworthy that this common setup is a simplification of actual systems for the purpose of highlighting the idea without being drowned into great technical details. To tackle the practical problems, the real data will be included in the future work.

Chapter 5

Conclusion

In this chapter, we summarise results of the present work and gives potential future research directions to the localisation of MD.

5.1 Summary of Thesis

In Chapter 1, we settle the main stream of the thesis. The localisation is challenge in presence of multipath. Even though it has been studied in the literature, there are still some unresolved open issues. Then, we discuss the typical methods to deal with localisation problem in different prior knowledge and assumptions. Finally, we summarise challenges of localisation in multipath environments.

In Chapter 2, we discuss the basic idea of super-resolved localisation which is based on super-resolution technique. The super-resolution technique shows that line spectral estimation problem can be formulated as SDP with finite many variables and constraints. However, this framework cannot be directly applied to our localisation problems, because the received signal is not the superposition of simple exponential form. In order to avoid data association and error propagation, our model is direct mapping from unknown parameters to received signal, the atom is formulated with irregular trigonometric

polynomial form. Efficient algorithms for solving it are not known in full generality.

In Chapter 3, we present a novel super-resolved localisation method that estimate multiple MDs and perform self-calibration to correct possible errors simultaneously. The array directional error implies deviations in the orientation of the array placement which results in additional angular bias. We first formulate a direct localisation model with self-calibration on these errors and based on this model, group sparsity is exploited to improve the performance. Then, we propose a method called GSE-SC to solve this non-convex optimisation problem. Particularly, MD is added to the support set one by one, while all the unknowns are improved locally by a two-loop alternating gradient descent to bring the cost function down. In addition, we analyse the ambiguity problem and provide a sufficient condition that guarantees no ambiguity. Simulation results demonstrate that GSE-SC method outperforms existing methods including MF, GCS and ADCG methods.

In Chapter 4, we provide geometric insight to demonstrate the usage of both LoS and NLoS path for localising MD. In the literature, the NLoS path is commonly treated as a main issue to degraded localisation performance. A MuG localisation scheme is proposed. Concretely, it is an end-to-end localisation scheme to exploit information carried by multipath. We address the localisation problem by three key elements. We first study the geometry of the localisation in multipath environment. We identify and analyse scenarios that localisation requires exploitation of NLoS paths and incorporating scatters' unknown locations. Motivated by the geometric insight, we then treat the LoS paths as NLoS paths by the usage of virtual scatters so that hard-decision, separation, and removal of NLoS paths are avoided. Then we formulate the localisation problem as a general sparse recovery problem which is different from previous works. Finally, the modern grid-free technique is adapted to solve the sparse inverse problem. The simulation results show that the proposed MuG scheme is capable to handle different system configurations and all three propagation environments. We also demonstrate superior performance of multipath exploitation strategy in case of NLoS paths dominating in the environment. Our approach addresses long-standing issues not completely solved in the literature, and

achieves good localisation accuracy.

5.2 Future Work

For localisation problem in multipath environments, the NLoS bias is one of the main issues to degrade estimation performance. In order to enhance localisation performance, the common strategy of existing method is still dealing with NLoS bias. Many methods have been proposed in LoS/NLoS separation. Nonetheless, these methods may not satisfy for industrial application due to computational cost and positioning accuracy. Hence, a new, efficient and simple method is required. The usage of machine learning techniques to train a classifier for LoS and NLoS path could be an alternative research direction. In particular deep learning, is known to be very effective when the underlying model is hard to approximate or unknown [120]. On the other hand, many methods assume that LoS path should not be blocked in order to apply trilateration. This loss of LoS can be addressed using a distributed antenna array, since the probability that the LoS link with the majority of antennas is broken is much smaller in this case. Therefore, distributed localisation system should be studied and evaluated if they do in fact improve the robustness of the positioning system in LoS paths.

While the LoS path is well studied and multiple methods are reported in the literature (e.g., angle- and time-of-arrival based predictions and trilateration methods) suitable solutions for the NLoS path are still open for research. Therefore, a potential future research direction is to address the more challenging NLoS case which covers a huge variety of different scenarios. In Chapter 3, we propose a method to calibrate array directional error during localising MDs. As for another error i.e., NLoS bias, it can be formulated into array directional error for calibrating instead of discarded. But this formulation may lead to underdetermined problem, in which there might be infinite number of solutions and finding its solution is computationally difficult. In Chapter 4, geometrical insight of NLoS path localisation is provided which can be viewed as a guidance for future work.

Appendices

Appendix A

Appendices of Chapter 3

A.1 Gradients w.r.t. $\boldsymbol{\Gamma}$

In the t-th step, we use $\boldsymbol{c}_{i,j}$ and $\mathcal{C}^t$ to denote atoms $\boldsymbol{a}_j(\boldsymbol{m}_i, \delta_j)$ and the cost function $\boldsymbol{C}(\boldsymbol{M}, \boldsymbol{\Gamma}, \boldsymbol{\delta})$ as

$$\boldsymbol{a}_{i,j} = \left[e^{\sqrt{-1}2\pi\frac{d_j}{\lambda}\sin\theta_{i,j}\cdot 0}, \ldots, e^{\sqrt{-1}2\pi\frac{d_j}{\lambda}\sin\theta_{i,j}\cdot(N_m-1)}\right]^T, \tag{A.1}$$

$$\mathcal{C}^t = \frac{1}{2}\sum_{j=1}^{J}\left\|\boldsymbol{r}_j - \sum_{i=1}^{t}\gamma_{i,j}\boldsymbol{a}_{i,j}\right\|_2^2 + \mu\sum_{i=1}^{t}\|\boldsymbol{\gamma}_i\|_2, \tag{A.2}$$

where

$$\sin\theta_{i,j} = \frac{\sin\alpha_j(m_i^x - AP_j^x) - \cos\alpha_j(m_i^y - AP_j^y)}{\sqrt{(m_i^x - AP_j^x)^2 + (m_i^y - AP_j^y)^2}}, \tag{A.3}$$

$$\|\boldsymbol{\gamma}_i\|_2 = \sqrt{\sum_{m=1}^{M}|\gamma_{i,j}|^2}, \tag{A.4}$$

$\boldsymbol{\gamma}_i = [\gamma_{i,1}, ..., \gamma_{i,J}]^T$ and $\alpha_j = \alpha_j' - \delta_j$. Denote that $\boldsymbol{n}_m = [0, 1, \ldots, N_m - 1]^T$. Then, the $\partial\mathcal{C}^t/\partial\gamma_{i,j}^*$ can be expressed as

$$\begin{aligned}\frac{\partial \mathcal{C}^t}{\partial \gamma_{i,j}^*} &= \frac{\partial \frac{1}{2}\sum_{j=1}^{J}\left\|\boldsymbol{r}_j - \sum_{i=1}^{t}\gamma_{i,j}\boldsymbol{a}_{i,j}\right\|_2^2}{\partial \gamma_{i,j}^*} + \frac{\partial \mu \sum_{i=1}^{t}\|\boldsymbol{\gamma}_i\|_2}{\partial \gamma_{i,j}^*} \\ &= \frac{1}{2}\frac{\partial\left\|\boldsymbol{r}_j - \sum_{i=1}^{t}\gamma_{i,j}\boldsymbol{a}_{i,j}\right\|_2^2}{\partial \gamma_{i,j}^*} + \mu\frac{\partial\|\boldsymbol{\gamma}_i\|_2}{\partial \gamma_{i,j}^*}. \end{aligned} \tag{A.5}$$

The first term in (A.5) implies

$$\begin{aligned}&\frac{1}{2}\frac{\partial\left\|\boldsymbol{r}_j - \sum_{i=1}^{t}\gamma_{i,j}\boldsymbol{a}_{i,j}\right\|_2^2}{\partial \gamma_{i,j}^*} \\ &= \frac{1}{2}\frac{\partial\left(\boldsymbol{r}_j^H - \sum_{i=1}^{t}\gamma_{i,j}^*\boldsymbol{a}_{i,j}^H\right)\left(\boldsymbol{r}_j - \sum_{i=1}^{t}\gamma_{i,j}\boldsymbol{a}_{i,j}\right)}{\partial \gamma_{i,j}^*} \\ &= -\frac{1}{2}\frac{\partial\gamma_{i,j}^*\boldsymbol{a}_{i,j}^H\left(\boldsymbol{r}_j - \sum_{i=1}^{t}\gamma_{i,j}\boldsymbol{a}_{i,j}\right)}{\partial \gamma_{i,j}^*} \\ &= -\frac{1}{2}\boldsymbol{a}_{i,j}^H\left(\boldsymbol{r}_j - \sum_{i=1}^{t}\gamma_{i,j}\boldsymbol{a}_{i,j}\right). \end{aligned} \tag{A.6}$$

Through extending the second term in (A.5), we have

$$\mu\frac{\partial\|\boldsymbol{\gamma}_i\|_2}{\partial \gamma_{i,j}^*} = \mu\frac{1}{2\|\boldsymbol{\gamma}_i\|_2}\frac{\partial\gamma_{i,j}^*\gamma_{i,j}}{\partial \gamma_{i,j}^*} = \mu\frac{\gamma_{i,j}}{2\|\boldsymbol{\gamma}_i\|_2}. \tag{A.7}$$

Combining (A.6) and (A.7) yields

$$\frac{\partial \mathcal{C}^t}{\partial \gamma_{i,j}^*} = -\frac{1}{2}\boldsymbol{a}_{i,j}^H\left(\boldsymbol{r}_j - \sum_{i=1}^{t}\gamma_{i,j}\boldsymbol{a}_{i,j}\right) + \mu\frac{\gamma_{i,j}}{2\|\boldsymbol{\gamma}_i\|_2}. \tag{A.8}$$

A.2 Gradients w.r.t. $\boldsymbol{M}$

Here we calculate $\partial \mathcal{C}^t / \partial \boldsymbol{m}_i$. To this aim, we denote by

$$\mathcal{C}_j^t := \left\| \boldsymbol{r}_j - \sum_{i=1}^{t} \gamma_{i,j} \boldsymbol{a}_{i,j} \right\|_2^2. \tag{A.9}$$

Hence, we have

$$\frac{\partial \mathcal{C}^t}{\partial \boldsymbol{m}_i} = \frac{1}{2} \sum_{j=1}^{J} \frac{\partial \mathcal{C}_j^t}{\partial \boldsymbol{m}_i}. \tag{A.10}$$

Particularly, we denote $\boldsymbol{v}_j := \boldsymbol{r}_j - \sum_{i=1}^{t} \gamma_{i,j} \boldsymbol{a}_{i,j}$. Then we have $\mathcal{C}_j^t = \boldsymbol{v}_j^H \boldsymbol{v}_j$ and the partial derivative becomes

$$\begin{aligned} \frac{\partial \mathcal{C}_m^t}{\partial \boldsymbol{m}_i} = \frac{\partial \boldsymbol{v}_j^H \boldsymbol{v}_j}{\partial \boldsymbol{m}_i} &= \boldsymbol{v}_j^H \frac{\partial \boldsymbol{v}_j}{\partial \boldsymbol{m}_i} + \frac{\partial \boldsymbol{v}_j^H}{\partial \boldsymbol{m}_i} \boldsymbol{v}_j \\ &= \boldsymbol{v}_j^H \frac{\partial \boldsymbol{v}_j}{\partial \boldsymbol{m}_i} + \left(\frac{\partial \boldsymbol{v}_j}{\partial \boldsymbol{m}_i} \right)^H \boldsymbol{v}_j \\ &= 2\mathrm{Re} \left(\boldsymbol{v}_j^H \frac{\partial \boldsymbol{v}_j}{\partial \boldsymbol{m}_i} \right), \end{aligned} \tag{A.11}$$

where

$$\frac{\partial \boldsymbol{v}_j}{\partial \boldsymbol{m}_i} = \frac{\partial \left(\boldsymbol{r}_j - \sum_{i=1}^{t} \gamma_{i,j} \boldsymbol{a}_{i,j} \right)}{\partial \boldsymbol{m}_i} = -\gamma_{i,j} \frac{\partial \boldsymbol{a}_{i,j}}{\partial \boldsymbol{m}_i}. \tag{A.12}$$

In the 2D plane, we have $\boldsymbol{m}_i = [m_i^x, m_i^y]^{\mathrm{T}}$. Based on (A.1) and (A.3), here we calculate the gradients w.r.t m_i^x and m_i^y, respectively, given by

$$\begin{aligned} \frac{\partial \left[\boldsymbol{a}_{i,j} \right]_n}{\partial m_i^x} &= \left[\boldsymbol{a}_{i,j} \right]_n \sqrt{-1} 2\pi \frac{d_j}{\lambda} n \frac{\partial \sin\theta_{i,j}}{\partial m_i^x} \\ &= \left[\boldsymbol{a}_{i,j} \right]_n \sqrt{-1} 2\pi \frac{d_j}{\lambda} n \frac{1}{R_{sp}} \Big(\sin\alpha_j - (m_i^x - AP_j^x) \\ &\quad \cdot \frac{\sin\alpha_j (m_i^x - AP_j^x) - \cos\alpha_j (m_i^y - AP_j^y)}{R_{sp}^2} \Big) \\ &= \left[\boldsymbol{a}_{i,j} \right]_n \sqrt{-1} 2\pi \frac{d_j}{\lambda} n (m_i^y - AP_j^y) R_{th} / R_{sp}^2, \end{aligned} \tag{A.13}$$

where R_{th} and R_{sp} are defined as

$$R_{th} := \frac{\sin\alpha_j(m_i^y - AP_j^y) + \cos\alpha_j(m_i^x - AP_j^x)}{R_{sp}}, \tag{A.14}$$

$$R_{sp} := \sqrt{(m_i^x - AP_j^x)^2 + (m_i^y - AP_j^y)^2}. \tag{A.15}$$

Then, by substituting (A.13) into (A.12), substituting (A.12) into (A.11), and substituting (A.11) into (A.10), we have

$$\frac{\partial \mathcal{C}^t}{\partial m_i^x} = \sum_{j=1}^{J} \mathrm{Re}\Bigg(-\left(\boldsymbol{r}_j^H - \sum_{i=1}^{t} \gamma_{i,j}^* \boldsymbol{a}_{i,j}^H\right) \gamma_{i,j} \left(\boldsymbol{a}_{i,j} \odot \boldsymbol{n}_m\right) \cdot \sqrt{-1} 2\pi \frac{d_j}{\lambda} R_{th}/R_{sp}^2 \cdot (m_i^y - AP_j^y)\Bigg). \tag{A.16}$$

Similarly, we calculate the corresponding gradient w.r.t. m_i^y, given by

$$\frac{\partial \mathcal{C}^t}{\partial m_i^y} = \sum_{m=1}^{M} \mathrm{Re}\Bigg(-\left(\boldsymbol{r}_j^H - \sum_{i=1}^{t} \gamma_{i,j}^* \boldsymbol{a}_{i,j}^H\right) \gamma_{i,j} \left(\boldsymbol{a}_{i,j} \odot \boldsymbol{n}_m\right) \cdot \sqrt{-1} 2\pi \frac{d_j}{\lambda} R_{th}/R_{sp}^2 \cdot (AP_j^x - m_i^x)\Bigg), \tag{A.17}$$

Here, we are ready to provide the partial gradient w.r.t. $\boldsymbol{m}_i$ as

$$\frac{\partial \mathcal{C}^t}{\partial \boldsymbol{m}_i} = \sum_{m=1}^{M} \mathrm{Re}\Bigg(-\left(\boldsymbol{r}_j^H - \sum_{i=1}^{t} \gamma_{i,j}^* \boldsymbol{a}_{i,j}^H\right) \gamma_{i,j} \left(\boldsymbol{a}_{i,j} \odot \boldsymbol{n}_m\right) \cdot \sqrt{-1} 2\pi \frac{d_j}{\lambda} R_{th}/R_{sp}^2 \cdot \left[m_i^y - AP_j^y, AP_j^x - m_i^x\right]^T\Bigg). \tag{A.18}$$

A.3 Gradients w.r.t. $\boldsymbol{\delta}$

Here we calculate $\partial \mathcal{C}^t / \partial \delta_j$. Similar with the derivation from (A.9) to (A.12), we directly have

$$\begin{aligned}\frac{\partial \mathcal{C}^t}{\partial \delta_j} &= \mathrm{Re}\left(- \left(\boldsymbol{r}_j^H - \sum_{i=1}^{t} \gamma_{i,j}^* \boldsymbol{a}_{i,j}^H \right) \frac{\partial \sum_{i=1}^{t} \gamma_{i,j} \boldsymbol{a}_{i,j}}{\partial \delta_j} \right) \\ &= \sum_{i=1}^{t} \mathrm{Re}\left(- \left(\boldsymbol{r}_j^H - \sum_{i=1}^{t} \gamma_{i,j}^* \boldsymbol{a}_{i,j}^H \right) \gamma_{i,j} \frac{\partial \boldsymbol{a}_{i,j}}{\partial \delta_j} \right). \end{aligned} \tag{A.19}$$

Based on $\alpha_j = \alpha_j' - \delta_j$ and (A.3), we then calculate

$$\begin{aligned}\frac{\partial [\boldsymbol{a}_{i,j}]_n}{\partial \delta_j} &= [\boldsymbol{a}_{i,j}]_n \sqrt{-1} 2\pi \frac{d_j}{\lambda} n \frac{\partial \sin\theta_{i,j}}{\partial \delta_j} \\ &= -[\boldsymbol{a}_{i,j}]_n \sqrt{-1} 2\pi \frac{d_j}{\lambda} n \\ &\quad \cdot \frac{\cos\alpha_j (m_i^x - A P_j^x) + \sin\alpha_j (m_i^y - A P_j^y)}{\sqrt{(m_i^x - A P_j^x)^2 + (m_i^y - A P_j^y)^2}} \\ &= -[\boldsymbol{a}_{i,j}]_n \sqrt{-1} 2\pi \frac{d_j}{\lambda} n R_{th}. \end{aligned} \tag{A.20}$$

Combining (A.19) and (A.20) yields

$$\begin{aligned}\frac{\partial \mathcal{C}^t}{\partial \delta_j} = \sum_{i=1}^{t} \mathrm{Re}\Bigg(& \left(\boldsymbol{r}_j^H - \sum_{i=1}^{t} \gamma_{i,j}^* \boldsymbol{a}_{i,j}^H \right) \gamma_{i,j} (\boldsymbol{a}_{i,j} \odot \boldsymbol{n}_m) \\ & \cdot \sqrt{-1} 2\pi \frac{d_j}{\lambda} R_{th} \Bigg). \end{aligned} \tag{A.21}$$

Appendix B

Appendices of Chapter 4

B.1 Cramér-Rao bound

As for (q, w)-th entry of Fisher information matrix (FIM) $\boldsymbol{F}(\boldsymbol{p})$, the general expression can be written as

$$\boldsymbol{F}\left(\boldsymbol{p}\right) = \mathrm{E}\left(\frac{\partial \ln L(\boldsymbol{R};\boldsymbol{p})}{\partial \boldsymbol{p}} \frac{\partial \ln L^{\mathrm{H}}(\boldsymbol{R};\boldsymbol{p})}{\partial \boldsymbol{p}}\right), \tag{B.1}$$

The partial derivatives of the log-likelihood function $\ln L(\boldsymbol{R};\boldsymbol{p})$ with respect to the q-th unknown parameter of $\boldsymbol{p}$ is given by

$$\begin{aligned} &\frac{\partial \ln L(\boldsymbol{R};\boldsymbol{p})}{\partial p_q} \\ &= \sum_j \sum_n \sum_{n_r} -\frac{1}{\sigma^2}\left(\frac{\partial \left(r_j(n,n_r) - f_j(\boldsymbol{p},n,n_r)\right)^*}{\partial p_q} v_j(n,n_r)\right) \\ &\quad - \frac{1}{\sigma^2}\left(v_j^*(n,n_r)\frac{\partial \left(r_j(n,n_r) - f_j(\boldsymbol{p},n,n_r)\right)}{\partial p_q}\right) \end{aligned} \tag{B.2}$$

$$= \frac{1}{\sigma^2}\sum_j \sum_n \sum_{n_r} \frac{\partial f_j^*(\boldsymbol{p},n,n_r)}{\partial p_q} v_j(n,n_r) + v_j^*(n,n_r)\frac{\partial f_j(\boldsymbol{p},n,n_r)}{\partial p_q}. \tag{B.3}$$

By substituting (B.3) into (B.1), we have

$$\boldsymbol{F}\left(\boldsymbol{p}\right)_{q,w} = \mathrm{E}\left(\frac{\partial \ln L(\boldsymbol{R};\boldsymbol{p})}{\partial p_q}\frac{\partial \ln L(\boldsymbol{R};\boldsymbol{p})}{\partial p_w}\right) \tag{B.4}$$

$$= \frac{1}{\sigma^4}\sum_j\sum_n\sum_{n_r}\mathrm{E}\left(\left(\frac{\partial f_j^*(\boldsymbol{p},n,n_r)}{\partial p_q}v_j(n,n_r) + v_j^*(n,n_r)\frac{\partial f_j(\boldsymbol{p},n,n_r)}{\partial p_q}\right)\cdot\left(\frac{\partial f_j^*(\boldsymbol{p},n,n_r)}{\partial p_w}v_j(n,n_r) + v_j^*(n,n_r)\frac{\partial f_j(\boldsymbol{p},n,n_r)}{\partial p_w}\right)\right) \tag{B.5}$$

$$\overset{(a)}{=} \frac{1}{\sigma^2}\sum_j\sum_n\sum_{n_r}\left(\frac{\partial f_j^*(\boldsymbol{p},n,n_r)}{\partial p_q}\frac{\partial f_j(\boldsymbol{p},n,n_r)}{\partial p_w} + \frac{\partial f_j(\boldsymbol{p},n,n_r)}{\partial p_q}\frac{\partial f_j^*(\boldsymbol{p},n,n_r)}{\partial p_w}\right) \tag{B.6}$$

$$= \frac{2}{\sigma^2}\mathrm{Re}\left(\sum_j\sum_n\sum_{n_r}\frac{\partial f_j(\boldsymbol{p},n,n_r)}{\partial p_q}\frac{\partial f_j^*(\boldsymbol{p},n,n_r)}{\partial p_w}\right), \tag{B.7}$$

where (a) is following the fact that the complex noise samples are i.i.d with the variance σ^2 and $\mathrm{E}\left(v_j v_j\right) = \mathrm{E}\left(v_j^* v_j^*\right) = 0$.

In order to obtain entries of FIM $\boldsymbol{F}(\boldsymbol{p})$, we need to compute derivative of observation $f_j(\boldsymbol{p},n,n_r)$ with respect to the q-th unknown parameter of $\boldsymbol{p}$,

$$\frac{\partial f_j(\boldsymbol{p},n,n_r)}{\partial p_q} = \frac{\partial \sum_{k=1}^K b_j\left(\boldsymbol{m},\boldsymbol{s}_k,n,n_r\right)\gamma_{j,k}}{\partial p_q}. \tag{B.8}$$

For $p_q \in \{m^x, m^y, s_1^x, s_1^y, ..., s_K^x, s_K^y\}$, the corresponding first-order derivative (B.8) is

$$\begin{aligned}&\frac{\partial \sum_{k=1}^K b_j\left(\boldsymbol{m},\boldsymbol{s}_k,n,n_r\right)\gamma_{j,k}}{\partial p_q}\\ &= \sum_k^K \frac{\partial \tau_j(\boldsymbol{m},\boldsymbol{s}_k)}{\partial p_q} b_j\left(\boldsymbol{m},\boldsymbol{s}_k,n,n_r\right)(-i2\pi f n)\gamma_{j,k}\\ &+ \sum_k^K \frac{\partial \theta_j(\boldsymbol{m},\boldsymbol{s}_k)}{\partial p_q} b_j(\boldsymbol{m},\boldsymbol{s}_k,n,n_R)(i\frac{2\pi}{\lambda}L\cos(\theta_j)(n_r-1))\gamma_{j,k}.\end{aligned} \tag{B.9}$$

The detailed computations of $\frac{\partial \tau_j(\boldsymbol{m},\boldsymbol{s}_k)}{\partial p_q}$ and $\frac{\partial \theta_j(\boldsymbol{m},\boldsymbol{s}_k)}{\partial p_q}$ are similar with (4.78) - (4.86).

For $p_q = \mathrm{Re}(\gamma_{j,k})$, the corresponding first-order derivative (B.8) is

$$\frac{\partial \sum_{k=1}^K b_j(\boldsymbol{m},\boldsymbol{s}_k,n,n_R)\gamma_{j,k}}{\partial p_q} = b_j(\boldsymbol{m},\boldsymbol{s}_k,n,n_r) \tag{B.10}$$

For $p_q = \mathrm{Im}(\gamma_{j,k})$, the corresponding first-order derivative (B.8) is

$$\frac{\partial \sum_{k=1}^{K} b_j(\boldsymbol{m}, \boldsymbol{s}_k, n, n_R)\gamma_{j,k}}{\partial p_q} = b_j^*(\boldsymbol{m}, \boldsymbol{s}_k, n, n_r). \tag{B.11}$$

Finally, the CRB can be obtained as the inverse of the FIM $\boldsymbol{F}(\boldsymbol{p})$.

Bibliography

[1] V. Y. Zhang and A. K.-s. Wong, "Combined AOA and TOA NLOS localization with nonlinear programming in severe multipath environments," in *2009 IEEE Wireless Communications and Networking Conference*. IEEE, 2009, pp. 1–6.

[2] C. K. Seow and S. Y. Tan, "Non-line-of-sight localization in multipath environments," *IEEE Transactions on Mobile Computing*, vol. 7, no. 5, pp. 647–660, 2008.

[3] A. Shahmansoori, G. E. Garcia, G. Destino, G. Seco-Granados, and H. Wymeersch, "Position and orientation estimation through millimeter-wave MIMO in 5g systems," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1822–1835, March 2018.

[4] H. Miao, K. Yu, and M. J. Juntti, "Positioning for NLOS propagation: Algorithm derivations and Cramer–Rao bounds," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 5, pp. 2568–2580, 2007.

[5] K. Papakonstantinou and D. Slock, "Direct location estimation using single-bounce NLOS time-varying channel models," in *2008 IEEE 68th Vehicular Technology Conference*, Sept 2008, pp. 1–5.

[6] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[7] C. Ververidis and G. Polyzos, "Mobile marketing using a location based service," in *Proceedings of the First International Conference on Mobile Business, Athens, Greece*, 2002, pp. 1–12.

[8] Z. Li, K. Liu, Y. Zhao, and Y. Ma, "MaPIT: an enhanced pending interest table for NDN with mapping bloom filter," *IEEE Communications Letters*, vol. 18, no. 11, pp. 1915–1918, 2014.

[9] Z. Li, L. Song, and H. Shi, "Approaching the capacity of K-user MIMO interference channel with interference counteraction scheme," *Ad Hoc Networks*, vol. 58, pp. 286–291, 2017.

[10] J. Riba and A. Urruela, "A non-line-of-sight mitigation technique based on ML-detection," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2. IEEE, 2004, pp. ii–153.

[11] L. Cong and W. Zhuang, "Non-line-of-sight error mitigation in TDOA mobile location," in *GLOBECOM'01. IEEE Global Telecommunications Conference (Cat. No. 01CH37270)*, vol. 1. IEEE, 2001, pp. 680–684.

[12] M. P. Wylie and J. Holtzman, "The non-line of sight problem in mobile location estimation," in *Proceedings of ICUPC - 5th International Conference on Universal Personal Communications*, vol. 2, Oct 1996, pp. 827–831 vol.2.

[13] Y. Wang, K. Gu, Y. Wu, W. Dai, and Y. Shen, "Nlos effect mitigation via spatial geometry exploitation in cooperative localization," *IEEE Transactions on Wireless Communications*, 2020.

[14] Y.-T. Chan, W.-Y. Tsui, H.-C. So, and P.-c. Ching, "Time-of-arrival based localization under NLOS conditions," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 1, pp. 17–24, 2006.

[15] N. Garcia, H. Wymeersch, E. G. Larsson, A. M. Haimovich, and M. Coulon, "Direct localization for massive MIMO," *IEEE Transactions on Signal Processing*, vol. 65, no. 10, pp. 2475–2487, 2017.

[16] Y. Qi, H. Kobayashi, and H. Suda, "Analysis of wireless geolocation in a non-line-of-sight environment," *IEEE Transactions on Wireless Communications*, vol. 5, no. 3, pp. 672–681, 2006.

[17] S. Gezici, Z. Tian, G. B. Giannakis, H. Kobayashi, A. F. Molisch, H. V. Poor, and Z. Sahinoglu, "Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 70–84, 2005.

[18] S. Venkatesh and R. M. Buehrer, "A linear programming approach to nlos error mitigation in sensor networks," in *Proceedings of the 5th International Conference on Information Processing in Sensor Networks*, 2006, pp. 301–308.

[19] S. Venkatesh and R. M. Buehrer, "Nlos mitigation using linear programming in ultrawideband location-aware networks," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 5, pp. 3182–3198, 2007.

[20] R. M. Vaghefi, J. Schloemann, and R. M. Buehrer, "Nlos mitigation in TOA-based localization using semidefinite programming," in *2013 10th Workshop on Positioning, Navigation and Communication (WPNC)*, 2013, pp. 1–6.

[21] H. Zhang, X. Qi, Q. Wei, and L. Liu, "TOA nlos mitigation cooperative localisation algorithm based on topological unit," *IET Signal Processing*, vol. 14, no. 10, pp. 765–773, 2021.

[22] Y. Liu, F. Guo, L. Yang, and W. Jiang, "An improved algebraic solution for TDOA localization with sensor position errors," *IEEE Communications Letters*, vol. 19, no. 12, pp. 2218–2221, 2015.

[23] V. Bianchi, P. Ciampolini, and I. De Munari, "Rssi-based indoor localization and identification for zigbee wireless sensor networks in smart homes," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 2, pp. 566–575, 2018.

[24] W. Wang and Q. Zhu, "Rss-based monte carlo localisation for mobile sensor networks," *IET Communications*, vol. 2, no. 5, pp. 673–681, 2008.

[25] F. Penna and D. Cabric, "Bounds and tradeoffs for cooperative doa-only localization of primary users," in *2011 IEEE Global Telecommunications Conference-GLOBECOM 2011.* IEEE, 2011, pp. 1–5.

[26] Q. Cheng, M. Munoz, A. Alomainy, and Y. Hao, "Compressive sensing applied to fingerprint-based localisation," in *2014 IEEE MTT-S International Microwave Workshop Series on RF and Wireless Technologies for Biomedical and Healthcare Applications (IMWS-Bio2014).* IEEE, 2014, pp. 1–3.

[27] O. Bialer, D. Raphaeli, and A. J. Weiss, "Maximum-likelihood direct position estimation in dense multipath," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 2069–2079, 2012.

[28] F. Yin, C. Fritsche, F. Gustafsson, and A. M. Zoubir, "EM-and JMAP-ML based joint estimation algorithms for robust wireless geolocation in mixed LOS/NLOS environments," *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 168–182, 2013.

[29] Y. Wang, K. Gu, Y. Wu, W. Dai, and Y. Shen, "Exploiting NLOS bias correlation in cooperative localization," in *2019 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2019, pp. 1–5.

[30] Y. Wang, Y. Wu, and Y. Shen, "On the resolution limits for MIMO localization," *IEEE Communications Letters*, vol. 23, no. 3, pp. 462–465, 2019.

[31] ——, "Joint spatiotemporal multipath mitigation in large-scale array localization," *IEEE Transactions on Signal Processing*, vol. 67, no. 3, pp. 783–797, 2019.

[32] V. Dizdarevic and K. Witrisal, "On impact of topology and cost function on lse position determination in wireless networks," in *Proc. Workshop on Positioning, Navigation, and Commun.(WPNC)*. Citeseer, 2006, pp. 129–138.

[33] J. Borras, P. Hatrack, and N. B. Mandayam, "Decision theoretic framework for NLOS identification," in *VTC'98. 48th IEEE Vehicular Technology Conference. Pathway to Global Wireless Revolution (Cat. No. 98CH36151)*, vol. 2. IEEE, 1998, pp. 1583–1587.

[34] P.-C. Chen, "A non-line-of-sight error mitigation algorithm in location estimation," in *WCNC. 1999 IEEE Wireless Communications and Networking Conference (Cat. No. 99TH8466)*, vol. 1. IEEE, 1999, pp. 316–320.

[35] S. Gezici and Z. Sahinoglu, "UWB geolocation techniques for ieee 802.15. 4a personal area networks," *MERL Technical report*, 2004.

[36] K. Yu and Y. J. Guo, "Improved positioning algorithms for nonline-of-sight environments," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pp. 2342–2353, 2008.

[37] D. B. Jourdan, D. Dardari, and M. Z. Win, "Position error bound for uwb localization in dense cluttered environments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 2, pp. 613–628, 2008.

[38] B. Denis, J.-B. Pierrot, and C. Abou-Rjeily, "Joint distributed synchronization and positioning in uwb ad hoc networks using TOA," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 4, pp. 1896–1911, 2006.

[39] J. Zheng, Y.-e. Sun, Y. Huang, Y. Wang, and Y. Xiao, "Error analysis of range–based localisation algorithms in wireless sensor networks," *International Journal of Sensor Networks*, vol. 12, no. 2, pp. 78–88, 2012.

[40] S. Zhong, B. Zhang, J. Li, and Q. Li, "High-precision localisation algorithm in wireless sensor networks," *International Journal of Computer Applications in Technology*, vol. 41, no. 1-2, pp. 150–155, 2011.

[41] Y.-T. Chan, H. Y. C. Hang, and P.-c. Ching, "Exact and approximate maximum likelihood localization algorithms," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 1, pp. 10–16, 2006.

[42] C. Chang and A. Sahai, "Estimation bounds for localization," in *2004 First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004.* IEEE, 2004, pp. 415–424.

[43] W. Kim, J. G. Lee, and G.-I. Jee, "The interior-point method for an optimal treatment of bias in trilateration location," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 4, pp. 1291–1301, 2006.

[44] J. Caffery and G. L. Stuber, "Subscriber location in cdma cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 47, no. 2, pp. 406–416, 1998.

[45] R. M. Vaghefi and R. M. Buehrer, "Cooperative sensor localization with nlos mitigation using semidefinite programming," in *2012 9th Workshop on Positioning, Navigation and Communication*, 2012, pp. 13–18.

[46] H. V. Poor, *An introduction to Signal Detection and Estimation.* Springer Science & Business Media, 2013.

[47] N. Garcia, A. M. Haimovich, J. A. Dabin, M. Coulon, and M. Lops, "Direct localization of emitters using widely spaced sensors in multipath environments," in *2014 48th Asilomar Conference on Signals, Systems and Computers.* IEEE, 2014, pp. 695–700.

[48] H. Liu, W. Dai, and Y. Shen, "Super-resolved localisation without identifying LoS/NLoS paths," *arXiv preprint arXiv:1910.12662*, 2019.

[49] I. Guvenc and C. C. Chong, "A survey on TOA based wireless localization and nlos mitigation techniques," *IEEE Communications Surveys Tutorials*, vol. 11, no. 3, pp. 107–124, 2009.

[50] E. J. Candès and C. Fernandez-Granda, "Super-resolution from noisy data," *Journal of Fourier Analysis and Applications*, vol. 19, no. 6, pp. 1229–1254, 2013.

[51] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[52] H. Liu, W. Dai, and Y. Shen, "MuG: A multipath-exploited and grid-free localisation method," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 4665–4669.

[53] H. Greenspan, "Super-resolution in medical imaging," *The Computer Journal*, vol. 52, no. 1, pp. 43–63, 2009.

[54] K. G. Puschmann and F. Kneer, "On super-resolution in astronomical imaging," *Astronomy and Astrophysics*, vol. 436, no. 1, pp. 373–378, 2005.

[55] C. W. Mccutchen, "Superresolution in microscopy and the abbe resolution limit," *JOSA*, vol. 57, no. 10, pp. 1190–1192, 1967.

[56] P. L. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets strang–fix," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1741–1757, 2007.

[57] M. M. Abo-Zahhad, A. I. Hussein, A. M. Mohamed *et al.*, "Compressive sensing algorithms for signal processing applications: A survey," *International Journal of Communications, Network and System Sciences*, vol. 8, no. 06, p. 197, 2015.

[58] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[59] M. D. Plumbley, T. Blumensath, L. Daudet, R. Gribonval, and M. E. Davies, "Sparse representations in audio and music: from coding to source separation," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 995–1005, 2009.

[60] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[61] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[62] ——, "For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[63] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, p. 717, 2009.

[64] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

[65] R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *Radar Conference, 2007 IEEE*. IEEE, 2007, pp. 128–133.

[66] A. C. Fannjiang, T. Strohmer, and P. Yan, "Compressed remote sensing of sparse objects," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 595–618, 2010.

[67] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

[68] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[69] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, 2011.

[70] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *Applied and Computational Harmonic Analysis*, vol. 35, no. 1, pp. 111–129, 2013.

[71] Z. Yang and L. Xie, "On gridless sparse methods for line spectral estimation from complete and incomplete data," *IEEE Transactions on Signal Processing*, vol. 63, no. 12, pp. 3139–3153, 2015.

[72] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[73] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," 2014.

[74] Y. Chi and M. Ferreira Da Costa, "Harnessing sparsity over the continuum: Atomic norm minimization for superresolution," *IEEE Signal Processing Magazine*, vol. 37, no. 2, pp. 39–57, 2020.

[75] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5987–5999, 2013.

[76] G. Tang, B. N. Bhaskar, and B. Recht, "Near minimax line spectral estimation," *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 499–512, 2014.

[77] T. K. Sarkar and O. Pereira, "Using the matrix pencil method to estimate the parameters of a sum of complex exponentials," *IEEE Antennas and Propagation Magazine*, vol. 37, no. 1, pp. 48–55, 1995.

[78] T. T. Georgiou, "The carathéodory–fejér–pisarenko decomposition and its multivariable counterpart," *IEEE Transactions on Automatic Control*, vol. 52, no. 2, pp. 212–228, 2007.

[79] L. Qiu and E. J. Davison, "The stability robustness of generalized eigenvalues," in *Proceedings of the 28th IEEE Conference on Decision and Control,*. IEEE, 1989, pp. 1902–1907.

[80] R. Patel and M. Toda, "Quantitative measures of robustness for multivariable systems," in *Joint Automatic Control Conference*, no. 17, 1980, p. 35.

[81] L. Qiu and E. Davison, "A new method for the stability robustness determination of state space models with real perturbations," in *Proceedings of the 27th IEEE Conference on Decision and Control*. IEEE, 1988, pp. 538–543.

[82] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4680–4688, 2011.

[83] K.-C. Toh, M. J. Todd, and R. H. Tütüncü, "SDPT3a MATLAB software package for semidefinite programming, version 1.3," *Optimization Methods and Software—, volume=11, number=1-4, pages=545–581, year=1999, publisher=Taylor & Francis.*

[84] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11, no. 1-4, pp. 625–653, 1999.

[85] K. R. Pattipati, S. Deb, Y. Bar-Shalom, and R. B. Washburn, "A new relaxation algorithm and passive sensor data association," *IEEE Transactions on Automatic Control*, vol. 37, no. 2, pp. 198–213, 1992.

[86] M. R. Chummun, T. Kirubarajan, K. R. Pattipati, and Y. Bar-Shalom, "Fast data association using multidimensional assignment with clustering," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 3, pp. 898–913, 2001.

[87] Y.-M. Chen, J.-H. Lee, and C.-C. Yeh, "Two-dimensional angle-of-arrival estimation for uniform planar arrays with sensor position errors," in *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, no. 1. IET, 1993, pp. 37–42.

[88] Z. Xiaofei and X. Dazhuan, "Antenna array self-calibration algorithm with sensor location errors," in *6th International Symposium on Antennas, Propagation and EM Theory, 2003. Proceedings. 2003*, 2003, pp. 225–228.

[89] Y. Liu, C. Liu, Y. Zhao, and J. Zhu, "Wideband array self-calibration and doa estimation under large position errors," *Digital Signal Processing*, vol. 78, pp. 250–258, 2018.

[90] Xiaoning Lu and K. C. Ho, "Taylor-series technique for source localization using AOA s in the presence of sensor location errors," in *Fourth IEEE Workshop on Sensor Array and Multichannel Processing, 2006.*, 2006, pp. 190–194.

[91] C. Qu, Z. Xu, and C. Wang, "Novel passive localization algorithm based on weighted restricted total least square," *Journal of Systems Engineering and Electronics*, vol. 24, no. 4, pp. 592–599, 2013.

[92] Y. Wang and K. C. Ho, "An asymptotically efficient estimator in closed-form for 3-d AOA localization using a sensor network," *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 6524–6535, 2015.

[93] A. Amar and A. J. Weiss, "Advances in direct position determination," in *Sensor Array and Multichannel Signal Processing Workshop*, 2004, pp. 584–588.

[94] G. Wu, M. Zhang, and F. Guo, "Self-calibration direct position determination using a single moving array with sensor gain and phase errors," *Signal Processing*, p. 107587, 2020.

[95] J. Yin, D. Wang, Y. Wu, and T. Tang, "Single-step localization using multiple moving arrays in the presence of observer location errors," *Signal Processing*, vol. 152, pp. 392–410, 2018.

[96] D. Wang, J. Yin, R. Liu, H. Yu, and Y. Wang, "Performance analysis and improvement of direct position determination based on doppler frequency shifts in presence of model errors: case of known waveforms," *Multidimensional Systems and Signal Processing*, vol. 30, no. 2, pp. 749–790, 2019.

[97] Z. Yang, D. Wang, B. Yang, and F. Wei, "Robust direct position determination against sensor gain and phase errors with the use of calibration sources," *Multidimensional Systems and Signal Processing*, pp. 1–34, 2020.

[98] Y. Chi, "Joint sparsity recovery for spectral compressed sensing," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3938–3942, 2014.

[99] Y. Wang and K. C. Ho, "Unified near-field and far-field localization for AOA and hybrid AOA-TDOA positionings," *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 1242–1254, 2018.

[100] J. Huang and T. Zhang, "The benefit of group sparsity," *Annals of Statistics*, vol. 38, no. 4, pp. 1978–2004, 2010.

[101] M. B. Wakin, M. F. Duarte, S. Sarvotham, D. Baron, and R. G. Baraniuk, "Recovery of jointly sparse signals from few random projections," in *Advances in Neural Information Processing Systems*, 2006, pp. 1433–1440.

[102] C. R. Berger and J. M. F. Moura, "Noncoherent compressive sensing with application to distributed radar," in *2011 45th Annual Conference on Information Sciences and Systems*, 2011, pp. 1–6.

[103] M. Wei, "Group sparsity techniques for data fusion of a passive miso radar network," in *2016 17th International Radar Symposium (IRS)*. IEEE, 2016, pp. 1–5.

[104] Z. Yang and L. Xie, "Exact joint sparse frequency recovery via optimization methods," *IEEE Transactions on Signal Processing*, vol. 64, no. 19, pp. 5145–5157, 2016.

[105] Y. Chi and Y. Chen, "Compressive two-dimensional harmonic retrieval via atomic norm minimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 4, pp. 1030–1042, 2015.

[106] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[107] G. Schwarz *et al.*, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[108] N. Boyd, G. Schiebinger, and B. Recht, "The alternating descent conditional gradient method for sparse inverse problems," *SIAM Journal on Optimization*, vol. 27, no. 2, pp. 616–639, 2017.

[109] N. Rao, P. Shah, and S. Wright, "Forwardbackward greedy algorithms for atomic norm regularization," *IEEE Transactions on Signal Processing*, vol. 63, no. 21, pp. 5798–5811, 2015.

[110] N. Boumal and P.-A. Absil, "Low-rank matrix completion via preconditioned optimization on the grassmann manifold," *Linear Algebra and its Applications*, vol. 475, pp. 200–239, 2015.

[111] J. L. Horner and P. D. Gianino, "Phase-only matched filtering," *Applied Optics*, vol. 23, no. 6, pp. 812–816, 1984.

[112] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "A sparse-group lasso," *Journal of Computational and Graphical Statistics*, vol. 22, no. 2, pp. 231–245, 2013.

[113] S. M. Kay, *Fundamentals of Statistical Signal Processing.* Prentice Hall PTR, 1993.

[114] B. Gear, E. Mellios, A. Nix, and J. McGeehan, "A maximum likelihood location estimator for non-line of sight geolocation of radio emitters," in *2019 13th European Conference on Antennas and Propagation (EuCAP)*, March 2019, pp. 1–5.

[115] J. C. Chen, R. E. Hudson, and K. Yao, "Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field," *IEEE Transactions on Signal Processing*, vol. 50, no. 8, pp. 1843–1854, 2002.

[116] A. J. Weiss, "Direct position determination of narrowband radio frequency transmitters," *IEEE Signal Processing Letters*, vol. 11, no. 5, pp. 513–516, 2004.

[117] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.

[118] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, "Distributed compressive sensing," *arXiv preprint arXiv:0901.3403*, 2009.

[119] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Conference Record of the Thirty-Ninth Asilomar Conference onSignals, Systems and Computers, 2005.*, 2005, pp. 1537–1541.

[120] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.