Algorithms for Indoor Positioning Systems Using Ultra-Wideband Signals

J. Yan

Algorithms for Indoor Positioning Systems Using Ultra-Wideband Signals

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. ir. K. C. A. M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op donderdag 23 december 2010 om 10.00 uur

door

Junlin Yan

Master of Science in Engineering
Chalmers University of Technology, Gothenburg, Sweden
geboren te Shanghai, China

Dit proefschrift is goedgekeurd door de promotor: Prof. dr. P. J. G. Teunissen

Copromotor: Dr. C. C. J. M. Tiberius

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. P. J. G. Teunissen,	Technische Universiteit Delft, promotor
Dr. C. C. J. M. Tiberius,	Technische Universiteit Delft, copromotor
Prof. dr. M. G. Di Benedetto,	University of Rome La Sapienza
Prof. dr. K. G. Langendoen,	Technische Universiteit Delft
Prof. dr. P. J. Oonincx,	Netherlands Defense Academy
Prof. dr. P. J. M. van Oosterom,	Technische Universiteit Delft

Yan, Junlin

Algorithms for indoor positioning systems using ultra-wideband signals

Remote Sensing (RS),
Mathematical Geodesy and Positioning (MGP),
Delft University of Technology
ISBN-978-94-6113-029-7

This dissertation is dedicated to:
my parents, Liegen and Jinye,
my girl friend, Juan,
for their love and support.

Abstract

Algorithms for Indoor Positioning Systems Using Ultra-Wideband Signals

Positioning systems and techniques have attracted more and more attention in recent years, in particular with satellite navigation technology as a tremendous enabler, and developments in indoor navigation. The work presented in this thesis has been conducted within the research project: "HERE: indoor positioning based on UWB radio signals", which aims at developing an alternative solution to the indoor positioning problem, since the Global Navigation Satellite Systems (GNSSs), e.g. the Global Positioning System (GPS), generally can not provide reliable positioning services in indoor areas, due to strong signal attenuation and dense multipath effects. The project focuses on range-based technologies, which are mainly composed of two parts: 1) ranging based on information such as Time of Arrival (ToA), Time Difference of Arrival (TDoA) and Received Signal Strength (RSS); 2) positioning using the obtained ranging results. This four-year project has been carried out in a team composed of two Ph.D. candidates and two supervisors. As one of the Ph.D. candidates, the author of this thesis focuses on the positioning part of the project, with the main covered issues summarized as follows.

Iterative Descent Methods The majority of existing systems, e.g. GPS, solve the positioning problem using Iterative Descent (ID) methods based on non-linear least-squares. Applying these ID methods for indoor positioning purpose faces the following three major problems:

- It is more difficult to obtain a good initial guess, which is critical for the ID methods to converge to the correct solution and converge faster. For on-earth satellite navigation applications, a good initial guess can be very easily obtained by choosing the Earth's center since the other (local) solution (if any) usually lies far in space. On the contrary, for indoor positioning, it is not as easy to obtain a good initial guess since the information on the user position is rather limited beforehand. In this thesis, we propose to use the so called direct methods or simply the geometric center of the seen transmitters, for initial guess. A thorough study of the existing direct methods is also given.

- The non-linear least-squares estimator is inherently biased, even with unbiased range measurements, since the expectation of the higher order terms in the final estimator is non-zero. This bias is generally negligible for GNSSs, with extremely large satellite-user distance, but can be problematic for indoor applications with reduced geometric system scale. Based on an analysis of the bias due to nonlinearity, a scheme is proposed to test the significance of the bias. The corresponding work is validated

with measurements obtained using UWB acoustic testbed.

- The iterative nature of the ID methods may be computationally too heavy for indoor applications, which usually require low power systems. Aiming at reducing the computational load of traditional iterative descent algorithms, a new framework is proposed for position estimation. The multidimensional non-linear localization problem is first transformed to a lower dimension and then solved iteratively. In three dimensional positioning systems, the achievable reduction on the amount of computation in each iteration is 67%. On the other hand, accurate positioning results can be obtained with this low-complexity framework, especially with TDoA measurements.

Direct Methods In general, strict direct (non-iterative) least-squares solutions to non-linear problems do not exist. However, with some assumptions or simplifications, direct least-squares positioning algorithms can be developed. In fact, there is a large number of direct methods documented in literature, scattered across the fields of radar, aerospace engineering, oceanic engineering, (acoustic) signal processing and wireless communications. Some of the documented methods are, surprisingly, identical, though the derivations are often greatly different. A deep study and a proper classification of the methods helps to achieve a better understanding, which is one of the central contributions of this thesis. It can be used to assist researchers and developers in the field to find the right choice for their applications. The direct methods provide simple to compute estimates, but they are suboptimal with the introduced simplifications. Based on the survey of the existing direct methods, a new non-iterative algorithm has been developed to improve the positioning accuracy. Theoretical proof is given that the method provides a better estimator in the sense that it corresponds to an equal or smaller value of the original least-squares objective function. It does so by exploiting two similar fully constrained models. Meanwhile, the non-iterative nature makes the algorithm attractive for low cost, low power applications.

NLoS Identification/Mitigation It is widely known that the Non-Line of Sight (NLoS) effect is one of the main degraders for the position estimation accuracy. Hence, NLoS identification and mitigation is another important and hot subject regarding indoor positioning. A review of several existing NLoS identification and mitigation schemes is provided, and four new schemes are described, which are based on systematic positioning model hypothesis testing. The idea is to combine the statistics of timing- and RSS-based range measurements, and in the mean time exploit the fact that all collected measurements are related to the same unknown position. The four proposed schemes all use the combination of timing and RSS measurements because 1) the timing measurements are usually very accurate compared to RSS measurements, 2) the distributions of RSS measurements under LoS and NLoS conditions are well separated. The computational load of the schemes decreases as more simplifications are introduced, and the performance, however, is also degraded in general, except the case where most of the links are NLoS. Validation results show that, under full UWB signal bandwidth of 7.5 GHz, up to 99% correct decision rate can be achieved.

Contents

1 Introduction

Positioning and Navigation have been of long-standing interest and have a fascinating history. In old times, our forebears crossed vast continents and oceans guided only by stars in the sky which could not be counted on to be always visible. Today, people rely on radio signals received from artificial stars (i.e., satellites) for navigation purpose.

In modern society, besides its value in military events, positioning services find their applications in a variety of areas, such as industry, commerce, science, sports, and personal everyday life, and these applications are being developed at an astonishing rate.

Generally, positioning services may be divided into two categories: outdoor and indoor.

Outdoor There are many outdoor scenarios where positioning services fit in: surveying, exploitation of natural resources, public safety, transportation on land, via sea, and in the air, hiking, cycling, gaming, running, atmospheric science, geodesy, and wild life tracking, etc.

As the first global navigation satellite system to become operational, the Global Positioning System (GPS) is now being widely used in our daily life and supports most of the applications, if not all, listed above. Buy a $100 GPS receiver and a map, and you would not be lost as long as you have a clear view of the sky. Or, buy a pair of more capable receivers for about $5000 each and, with a careful analysis of the measurements, you would be able to tell if the earth under your feet moved a few millimeters while you were not looking [1].

No wonder, GPS is a great technological success with far greater impact on the military and civil worlds than could have been foreseen by the designers of the system in the 1970's. However, as wonderfully as it works in outdoor environments, it does not work well when it comes to indoor cases.

Indoor With the outdoor applications discussed above, it should not be too hard to imagine how vast the market for indoor positioning service would be just by simply thinking of how much more time most people spend indoors than outdoors. Indoor positioning service may be desirable at the places listed below, both public and private: large department stores, supermarkets, airports, railway stations, subways, universities, hospitals, offices etc. The corresponding applications could be logistics such as package tracking, security applications such as localizing authorized persons in high-security areas, medical purposes such as monitoring patients, family communication/supervision of children, search and rescue such as locating firefighters in accidents like fires, avalanches and earthquakes [2].

Unlike in outdoor open areas, a signal in indoor environments typically reaches the receiver via more than one path because of the structures in the vicinity of it such as doors, walls,

furniture, floor, ceiling, people and so on. These structures have two major effects on the received signal:

- when there is no blockage, the received signal will consist of a Line of Sight (LoS) direct path and the delayed and usually attenuated replicas due to reflections and other phenomena, which is defined as multipath propagation,
- when blockages occur, only reflected replicas and possibly the direct path with extra attenuation and delay (if not absolutely blocked, e.g. just a brick wall in the way) are received, which is defined as Non-Line of Sight (NLoS) propagation.

First, we explore how multipath propagation degrades GPS performance. As we know, GPS is a positioning system based on pseudorange [1] measurements obtained using timing information. A conventional GPS receiver uses a correlation based algorithm where the time shift of the template signal, generated by the receiver, that corresponds to the maximum correlation peak with the received signal is used as the pseudo Time of Arrival (ToA) estimate. In indoor cases as we described above, the received replicas of the transmitted signal due to reflections may partially overlap with the direct path. This is typical for a narrow band system like GPS. As a result, the output of the correlator contains not only the correlation between the template and the first path signal, but also the correlations between the template and the replicas. The latter will shift the position of the correlation peak (i.e., the output of the correlator does not peak at the correct timing), and therefore will introduce errors in the final result.

Second, NLoS propagation is another deadly issue for GPS to work properly in indoor spaces. With strong attenuation, the detection of a satellite signal in indoor environments usually requires a 40+ dB processing gain, on top of the typical 33 dB gain required in outdoor areas, which generally can not be fulfilled with a standard GPS receiver [3]. Hence, even when the GPS signal can reach users in buildings, the available signal power will be too weak for fast and reliable signal acquisition at the GPS receivers [4].

In indoor environments, the GPS positioning error, the distance between an estimate and the true position, can easily go beyond 100 m [5], even with advanced high sensitivity GPS receivers [6], this is due to the fact that:

1. The weak GPS signal power and dense indoor multipaths effects introduce large errors in the pseudoranges, and propagate to final position estimates.
2. Due to NLoS effects, the GPS signal usually can only be received in a limited area, e.g. an area close to the windows. This means that the satellites seen by the receiver are generally on one side of the receiver, and positioning is performed in a poor geometry.

To this end, we see that the requirements for an indoor positioning system include

1. Robustness to dense multipaths effects due to indoor structures.
2. Good penetration through structures, so that positioning can be done in a favorable geometry where the receiver is surrounded by transmitters.

3. Sub-meter or even centimeter level accuracy because of the natural scale of indoor spaces.

Fortunately, Ultra-WideBand (UWB) signals can potentially support a positioning system that meets the requirements listed above [7, 8]. UWB is a promising candidate for future indoor positioning systems and its definition and properties are discussed next.

1.1 UWB Signal

UWB technology is not a new concept and its application in radar dates back to the early 20th century. By the late 80's, UWB has found more applications in radar, communications, and positioning systems [9], and it has drawn significant attention [10], after the Federal Communications Commission (FCC) announced its regulations on the use of unlicensed UWB communications in February 2002 [11]. According to the FCC, the UWB signal is defined as any signal that has a fractional bandwidth greater than 0.2 or occupies 500 MHz or more of spectrum. The formula proposed by the Commission for calculating the fractional bandwidth is $2(f_H - f_L)/(f_H + f_L)$ where f_H and f_L are the upper and lower frequency of the -10 dB emission point, respectively, of the Equivalent Isotropically Radiated Power (EIRP).

With the large bandwidth, UWB signals inevitably use spectrum parts that have already been assigned to other systems. Therefore, the FCC defined a set of spectrum limits for both average and peak measured EIRP, concerning different applications. The adopted limits are 1) a limit on the average power set to an EIRP of -41.3 dBm when measured with a spectrum analyzer of a 1 MHz Resolution Bandwidth (RBW) and 2) a limit on the peak power set to an EIRP of 0 dBm when measured with a 50 MHz RBW. For an analyzer with a different RBW, the differences of measured power are based on a $20\log_{10}$ relationship rather than $10\log_{10}$ (experimental observation), for systems operating with pulse repetition frequencies less than 6.7 MHz [11]. Since peak measurements using a spectrum analyzer with a 3 MHz RBW are relatively straight-forward and can be performed using existing measurement procedures, the limit on peak power can be converted to $20\log_{10}(3/50) = -24.4$ dBm with a 3 MHz RBW analyzer [11].

In Figure 1.1, the FCC limits on average UWB EIRP for indoor applications are shown, with the main agreed bandwidth from 3.1 - 10.6 GHz. For the whole spectrum, the limit is at or below -41.3 dBm. For spectrum parts where particularly sensitive systems with low received power are allocated, e.g. GPS, the limit is set much lower to -75.3 dBm.

In Europe, the Electronic Communications Committee (ECC) of European Conference of Postal and Telecommunications Administrations (CEPT) has also released an official EIRP mask for applications using UWB. In Figure 1.2, the CEPT limits on both average and peak UWB EIRP for applications in bands below 10.6 GHz are shown [12], which are more strict than the FCC mask.

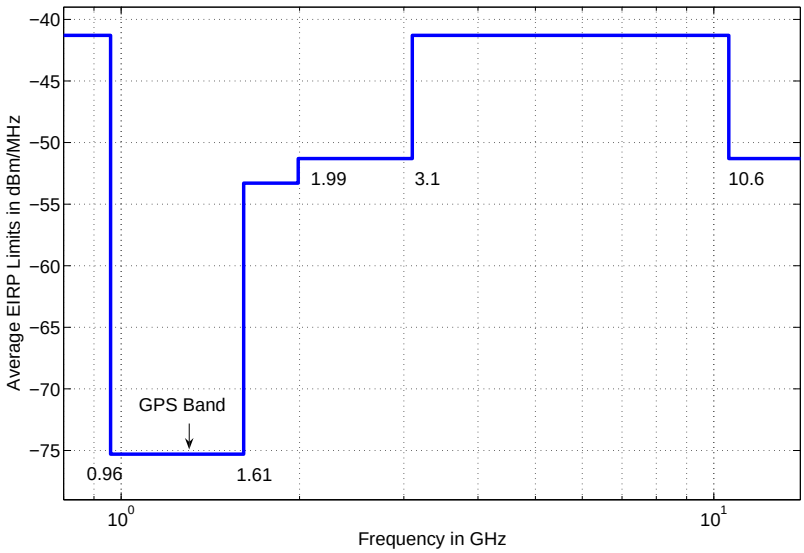

Figure 1.1: FCC UWB maximum average EIRP limit for Indoor Systems. Source [11].

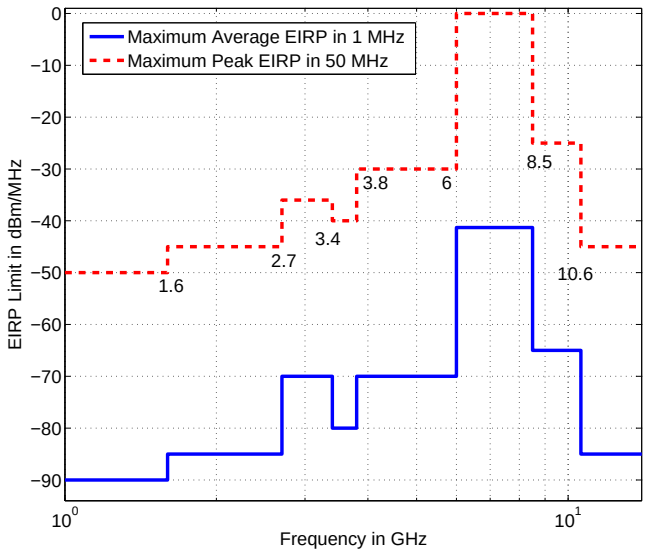

Figure 1.2: CEPT UWB maximum average and peak EIRP limit for applications in bands below 10.6 GHz. Source [12].

1.1.1 Advantages of UWB for Indoor Positioning

It is stated that the UWB technique can be advantageously applied in the field of localization and navigation [13, 14]. A brief comparison between GPS and UWB in the context of indoor positioning can be found in [15].

The major advantages of an UWB signal as a candidate for indoor positioning against narrow band signals (e.g., GPS signal) can be described based on the following issues.

Fine Time Resolution vs Multipath Effects As mentioned before, multipath propagation is one of the key issues to indoor positioning applications. To combat multipath effects, an easy option would be removing the obstacles that may introduce high level reflections or putting the receiver in a more open area. This might help at some outdoor places but is not applicable in most indoor scenarios. Hence, a more feasible option would be using large bandwidth signals that have very short pulses in the time domain, since the peak width of the autocorrelation function is inversely proportional to the signal bandwidth [16], and a narrow correlation peak enhances the ability to accurately estimate the ToA of a signal and separate the multipath contributions.

Obviously, a UWB signal is a very good choice in this sense. Corresponding to the potential maximum absolute bandwidth of 7.5 GHz and a minimum absolute bandwidth of 500 MHz, the pulse durations (approximated by the inverse of the bandwidth) are of the order 133 picoseconds and 2 nanoseconds, respectively. This means in the worst case, the main lobes of the received replicas from two paths will not overlap if they arrive at time instances separated by 2 nanoseconds (equivalent to a distance of 60 cm). Hence, regarding the target ranging accuracy, multipaths are usually resolvable without very complex algorithms [2]. On the other hand, the chip duration of the GPS C/A code is roughly 1 microsecond which makes the GPS signal much less robust to multipath effects. Please note that the above reasoning only holds for LoS cases, and is meant to roughly show the great improvement UWB technology can potentially bring.

Large Bandwidth vs NLoS Effects With NLoS propagation, the direct path is severely attenuated and delayed or even absent at the receiver. This phenomenon poses the biggest challenge for indoor positioning because

1. it introduces large NLoS delays in the obtained ToA results and these errors further propagate to the positioning results, if they are not appropriately accounted for.

2. NLoS effects may introduce situations where positioning can not be performed since the number of transmitters seen by the receiver is less than the number of unknown position coordinates.

Therefore, it is preferable to choose signals that have good penetration properties. Considering the assigned very wide bandwidth, UWB signals have a large variety of frequency components and the probability that at least some of them can pass through or around structures is large.

One should however understand that UWB does not provide fine time resolution and good penetration at the same time, but UWB waveforms present a better choice than do conventional radio signals [8].

1.2 Positioning Schemes

The majority of positioning schemes estimate the unknown receiver (user) position by exploiting geometric relations between transmitters and the receiver, and they can be characterized by the used measurements, including timing information (ToA and Time Difference of Arrival (TDoA)), Received Signal Strength (RSS), and Angle of Arrival (AoA). There also exist approaches that utilize location-dependent characteristics, instead of geometric relations, of these measurements to find the position. Such techniques are usually called location fingerprinting.

In this section, an overview of the mentioned positioning schemes is provided, and by comparing their advantages and disadvantages, the schemes to focus on in this thesis are selected.

Please note that in this thesis, our discussion by default concerns self positioning systems (like GPS), but the discussion also applies for tracking systems in which the user sends signals to a set of stations where the signal processing and the position estimation are accomplished.

1.2.1 Timing-based Positioning (ToA and TDoA)

ToA The ToA based positioning scheme combines the ToA measurements between the receiver and a number of transmitters to obtain the position, since the range between a transmitter and a receiver can be readily obtained by multiplying the ToA with the speed of light. This scheme requires the receiver to know the accurate time of transmission, which can be achieved e.g. in a synchronized system where transmitters and the receiver have a common clock, or in a system where all the units are transceivers with capability of estimating round-trip-time [17, 18, 19, 20]. In the latter case, one unit sends a signal to a second unit, who immediately transmits the signal back. The round trip time is proportional to twice the propagation time plus processing time required at the second unit.

The advantages of ToA based schemes include:

- ✓ The fine time resolution of the UWB signal can be well exploited. In a synchronized system, very accurate ToA estimates can be obtained [18], especially when the UWB signal is used [21]. The alternative solution of a transceiver system is simpler to implement and less costly, but the accuracy of the corresponding ToA estimates can be degraded compared to that of a synchronized system [17].
- ✓ Timing based multiple access schemes, such as Code Division Multiple Access (CDMA) and Time Division Multiple Access (TDMA), are already widely implemented in existing systems, which allow high-accuracy ToA estimation [22, 23] and make timing based location schemes very attractive [24].

But the disadvantages are:

× Synchronization among all the units in a system is in general difficult and costly to achieve for wireless radio systems. In [25], the reported synchronization in wireless sensor networks is on the order of 10 μs, corresponding to 3 km in distance, which is not good enough for accurate radio ranging.

  For an acoustic system, synchronization can be achieved in a much easier and less costly way. The transmitter first sends a radio signal, only for synchronization purpose, and then the acoustic signal for ranging. The receiver obtains the ToA of the acoustic signal by calculating the difference in arrival times of the radio and acoustic signal. The background idea is that the travel time of the radio signal negligible, as compared to that of the acoustic signal, due to the fact that the speed of light is much larger than the speed of sound. However, compared to UWB radio systems, acoustic systems face other disadvantages including complete loss of signal due to obstruction; vulnerable to interference such as human voice, keys jangling, rustling paper etc. [26], depending on the applied frequency band.

× In the round trip time approach, ToA estimates are obtained using two way ranging which requires that all the units in the system are transceivers. This can be a less costly way, compared to a synchronized system, to realize ToA ranging, but still implies complicated and quite expensive structures of the system devices, especially in case of an UWB radio system where hardware requirement on the receiver side is known to be demanding [2, 16].

TDoA The demanding requirement on full synchronization is relieved in a TDoA system, where only the transmitters are synchronized, and the receiver does not know the time of transmission. A perfect example of a TDoA positioning system is the GPS. The GPS satellites (transmitters) are synchronized by means of 1) using extremely expensive atomic clocks and 2) monitoring through the ground control segment. However, a conventional GPS receiver can not use such an advanced scheme due to cost and size limits, and therefore is asynchronous to the satellites.

Due to the misalignment between the receiver and transmitter clocks, the corresponding measured ranges may thus be all too short, or too long, by a common amount, and are called *pseudoranges* [1]. The receiver clock bias can be treated in two ways:

1. The clock bias can be canceled out by taking the difference of the times of arrival of signals from two transmitters, and hence the name TDoA. In this case, only coordinates of the receiver need to be estimated since the clock bias is absent in TDoA measurements. However, by taking the difference between ToA measurement pairs, one ToA measurement has been consumed, and the number of 'new' TDoA measurements is reduced by one.

2. The clock bias can also be treated as an unknown parameter, in addition to the coordinates of the unknown receiver position. In this case, the number of measurements is unchanged, but the number of unknown parameters is increased by one.

In fact, since taking the difference of a pair of two ToA measurements is a linear operation, the least-squares solution of the two listed approaches are always equivalent for a fully determined measurement set, and are equivalent under some conditions for over determined measurement sets. This will be elaborated on in Chapter 4.

In literature, the second way is sometimes classified as a ToA approach, e.g. in [16, 27]. However, in this thesis, the terms ToA and TDoA are used according to the presence of a receiver clock bias in a timing based positioning system, i.e. ToA for the cases where all clocks are synchronized, and TDoA for cases where the user receiver clock is asynchronous to the synchronized transmitter clocks.

In general, with the intrinsic similarities, a TDoA scheme shares most of the advantages and disadvantages of a ToA scheme, but with some particular differences listed as follows:

- ✓ Synchronizing only the transmitters with known positions in a TDoA scheme is less expensive than synchronizing all the units in a ToA scheme [17].
- × The presence of the clock bias either introduces an extra unknown to the system or needs to be canceled out by consuming a measurement. The positioning accuracy is worse than a ToA system with the same system geometry.
- × The clock bias is involved in the range equations in a linear manner, which is different from the non-linear manner, through which unknown position coordinates are involved. This fact makes the system geometry design of a TDoA scheme more complicated than a ToA scheme, since a rank defect situation can happen in more geometries with a TDoA system [1]. For example, in 3-dimensional (3-D) space, if the distance vectors between the transmitters and the receiver form a cone, see Figure 1.3 for details, the TDoA position problem is not solvable, since there exists an unlimited number of possible values for the clock bias, corresponding to an unlimited number of possible positions of the receiver, locating along the cone's line of symmetry. But a set of ToA equations corresponding to this type of geometry is still solvable.

1.2.2 RSS

RSS is defined as the power measured by a received signal strength indicator circuit, or equivalently reported as the squared magnitude of the signal [16]. It can be related to the distance and eventually be used for positioning purposes, if its path loss model is known. The major advantages and disadvantages of a RSS scheme are summarized as follows.

Advantages:

- ✓ RSS techniques are relatively cheap and easy to implement, and can be obtained using low-complexity algorithms [10, 28, 29, 30].
- ✓ Expensive time synchronization among the system units is not required and also no clock bias exists in RSS systems.

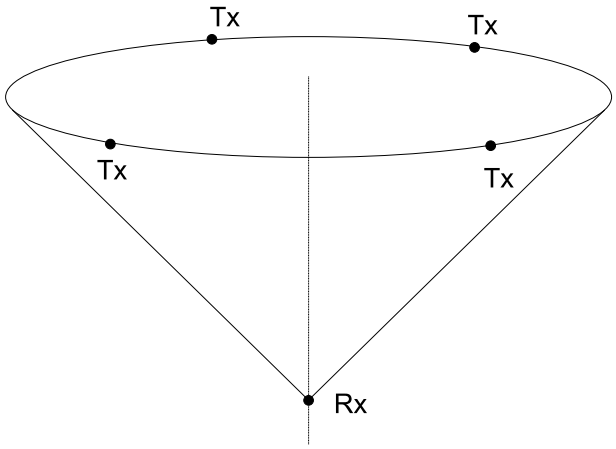

Figure 1.3: A 3-D geometry where the distance vectors between the transmitters and the receiver form a cone. The transmitters are located on the circle and the receiver is located on the vertex. In this case, the position of the receiver can not be determined with TDoA, since there exists an unlimited number of possible solutions along the axis of symmetry, with each corresponding to a different clock offset.

Disadvantages:

- × RSS-based positioning accuracy is usually much poorer than for other techniques and the scheme does not fully exploit the benefit brought by the wide bandwidth of UWB signals [2].
- × The determination of relevant parameters of the path loss model requires a measurement campaign. This, however, is not a strong disadvantage, since other techniques, e.g. ToA-based, may also need some amount of pre-measured samples to estimate measurement variance, which is helpful to achieve a better accuracy in the positioning stage.
- × Although no time synchronization is required, an accurately measured power value at a known distance is wanted, which is later used as a reference in an RSS path loss model. In a 3-D positioning system, the transmitting direction towards the user is location dependent. This makes it practically impossible to find an accurate global reference power for all positions, due to the fact that an antenna does not transmit signals with equal power in all directions [31, 32]. A study of the effects of antenna directivity on path loss in UWB indoor wireless channels can be found in [31].
- × The RSS technique is only suitable for short distance applications. For long-distance applications, such as satellite navigation, ranging and positioning results would have extremely poor accuracy due to the fact that such cases correspond to the flat tail area of the log-shaped pass-loss curve, where a very small change in RSS translates into a huge change in estimated distance.

1.2.3 AoA

By using direction information, the AoA scheme provides positioning results complementary to the timing-based and RSS-based schemes discussed above. The key difference of the

AoA scheme, as compared to the other two schemes, is that it requires an antenna array (instead of a single antenna) at the receiver to estimate the AoA of the received signal. However, we should note that the AoA scheme also has some common features with the other two schemes, since AoA of the received signal regarding a certain transmitter is usually estimated based on the (pseudo)range differences among antenna components at the receiver.

AoA-based system has the advantages:

- ✓ Expensive time synchronization among the system units is not required since the AoA of one transmitter-receiver pair is obtained using the (pseudo)range differences of multiple antenna components at the same receiver. Meanwhile, there is no need to consider clock bias effect in AoA systems. This means that the AoA scheme works with three types of systems, where: 1) no synchronization is established between any two units; 2) only transmitters are synchronized; 3) all units are synchronized.
- ✓ Less transmitters are required to calculate the final position, than the range-based systems. For example, in Figure 1.4 with two transmitters and one receiver in a 2-D plane, assuming the coordinates of Tx_1 and Tx_2 are (x_1, y_1) and (x_2, y_2), with estimated a_1 and a_2, the unknown position of Rx can be obtained as: $(\frac{y_1-y_2+x_2\tan a_2+x_1\tan a_1}{\tan a_2+\tan a_1}, \frac{x_1-x_2+y_2\cot a_2+y_1\cot a_1}{\cot a_2+\cot a_1})$. For ToA and RSS schemes, two candidate solutions (Rx and Rx′) will be obtained, and one needs another transmitter or other additional information to pick the correct solution out of the two candidates. A TDoA scheme also does not work here with only two, since one extra transmitter is required to deal with the clock offset.

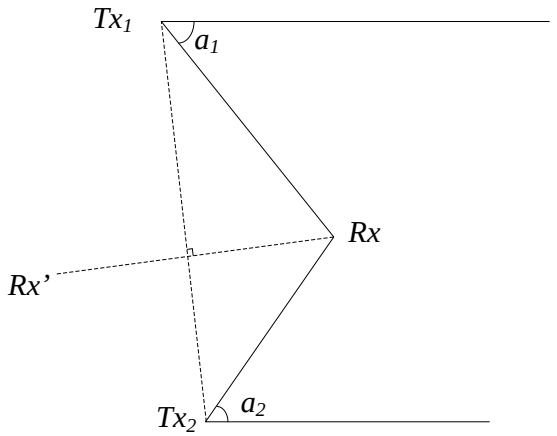

Figure 1.4: Geometric interpretation for 2-D AoA based systems.

And the disadvantages:

- × The antenna array required by AoA techniques increases the cost and the size of the device.
- × The dense scattering feature of indoor channels also imposes challenges on accurate AoA estimation [2].

1.2.4 Location Fingerprinting

Location fingerprinting refers to a technique that exploits the relationship between RSS and a specific location [33]. Generally, the deployment of fingerprinting based positioning systems can be divided into two phases [34].

The first phase of this technique is to collect location fingerprints by performing a site-survey of the RSS from multiple transmitters, usually called the calibration phase. The entire area is covered by a grid (usually rectangular) of points. The RSS values are measured with enough statistics to create a table of predetermined RSS values for each point of the grid. The set of RSS values assigned to a grid point is called the fingerprint of that point.

In the second phase, a user may report the measured RSS to a central server, where a positioning algorithm is performed to estimate the location and sends the estimate back to the user. This requires that both the user device and the central server are transceivers. However, if the user device has a local fingerprinting map and enough power for position computation, there is no need for the user to upload measured RSS values. The most common algorithm to estimate the location computes the Euclidean distance between the measured RSS vector and each fingerprint in the database [34]. The coordinates associated with the fingerprint that provides the smallest Euclidean distance is chosen as the estimate of the position.

The major advantages of the location fingerprinting technique include:

- ✓ Since the technique is based on RSS, it has the advantage of simple and low cost hardware implementation.

In practice, the quality of the final position estimate is heavily dependent on the following aspects, which compose the disadvantages of the fingerprinting technique:

- × The final positioning accuracy depends on the number and the distribution of the survey points. A big number of survey points may increase final positioning accuracy[1], but this requires more efforts at the first phase and increases the processing time at the positioning stage in the second phase.
- × The performance of fingerprinting type of techniques is known to be very sensitive to environmental changes, which, however, are inevitable in most of the indoor cases. Movements of people or rearrangement of furniture can easily degrade the positioning accuracy using fingerprinting.

1.2.5 Other Schemes

There also exist other schemes which do not exploit range information (timing-based, RSS, AoA) or scene survey (fingerprinting). One simple example is the cell identification technique [35], which relies on the fact that mobile cellular networks can identify the

[1] A grid with too dense points may not contribute to the final positioning accuracy, because the fingerprints of two points on the grid that are close to one another may be very similar.

approximate position of a mobile handset by knowing which cell site the device is using at a given time. We will not discuss these schemes in detail since the achievable positioning accuracy is generally too low for most indoor applications.

1.2.6 Schemes Used in This Work

Based on the advantages and disadvantages of the schemes listed above, we choose range-based positioning (timing and RSS schemes) in general, and timing-based positioning in particular, as the focus of this PhD work.

Special attention is given to the timing scheme since it best exploits the fine time resolution benefit offered by the very large bandwidth of UWB. Under LoS situations, timing-based ranging has been reported to have centimeter level accuracy [36][37][21][10]. Under NLoS situations, the ranging accuracy is degraded due to blockage, but this is a problem encountered by other schemes as well. The major disadvantage of the timing scheme, i.e. synchronization and transceiver implementation, can be potentially solved by a low complexity receiver concept as proposed in [38].

The RSS scheme is also adopted, with its advantage of low cost and easy implementation. Although RSS techniques do not provide centimeter level accuracy, it may well serve applications that are satisfied with decimeter or meter level accuracy but have strong constraints on system cost and complexity. More importantly, its low cost does make it attractive for possible hybrid positioning schemes.

The AoA scheme is not considered mainly due to the fact that an antenna array contributes high cost and large size to the system. The fingerprinting technique is currently also not our choice due to the fact that it is quite sensitive to environmental changes, which can easily happen in indoor areas. Moreover, an a-priori survey is required by this technique while in our work, the aim was to achieve positioning functionality with as less a-priori information as possible.

From now on, the term ToA/RSS will be used for the cases where all the clocks are synchronized or no timing information is required for ranging. In the coming section, a brief overview of the existing least-squares positioning algorithms is given. These algorithms will be further characterized in Chapter 4.

1.3 Least-Squares Positioning Algorithms

Numerical estimation of the unknown parameters of interest based on range measurements is usually done by maximizing or minimizing an objective function. Typical objective functions are sums of squared measurment residuals, likelihood functions, posterior density functions, risk functions, and robust loss functions. The two most common estimation criteria are least-squares and Maximum Likelihood (ML). In this thesis, we will focus on algorithms based on least-squares, since with a proper measurement weight matrix applied, the least-squares estimator and the ML estimator are equivalent, and its variance achieves the Cramer-Rao Lower Bound (CRLB) [39, 40], if the Probability Density Function (PDF) of the measurements is elliptically contoured, e.g. Gaussian. Furthermore, compared to

ML estimation, un-weighted least-squares estimation does not require any knowledge of the measurement PDF, and only first and second order moments of the measurement errors may be required in the weighted case. The details about least-squares estimation will be discussed in Chapter 3.

In fact, linear least-squares estimation is quite simple where a direct, closed-form solution can be easily obtained. The difficulty of applying least-squares theory for range-based positioning is that ranges and coordinates are involved in a non-linear relation. The currently existing least-squares positioning algorithms can be generally categorized as follows.

1.3.1 Iterative Descent Methods

The unknown Cartesian position coordinates are involved in the range equations in a non-linear way. The coordinates are usually estimated using iterative descent techniques, such as the Steepest-Descent method, the Newton method, the Trust-Region method, the Gauss-Newton method and the Levenberg Marquardt method. A detailed analysis of these methods can be found in [41]. The iterative descent methods solve the non-linear positioning problem *iteratively* by means of linear inference, i.e. they start with an initial guess of the unknown position, which is then updated, from iteration to iteration, based on the rule defined by the specific method. The final estimate results upon convergence. Traditional iterative descent methods, when successfully converged, provide estimators that closely approximate strict least-squares solutions. One drawback is that they require an initial guess of the solution to start. In the mean time, the heavy computations involved in the iterations can be problematic for low cost and low power indoor applications.

1.3.2 Direct

In general, without simplifications, direct (non-iterative) least-squares solutions to non-linear problems do not exist. However, with some assumptions or simplifications, direct least-squares positioning algorithms can be developed, which are extensively documented in literature, for both TDoA [42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54] and ToA/RSS [55, 56, 57, 58, 59, 60] systems. All these methods apply a squaring operation on the used range equations to get rid of the square root. After squaring, both linear and non-linear terms of the unknown parameters are obtained, and the non-linear term, is treated as an unknown parameter in addition to the unknown position coordinates. As we will discuss later in Chapter 4, these methods can be clearly classified based on the objective functions, and it can be shown that many different variants of the same method do exist, which can be traced back to just the use of different measurement weighting schemes. The direct methods provide simple to compute estimates, but they are not strict least-squares solutions to the used equations, since the relation between the non-linear term and the unknown parameters is not rigorously exploited.

1.3.3 Low-complexity Iterative

There also exists a third group of methods that solve the positioning problem iteratively but with only one-dimensional calculations in each iteration. In [61], the estimation accuracy is improved over the direct methods, by introducing a Lagrange Multiplier in the objective function to fully account for the constraint between the non-linear term and the original unknowns after squaring. Applying Eigenvalue Decomposition (EVD), the final estimate is obtained by iteratively searching the roots of a high order polynomial with a single unknown. This method was further investigated in [62]. Another method [63, 64] first transforms the multi-dimensional positioning problem to one-dimension, based on direct methods, then finds the estimate for the remaining single unknown iteratively using the original un-squared range equations. These methods perform better than the direct methods in terms of Root Mean Squared Error (RMSE), but they are again iterative and require an initial guess to start. All the mentioned methods will be discussed in this thesis.

1.4 NLoS Identification and Mitigation

According to the discussion in Section 1.1.1, the very large bandwidth of UWB signals allows for good penetration capability through obstacles. However, as mentioned before, UWB signals do not provide fine time resolution and good penetration at the same time, meaning that in NLoS cases, the direct path of the received signal will still be delayed and attenuated. Hence, the obtained ToA-range and RSS measurements contains additional delays and extra power loss, respectively. If these effects are not accounted for, the final positioning results obtained in NLoS cases will contain large errors as compared to LoS cases.

Due to the above reason, the NLoS identification and mitigation issue has attracted a lot of research interests, and will also be addressed in this thesis in Chapter 6.

1.5 "HERE: indoor positioning based on UWB radio signals"

"This research was supported by the Technology Foundation STW, applied science division of NWO and the technology program of the Ministry of Economic Affairs."

The work presented in this thesis has been conducted within the research project: "HERE: indoor positioning based on UWB radio signals", supported by the Dutch Technology Foundation STW (project no. 0.7343). This four years project has been carried out by a team composed of two Ph.D. candidates and two supervisors, in a cooperation between the Wireless and Mobile Communications Group of the Faculty of Electrical Engineering, Mathematics and Computer Sciences (EEMCS) of the Delft University of Technology, and the Mathematical Geodesy and Positioning Group of the Faculty of Aerospace Engineering of the same university.

The main research questions that the project aimed to answer can be summarized in the following points:

1. Identification of characteristic parameters of the UWB multipath channel in typical indoor environments, for the purpose of positioning.
2. Determination of statistical UWB channel models for these parameters based on measurements, and verification of the models.
3. Investigation of transmitter and receiver architectures, and signal processing algorithms for extraction of the required channel parameter information.
4. Design and development of a real-time ultra wideband positioning system testbed based on acoustic signals.
5. Investigation of system non-linearity and feasibility of applying nonlinear positioning algorithms, in typical indoor environments.
6. Analysis of existing positioning algorithms and development of possible improvements aiming at lower computational load or higher positioning accuracy.
7. Development of positioning model hypothesis tests for outlier/bias detection, based on the developed statistical ranging models.

The overall outcome of the project is the results of the close cooperation among all the people involved in the project; however, this thesis specifically presents results related to the points 4-7 of the list, which represent the main research focus of the author. The items 1-4 are covered by the Ph.D. candidate Giovanni Bellusci of the Wireless and Mobile Communications Group in his thesis entitled 'Ultra-Wideband Ranging for Low-Complexity Indoor Positioning Applications'.

1.6 Main Contributions

The main contributions of this work can be summarized in line with the research items 4 through 7 listed in Section 1.5.

1. Two of the major issues on applying non-linear least-squares for solving positioning problems are that 1) a good initial guess should be fed to the iterative techniques to help achieving convergence to the correct/global minimum and reduce the number of iterations and 2) the least-squares estimator is inherently biased in non-linear cases since the expectation of the higher order terms in the final estimator is non-zero. The property of unbiasedness does not carry over in general through a non-linear relation. The bias is related to the geometric scale (size) of the system and the quality of the position estimator.

   For on-earth satellite navigation applications, a good initial guess can be very easily obtained by choosing the Earth's center since the other (local) solution (if any) usually lies far in space. Moreover, with extremely large satellite-user distance, the bias is negligible (as will be shown in Chapter 3) even when the standard deviations (STD) in the position estimators are at the 100 m level (10 m is common today with standalone GPS). On the contrary, for indoor positioning, it is not as easy to

obtain a good initial guess since the information on the user position is rather limited beforehand and the bias in the position estimator can become problematic due to the greatly reduced geometric system scale. For this reason, a thorough analysis of these nonlinearity effects, in the context of indoor positioning, is carried out and two schemes are proposed, one to obtain a good initial guess and the other to test the significance of the bias due to non-linearity. The corresponding work is validated with the measurements obtained using the UWB acoustic testbed.

2. Regarding the subject of positioning algorithms, especially for indoor positioning, several tutorial type of articles exist in literature [65, 66, 67, 35, 68, 16, 2, 69, 70]. These works cover a wide range of aspects of the subject on a general level, and devote most of the attention to the signal processing techniques at the stage where TDoA, ToA and RSS are estimated. None of the papers discusses in detail the mathematical positioning algorithms, which compose one of the central contributions of this thesis. In fact, there is a large number of direct methods documented in literature, scattered across the fields of radar, aerospace engineering, oceanic engineering, (acoustic) signal processing and wireless communications. Some of the documented methods are, surprisingly, identical, though the derivations are often greatly different. With a deep study and a proper classification of the methods, a better understanding can be achieved. This thesis can be used to assist researchers and developers in the field to find the right choice for their applications.

3. Aiming at reducing the computational load of traditional iterative descent algorithms, a new framework is proposed for least-squares positioning based on estimated ranges, covering TDoA, ToA and RSS cases. The multidimensional non-linear localization problem is first transformed to a lower dimension and then solved iteratively. In three dimensional positioning systems, the achievable reduction on the amount of computations in each iteration is 67%. On the other hand, accurate positioning results can be obtained with this low-complexity framework, especially with TDoA measurements.

4. A new non-iterative algorithm has been developed to improve the positioning accuracy of a well known existing direct method. Theoretical proof is given that the method provides a better estimator in the sense that it corresponds to an equal or smaller value of the original least-squares objective function. It does so by exploiting two similar fully constrained models. Meanwhile, the non-iterative nature makes the algorithm attractive for low cost, low power applications.

5. A review of several existing NLoS identification and mitigation schemes is provided, and we also propose new schemes based on systematic positioning model hypothesis testing. The idea is to combine the statistics of timing- and RSS-based range measurements, and in the mean time exploit the fact that all collected measurements are related to the same unknown position. Validation results show that the correct detection rate can go up to 99% and the corresponding positioning accuracy with NLoS mitigation is much better than without.

1.7 Thesis Outline

The remainder of the thesis is arranged in the following way.

In Chapter 2, the range models, for TDoA, ToA and RSS are given. These range models are widely used in the later positioning stage for both LoS and NLoS situations. Moreover, an UWB radio lab system was used and an audio testbed was developed in this project. The audio system has full three dimensional positioning capability. These two systems are described in this chapter. Hundreds of actual channel/range measurements are collected using these two systems and are utilized later to validate the theoretical work.

In Chapter 3, the positioning models used in this thesis are given, followed by a brief review of the non-linear least-squares theory. The study on the initial guess and the bias due to non-linearity is started in this chapter. Two schemes are proposed, for initial guess choice and bias check, respectively. Validation work is done in the next chapter based on the actual UWB radio and audio measurements, since knowledge of positioning algorithms is required, which are introduced in the next chapter.

In Chapter 4, a *structured* and *detailed* review of range-based least-squares *positioning methods* is provided, with special attention given to *direct methods*. The basic principles, assumptions and approximations of these methods are discussed, their *relations* are pointed out, their *theoretical performance* are analyzed, and the *properties* of different types of methods are summarized.

In Chapter 5, the mentioned framework for low complexity least-squares localization with high accuracy and the new non-iterative positioning method are presented. Theoretical error analysis is provided for both methods and validation work is performed using UWB radio timing-based range measurements.

In Chapter 6, a review of several relevant NLoS identification/mitigation techniques is provided for a better understanding of the NLoS problem, which helped the development of our own schemes, also presented in this chapter. The results are validated with both timing- and RSS-based range measurements, obtained using UWB radio signals.

In Chapter 7, concluding remarks are given together with recommendations and ideas for further work.

2 Measurement Models and Systems

As mentioned in the previous chapter, this PhD work focuses on range-based positioning, which is done by exploiting relations between timing and/or RSS measurements and the corresponding range.

Among many of the HERE project outcomes, several range estimation and modeling techniques have been documented. Since this thesis focuses more on the positioning stage, where ranges are already available, this chapter only serves as a brief description of

- the models used in the rest of this thesis to relate ToA, TDoA and RSS measurements to ranges,
- two experimental systems that are developed/used by the HERE project to obtain actual measurements for the validation of theoretical work presented in later chapters.

More detailed information on this subject is available in several papers of the project team [30, 36, 71].

2.1 Range Models

2.1.1 Time of Arrival

The ToA is the measured time at which a signal first arrives at a receiving end. It is widely used for range estimation. A review of the ToA estimation technologies in the context of UWB signals can be found, e.g. in [2].

2.1.1.1 Line of Sight

In a ToA system, the receiver (with unknown position) knows accurately when the transmitters (with known positions) started transmissions. The range between the i-th transmitter and the receiver can be obtained by multiplying the difference between the ToA measurement $\underline{\tau}_{ToA,i}$ and the Time of Transmission (ToT) $\tau_{ToT,i}$ with the speed of light c:

$$\underline{y}_i = c(\underline{\tau}_{ToA,i} - \tau_{ToT,i}) = d_i + \underline{e}_i, \tag{2.1}$$

where the underscores indicate randomness and the error term $\underline{e}_i$ accounts for the difference between the range measurement $\underline{y}_i$ and its true value d_i.

Under LoS conditions, the major sources of errors that cause this difference include additive noise and multipath effects [16] and $\underline{e}_i$ is usually modeled as a simple zero mean Gaussian variable [1, 72]: $\underline{e}_i \sim \mathcal{N}(0, \sigma_i^2)$, which gives the statistics of the range measurement as:

$$\underline{y}_i \sim \mathcal{N}(d_i, \sigma_i^2). \tag{2.2}$$

Assuming that the signal reached the receiver via an Additive White Gaussian Noise (AWGN) channel, the CRLB on the STD σ_i of the unbiased measurement $\underline{y}_i$ is given in [2, 73] as:

$$\sigma_i \geq \frac{c}{2\sqrt{2}\pi \mathsf{SNR}\beta}, \tag{2.3}$$

where SNR is the signal to noise ratio, β is defined as the effective signal bandwidth, calculated as: $\beta \triangleq \sqrt{\int_{-\infty}^{+\infty} f^2 \left|S(f)\right|^2 df \Big/ \int_{-\infty}^{+\infty} \left|S(f)\right|^2 df}$ and $S(f)$ is the Fourier Transform of the transmitted signal.

Equation (2.3) shows us that the lowest achievable variance of the range measurement obtained from ToA is inversely proportional to the effective signal bandwidth. This means that the very large bandwidth of UWB signals can be well exploited by ToA ranging [2]. In fact, lots of reported results in literature show that σ_i can go down to centimeter level, e.g. when ranging is done using UWB signals, either acoustic [74, 75] or radio [21, 72].

Equation (2.2) gives a simple Gaussian distribution of $\underline{y}_i$, where the two important parameters, the mean and the variance, are constant for all LoS situations. In literature, it has be reported that the mean and variance of $\underline{y}_i$ are dependent on the signal bandwidth and the true distance, even in LoS cases [36, 72, 76, 77]. The bandwidth and distance dependencies for ToA ranging errors can be explained as follows 1) the time resolution of the ToA ranging is related to signal bandwidth, see (2.3) and 2) the spatial separation between the direct path and multipaths has the tendency to decrease as the true distance becomes larger [76]. Due to the two listed reasons, a worse multipath resolvability is expected with a smaller signal bandwidth and/or a larger true distances.

The bandwidth and distance dependencies are rigorously investigated in [36] based on actual UWB data, and the statistical characterization is also provided. The corresponding models are given as follows:

$$\underline{y}_i(B, d_i) \sim \mathcal{N}(\mathsf{E}\{\underline{y}_i(B, d_i)\}, \sigma_i^2(B, d_i)), \tag{2.4}$$

where B in GHz is the signal bandwidth and

$$\mathsf{E}\{\underline{y}_i(B, d_i)\} = d_i + 0.0148 \exp(-B/0.48) d_i, \tag{2.5}$$

$$\sigma_i(B, d_i) = = 0.016(0.64 \exp(-B/0.6) d_i^{1.5} + 1). \tag{2.6}$$

Please note that the coefficients in (2.5)-(2.6) are dependent on the actual environment, and are empirically determined based on the actual UWB radio measurements obtained with a system introduced later in Section 2.2. Furthermore, these two equations require knowledge of the true distance, which is unavailable in practice. As suggested in [36], one solution is to use the range measurement instead.

2.1.1.2 Non-Line of Sight

In NLoS situations, (2.1) should be rewritten as:

$$\underline{y}_i = c(\underline{\tau}_{ToA,i} - \tau_{ToT,i}) = d_i + \mu_i + \underline{e}_i, \tag{2.7}$$

where μ_i is the extra NLoS bias due to the blockage of the direct path, and $\underline{e}_i$ accounts for the additive noise and multipath effects, which is usually modeled as a zero-mean Gaussian random variable with a certain variance, just like in LoS cases. One important note is that the bias μ_i in the range measurement is always non-zero in NLoS cases.

In literature, the NLoS bias μ_i is mainly treated in the following two ways:

1. μ_i is treated as an unknown deterministic parameter which can be estimated at the later positioning stage. In this case, one always needs check if the system is solvable, i.e. the number of measurements is no less than the number of unknown parameters.

2. μ_i is treated as an random variable that follows a certain distribution and the most widely used distribution is the exponential one [72, 78, 79, 80, 36].

In the first case, the measurement can be modeled as:

$$\underline{y}_i \sim \mathcal{N}(d_i + \mu_i, \sigma_i^2), \tag{2.8}$$

and in the second case, the measurement $\underline{y}_i$ is, next to the unknown true distance d_i, a sum of an exponentially distributed variable and a zero-mean Gaussian random variable. Ignoring bandwidth and distance dependencies, the bias and variance of $\underline{y}_i$ in both cases are constant for all measurements, and need to be obtained experimentally.

If bandwidth and distance dependencies are considered, one can calculate the mean and variance in both cases of $\underline{y}_i(B, d_i)$ use the equations e.g. in [36]:

$$\begin{aligned} \mathsf{E}\{\underline{y}_i(B, d_i)\} &= d_i + (0.027\exp(-B/0.47) + 0.013)d_i + 0.019, & (2.9)\\ \sigma_i(B, d_i) &= 0.049(0.21\exp(-B/0.73)d_i^{1.5} + 1) & (2.10) \end{aligned}$$

However, we should note that $\underline{y}_i$ is a worse approximation of d_i in NLoS cases than in LoS cases, with the presence of μ_i. Moreover, the environment-dependent coefficients in the above two equations are obtained experimentally.

2.1.2 Time Difference of Arrival

2.1.2.1 Line of Sight

In LoS cases, when the receiver clock is asynchronous to the transmitter clock, the ToA seen by the receiver should be written as:

$$\underline{\tau}_{ToA,i} = \underline{\tau}'_{ToA,i} + \tau_c, \tag{2.11}$$

where $\underline{\tau}'_{ToA,i}$ is the ToA according to the i-th transmitter clock, τ_c is the clock misalignment between the receiver clock and the transmitter clock. In this case, one can only get a *pseudo*range by multiplying (2.11) with c:

$$\underline{y}_i = c(\underline{\tau}_{ToA,i} - \tau_{ToT,i}) = d_i + b_c + \underline{e}_i. \tag{2.12}$$

where $b_c = c\tau_c$ needs to be treated as an unknown parameter in addition to the receiver position. As mentioned in Chapter 1, in a system where all the transmitters are synchronized, e.g. GPS, b_c is a common term in all pseudoranges and it is canceled in TDoA measurements. A TDoA measurement is the difference of two ToA measurements minus the known ToT difference of the two involved transmitters which can be related to ranges as

$$\underline{\Delta}_i = \underline{y}_i - \underline{y}_r = d_i - d_r + \underline{e}_i - \underline{e}_r. \tag{2.13}$$

After the differencing operation, the number of available measurements is reduced by 1, but the unknown b_c is also gone. One can either use the pseudoranges to estimate the position and the offset b_c, or use the pseudorange differences to estimate the position only.

The statistics of $\underline{y}_i$ can be given based on (2.2), one only needs to further consider a clock bias in the mean of $\underline{y}_i$:

$$\underline{y}_i \sim \mathcal{N}(d_i + b_c, \sigma_i^2). \tag{2.14}$$

For pseudoranges, the bandwidth and distance dependencies can be accounted for based on (2.4):

$$\underline{y}_i(B, d_i) \sim \mathcal{N}(\mathsf{E}\{\underline{y}_i(B, d_i)\}, \sigma_i^2(B, d_i)), \tag{2.15}$$

where

$$\begin{aligned} \mathsf{E}\{\underline{y}_i(B, d_i)\} &= d_i + b_c + 0.0148\exp(-B/0.48)d_i, & (2.16)\\ \sigma_i(B, d_i) &= = 0.016(0.64\exp(-B/0.6)d_i^{1.5} + 1). & (2.17) \end{aligned}$$

For a range difference, the statistics without bandwidth and distance dependencies are given based on (2.14) as:

$$\underline{\Delta}_i \sim \mathcal{N}(d_i - d_r, \sigma_i^2 + \sigma_r^2). \tag{2.18}$$

where $\underline{y}_i$ and $\underline{y}_r$ are assumed independent.

With bandwidth and distance dependencies, a range difference is also Gaussian distributed:

$$\underline{\Delta}_i(B, d_i, d_r) \sim \mathcal{N}\left(\mathsf{E}\{\underline{y}_i(B, d_i)\} - \mathsf{E}\{\underline{y}_i(B, d_r)\}, \sigma_i^2(B, d_i) + \sigma_r^2(B, d_r)\right), \tag{2.19}$$

where

$$\begin{aligned} \mathsf{E}\{\underline{y}_i(B, d_i)\} - \mathsf{E}\{\underline{y}_i(B, d_r)\} &= (d_i - d_r)[1 + 0.0148\exp(-B/0.48)], & (2.20)\\ \sigma_i^2(B, d_i) + \sigma_r^2(B, d_r) &= [0.016(0.64\exp(-B/0.6)d_i^{1.5} + 1)]^2 \\ &+ [0.016(0.64\exp(-B/0.6)d_r^{1.5} + 1)]^2. & (2.21) \end{aligned}$$

A practical problem of applying bandwidth and distance dependent models for the TDoA case is that it is now not feasible to use $\underline{y}_i$ in practice as the approximation for d_i, to calculate the TDoA variance, due to the fact that the absolute value of b_c can be quite large. Since $d_i - d_r$ can be approximated by $y_i - y_r$, one possible solution to this practical problem is to generate a model based on $d_i - d_r$ rather than d_i, following the same procedures in [36]. This, however, is not carried out in this thesis, and is left for future work.

2.1.2.2 Non-Line of Sight

In NLoS cases, based on (2.7), we can rewrite the pseudorange (2.12) as:

$$\underline{y}_i = d_i + b_c + \mu_i + \underline{e}_i, \tag{2.22}$$

where μ_i accounts for the NLoS bias.

Similarly, the range difference equation (2.13) becomes:

$$\underline{\Delta}_i = \underline{y}_i - \underline{y}_r = d_i - d_r + \mu_i - \mu_r + \underline{e}_i - \underline{e}_r. \tag{2.23}$$

The statistics of $\underline{y}_i$ and $\underline{\Delta}_i$ are discussed as follows:

1. If μ_i is treated as a deterministic unknown parameter, then the statistics of $\underline{y}_i$ can be given in a similar way as (2.8):

   $$\underline{y}_i \sim \mathcal{N}(d_i + b_c + \mu_i, \sigma_i^2), \tag{2.24}$$

   and

   $$\underline{\Delta}_i \sim \mathcal{N}(d_i - d_r + \mu_i - \mu_r, \sigma_i^2 + \sigma_r^2). \tag{2.25}$$

2. If μ_i is treated as a random variable, then $\underline{y}_i$ is, next to the true distance plus the clock bias, a sum of an exponentially distributed variable and a zero-mean Gaussian random variable, similar to the ToA case. Hence, $\underline{\Delta}_i$ becomes, next to the difference $d_i - d_r$, a sum of the difference between two exponentially distributed variables and the difference between two zero-mean Gaussian random variables, with all four variables being mutually independent.

Furthermore, if the bandwidth and distance dependencies are considered, the distributions of $\underline{y}_i$ in the two cases are not changed, and the mean and variance of $\underline{y}_i$ can then be calculated as:

$$\begin{aligned}
\mathsf{E}\{\underline{y}_i(B, d_i)\} &= d_i + b_c + (0.027\exp(-B/0.47) + 0.013)d_i + 0.019, && (2.26)\\
\sigma_i(B, d_i) &= 0.049(0.21\exp(-B/0.73)d_i^{1.5} + 1) && (2.27)\\
\mathsf{E}\{\underline{\Delta}_i(B, d_i)\} &= (d_i - d_r)[1 + (0.027\exp(-B/0.47) + 0.013)], && (2.28)\\
\sigma_i^2(B, d_i) + \sigma_r^2(B, d_r) &= [0.049(0.21\exp(-B/0.73)d_i^{1.5} + 1)]^2 \\
&+ [0.049(0.21\exp(-B/0.73)d_r^{1.5} + 1)]^2 && (2.29)
\end{aligned}$$

Please note that in NLoS cases, the practical problem of approximating d_i with $\underline{y}_i$ also exists.

2.1.3 Received Signal Strength

For both LoS and NLoS cases, the RSS can be modeled as a linearly decreasing variable on a logarithmic scale of the distance [81, 82].

2.1.3.1 Line of Sight

Under LoS, the major sources of errors in RSS measurements include multipaths and shadowing effects [16]. The multipaths arrive at the receiver with different phases and cause the frequency selective fading effect. Fortunately, this effect can be greatly mitigated by using signals with large bandwidth [83], e.g. UWB, and then the errors in the RSS measurements are mainly caused by the shadowing effect. Thus, the RSS from the i-th transmitter can be formulated as

$$\underline{P}_i(d_i) = P_0 + 10n_i \log_{10} \frac{d_i}{d_0} + \underline{s}_i \tag{2.30}$$

where $\underline{P}_i(d_i)$ is the measured received power in dBm at distance d_i, d_0 is a reference distance, P_0 is the reference power received at the reference distance, n_i is the pathloss exponent, and $\underline{s}_i$ accounts for the shadowing effect. Typically, $\underline{s}_i$ is modeled as a zero mean Gaussian random variable in dB [10, 28, 29]: $\underline{s}_i \sim \mathcal{N}(0, \sigma_{s_i}^2)$, which gives:

$$\underline{P}_i(d_i) \sim \mathcal{N}(P_0 + 10n_i \log_{10} \frac{d_i}{d_0}, \sigma_{s_i}^2) \tag{2.31}$$

The model (2.30) can be used to estimate the distance between the i-th transmitter and the receiver as:

$$\underline{y}'_i = d_0 10^{\frac{\underline{P}_i(d_i) - P_0}{10n}} = d_i e^{\frac{\underline{s}_i \ln 10}{10n_i}}, \tag{2.32}$$

with mean and variance given as:

$$\mathsf{E}\{\underline{y}'_i\} = d_i e^{\frac{1}{2}(\frac{\sigma_{s_i} \ln 10}{10n_i})^2} \tag{2.33}$$

$$\sigma'^2_i = d_i^2 e^{(\frac{\sigma_{s_i} \ln 10}{10n_i})^2} (e^{(\frac{\sigma_{s_i} \ln 10}{10n_i})^2} - 1). \tag{2.34}$$

Based on (2.33), range measurement $\underline{y}'_i$ can be made bias free by pre-multiplying by $e^{-\frac{1}{2}(\frac{\sigma_{s_i} \ln 10}{10n_i})^2}$:

$$\underline{y}_i = \underline{y}'_i e^{-\frac{1}{2}(\frac{\sigma_{s_i} \ln 10}{10n_i})^2} = d_i e^{\frac{\underline{s}_i \ln 10}{10n_i}} e^{-\frac{1}{2}(\frac{\sigma_{s_i} \ln 10}{10n_i})^2} \tag{2.35}$$

and the mean and variance becomes:

$$\mathsf{E}\{\underline{y}_i\} = d_i \tag{2.36}$$

$$\sigma_i^2 = d_i^2 (e^{(\frac{\sigma_{s_i} \ln 10}{10n_i})^2} - 1). \tag{2.37}$$

Since $\underline{s}_i$ is a Gaussian random variable in dB, the RSS-based range measurement is a log-normal random variable:

$$\underline{y}_i \sim \text{Log}\mathcal{N}(d_i, \sigma_i^2), \tag{2.38}$$

with $\text{E}\{\underline{y}_i\}$ and σ_i given in (2.36)-(2.37).

Please note that from now on, to distinguish the two different type of measurements (with different distributions) in (2.30) and (2.35), we will refer to the measurement in (2.30) as *RSS* and the one in (2.35) as *RSS-range*.

Based on (2.38), the error in a RSS-range measurement, $\underline{e}_i = \underline{y}_i - d_i$ is equal to a log-normal random variable plus a constant; its mean is 0 and its STD is equal to σ_i.

The CRLB on the RSS-based range measurement STD σ_i is given in [73] as:

$$\sigma_i \geq \frac{\ln 10}{10} \frac{\sigma_{s_i}}{n} d_i. \tag{2.39}$$

The above equation shows us that the lowest achievable STD of RSS-based ranging is related to the passloss exponent, the shadowing effect and the true distance, but not to the signal bandwidth. Hence, unlike timing-based ranging, RSS-based ranging using the pathloss model (2.30) does not fully exploit the very large bandwidth of UWB, though the frequency selective fading effect is mitigated.

As discussed in Section 2.1.1, the multipath resolvability is dependent on the signal bandwidth and the true distance. Hence, the pathloss exponent and the shadowing effect in (2.30) are expected to be bandwidth and distance dependent. However, at the time the main content of this thesis was finished, a bandwidth and distance dependent model for RSS-ranging (like (2.5)-(2.6)) was still not available. Since developing such a model is beyond the scope of this thesis, only (2.30) is used in later chapters.

2.1.3.2 Non-Line of Sight

The RSS pathloss model in the NLoS case used in this thesis has exactly the same structure as (2.30), and $\underline{s}_i$ is also modeled as a zero-mean Gaussian random variable. The differences in parameters include that: 1) the absolute value of the pathloss exponent n_i is much larger in NLoS cases than in LoS cases, due to the blockage of the direct path and 2) the STD of $\underline{s}_i$ is also larger. More advanced asymmetric models, such as the Nakagami model, are not considered in this thesis for the following reasons:

- the positioning algorithms on which we focus are based on the non-linear least-squares theory. This means that the mean and variance of the range measurements play an important role in weighted cases while the actual distribution is less important.
- the distribution of NLoS RSS measurements is of importance, e.g. for the LoS/NLoS identification in Chapter 6. However, to distinguish between a LoS and a NLoS RSS measurement, the left side of the NLoS RSS distribution that corresponds to negative errors is more important, and can be well approximated using a simple log-normal model [80].

2.2 Measurement Systems

The actual UWB measurements used in this thesis are obtained using two measurement systems, one based on radio signals and one based on audio signals. The two systems are introduced in the following.

2.2.1 Radio System

2.2.1.1 System Setup

The radio system has one transmitter and one receiver, i.e. the system is only capable of 1-D ranging. These limitations arise from the size, price and availability of the system hardware.

The setup of the radio UWB system is shown in Figure 2.1. The generator fires a Gaussian-

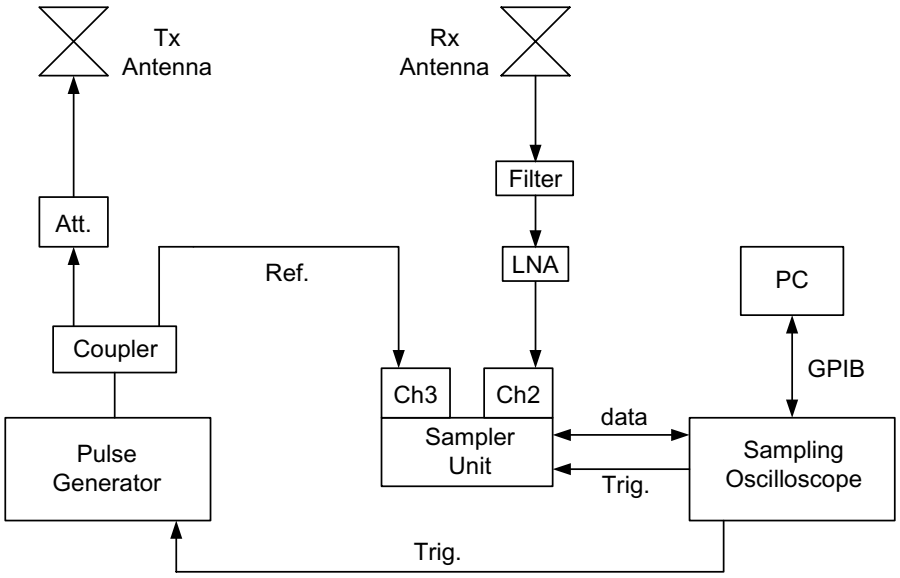

Figure 2.1: Setup for UWB radio system.

like pulse with a duration of 50 ps. A 6 dB attenuator is placed before the transmit antenna to avoid multiple antenna reflections. The receiver part includes a sampling oscilloscope, a sampler unit and a PC. The sampling oscilloscope controls both the sampling unit and the pulse generator with trigger pulses, implying that the transmitter and the receiver are synchronized. The sampling unit samples the received signals using a stroboscopic sampling mechanism [37], which has a virtual sampling rate of 100 GHz. A bandpass filter in the 3-20 GHz bandwidth, to reduce out of band interference and noise, and an LNA, working in the band between 0.1 and 12 GHz, follow the receiver antenna. As a result, it is possible to cover the bandwidth between 3.1 and 10.6 GHz allowed by the FCC for UWB radio transmissions [11]. Both the transmitter and the receiver use vertically polarized, omni-directional biconical antennas.

The Channel Impulse Response (CIR) $h(t)$ is estimated by deconvolving the received signal

Table 2.1: The intervals of zenith angle θ and azimuth angle ϕ.

i	θ	ϕ
1	$[\frac{\pi}{8} \quad \frac{3\pi}{8}]$	$[\frac{\pi}{8} \quad \frac{3\pi}{8}]$
2	$[\frac{5\pi}{8} \quad \frac{7\pi}{8}]$	$[\frac{5\pi}{8} \quad \frac{7\pi}{8}]$
3	$[\frac{9\pi}{8} \quad \frac{11\pi}{8}]$	$[\frac{\pi}{8} \quad \frac{3\pi}{8}]$
4	$[\frac{13\pi}{8} \quad \frac{15\pi}{8}]$	$[\frac{5\pi}{8} \quad \frac{7\pi}{8}]$
5	$[0 \quad \pi]$	$[0 \quad 2\pi]$

in the frequency domain using the inverse filtering technique. The spectrum of the received signal is divided by the one of the reference signal, measured at a distance of one meter, in the absence of reflections. In this way, it is possible to take into account the transmit and receive antenna transfer functions and the characteristics of the other system components. The division is done only for the parts of the spectrum within the frequency band of interest. The rest of the spectrum is filled with zeros. Taking the complex baseband Inverse Fast Fourier Transform (IFFT), the complex impulse response of the channel is obtained. To reduce the leakage problem when transforming the signal back to the time domain, a Hamming window is used, which provides sidelobes less than -43 dB. Note that the inverse filtering and the Hamming window are applied for easy and fast processing of the data.

The ToA is determined as the arrival time of the first path of the CIR. The first path is defined here as the first local maximum of the envelope of the estimated channel, with an amplitude within 20 dB from the strongest peak. The distance between the transmitter and the receiver is then computed using (2.1).

2.2.1.2 Measurement Setup

The measurements were carried out in the faculty building of EEMCS of Delft University of Technology. Both the transmitter and the receiver antenna are placed 1.5 m above the floor. The distances between the transmitter and the receiver are between 2 and 15 m for LOS, and between 3 and 12 m for NLOS. The physical obstruction under NLOS is caused by walls, between the transmitter and the receiver. The walls in the building are made of concrete or brick and the floors are of reinforced concrete. The measurements have been distributed between the minimum and maximum distance, to check distance dependency of the ranging results. In total, 321 LoS and 157 NLoS measurements are collected, corresponding to different transmitter and receiver positions. The corresponding true distances, measured with a laser-disto, are also recorded.

As the UWB radio measurement system is restricted to a single link (1 Tx, 1 Rx, synchronized), a 3-D positioning system is created virtually. Without loss of generality, the receiver is placed at the origin. From the total of actual range measurements, distributed in 2-15 m, a subset of 5 estimates is randomly chosen. For the i-th transmitter, spherical coordinates (d_i, θ_i, ϕ_i) are assigned, see Figure 2.2, where θ_i and ϕ_i are randomly chosen from the intervals shown in Table 2.1. Problematic geometries that make the positioning unsolvable, e.g. with all transmitters on a line, are avoided.

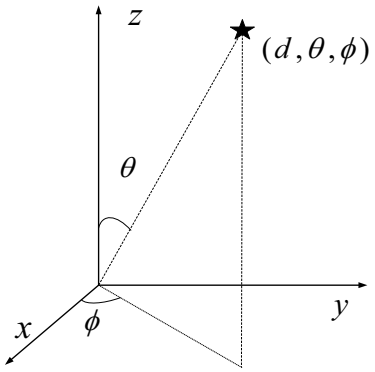

Figure 2.2: Spherical coordinates of a single measurement. The receiver is at the origin and the transmitter at (d, θ, ϕ).

2.2.1.3 Results

In this section, the ToA and RSS measurements that will be used in later chapters are given. The Gaussian fit for the ToA measurement and the log-normal pass loss model for the RSS measurements are also shown. Please note that the RSS value refers to the received power of the first path, which corresponds to the ToA. Low complexity way to measure the first path power was proposed in [71]. The reason to use first path power instead of traditional total power is because the former is directly related to the direct path under LoS situation. Hence, better ranging and positioning results can be obtained [84].

ToA The ToA range error, calculated as the difference between the range measurement obtained using (2.1) and the corresponding ground truth, is shown in Figure 2.3 for both LoS and NLoS cases. In the figure, only measurements obtained using the full 7.5 GHz UWB radio bandwidth are shown; the mean and STD of all the measurements together, obtained with 0.5 GHz are given in Table 2.2.

The LoS ToA range measurements obtained with 7.5 GHz of bandwidth are highly accurate, the errors are with zero mean and 1.5 cm STD. The CDF obtained with actual data is well modeled by a Gaussian CDF with mean and STD equal to those of the actual data. In the NLoS case, the measurements are biased due to the blockage and worse multipath effects, and the error STD is larger than for the LoS case. Moreover, the Gaussian model does not fit the data CDF as good as in the LoS case. However, the NLoS range error STD obtained with 7.5 GHz of bandwidth is even smaller than the LoS range error STD obtained with 0.5 GHz. This already shows that the ToA range errors are bandwidth dependent. The worse performance of the ranging results with 0.5 GHz of bandwidth is due to fact that the time domain resolution is reduced and ranging results are more affected by multipaths.

Please note that the verification of the bandwidth and distance dependent model is beyond the scope of this thesis, and we refer interested readers to [36].

RSS The RSS values obtained with 7.5 GHz of bandwidth are shown in Figure 2.4, together with the corresponding pathloss models for both LoS and NLoS cases. The pathloss exponents are estimated using the least-squares curve-fitting method, and the corresponding values are given in Table 2.3, together with the mean and STD (obtained with all measurements in LoS and NLoS, respectively) of the shadowing parameter s.

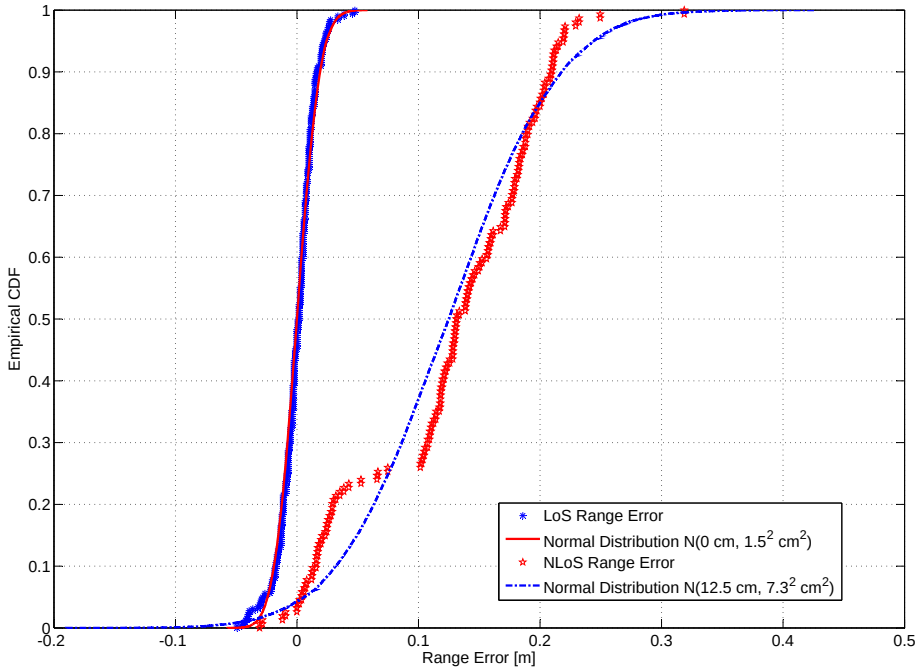

Figure 2.3: CDFs of the UWB radio LoS and NLoS ToA range errors compared with the corresponding Gaussian CDFs. The Gaussian CDFs are generated with equal mean and variance as the corresponding range errors. The measurements are obtained with full 7.5 GHz of bandwidth. The bumps in the NLoS range error curve is due to the fact that the measurements are obtained in different indoor environments, e.g. regular office, laboratory, corridor, with different blocking materials.

For 7.5 GHz of bandwidth, it can be observed from Figure 2.4 that the LoS RSS values are symmetrically distributed around the model, while in the NLoS case, the distribution of the RSS is a bit asymmetric. One important conclusion, however, is that the RSS values obtained under LoS and NLoS are well separated, meaning a potential for good LoS/NLoS identification based on RSS. This will be investigated further in Chapter 6.

In Table 2.3, for both tested bandwidths, the listed absolute value of the pathloss exponent in the LoS case is much smaller than the one in the NLoS case, as expected. Moreover, the pathloss under 7.5 GHz of bandwidth is greater than the loss under 0.5 GHz. This is due to the fact that in 0.5 GHz the first path power contains more contributions from the unresolved multipaths.

Table 2.2: Mean and STD of the ToA range errors under 7.5 and 0.5 GHz UWB radio bandwidths.

Bandwidth [GHz]	Condition	Mean [cm]	STD [cm]
7.5	LoS	0.00	1.50
	NLoS	12.50	7.30
0.5	LoS	2.03	11.01
	NLoS	18.97	16.50

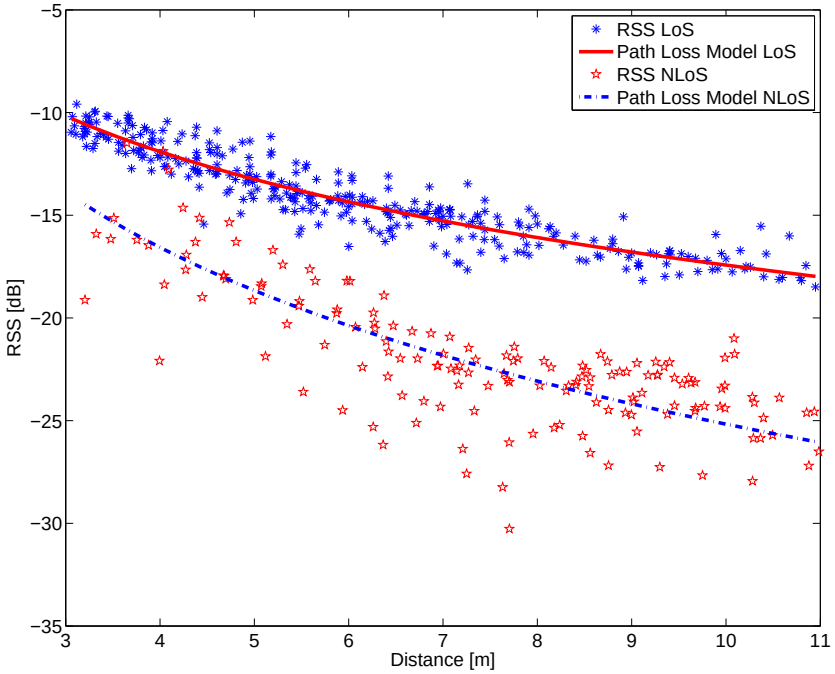

Figure 2.4: Actual UWB radio RSS measurements and log-normal pass loss models obtained under LoS and NLoS conditions, respectively. The signal bandwidth is 7.5 GHz.

Table 2.3: Pathloss exponent n, mean and STD of the shadowing parameter s under 7.5 and 0.5 GHz UWB radio bandwidths.

Bandwidth [GHz]	Condition	n	Mean of s [dB]	STD of s [dB]
7.5	LoS	-1.91	0.00	0.64
	NLoS	-2.94	-0.02	4.83
0.5	LoS	-1.81	0.00	2.11
	NLoS	-2.72	-0.03	4.91

2.2.2 Audio System

The reason we built an audio system is that building a radio UWB system with full positioning functionality is a difficult task at the current stage, due to the demanding hardware requirements and lack of mature components on the market. A test platform based on audio signals is attractive since it is well affordable and allows to have the flexibility required for research purposes.

2.2.2.1 Principle of Frequency Scaling

In order to get a similar interaction with the environment, the wavelengths of the audio signal should be comparable to the wavelengths of the radio signal. By down-scaling the frequency with a factor $\alpha = c/c_s \cong 8.7 \times 10^5$, where c and c_s are the speed of light and the speed of sound in air, respectively, the same wavelengths for the audio signal as for the

corresponding radio UWB signal can be obtained. Thus, the radio frequency band from 3.1 to 10.6 GHz, assigned by FCC for UWB communications [11], corresponds to the audio band from 3.563 to 12.184 kHz.

The audio system is only used for tests under LoS situations. NLoS situations are not tested, since the audio impedance of building materials is much higher than radio impedance, and the ability of audio signals to propagate through building materials is really poor. However, an audio system is still attractive in practice, given the fact that it is easy to implement and the components are cheap and widely available in the market, a lot of loudspeakers can be installed to provide a significant number of LoS signals.

Figure 2.5 shows the estimated impulse response for a radio UWB signal, when the transmitted pulse occupies the bandwidth between 3.1 and 10.6 GHz, together with impulse response for an audio signal occupying the band from 3.563 to 12.184 kHz, obtained with similar positions of the transmitter and the receiver. The the time axis of the audio CIR is scaled down by α. The analogies in the power delay profile (before 20 ns) are evident.

2.2.2.2 System Setup

The audio Ultra-Wideband system is controlled by a standard PC running Windows XP. The PC drives a 12-channel sound card, which can support up to 11 loudspeakers and 11 microphones. One input-output pair is reserved for a reference loop, which directly connects the output to the input. The corresponding received signal from the loop is used as the reference to correct the random delay added by the Windows XP operating system. Figure 2.6 shows the setup of the audio system. Currently, we use multiple loudspeakers as reference points with known position, to estimate the unknown position of the microphone.

The transmitted signal uses a Continuous Wave Frequency Modulation (FMCW), which is well known from radar applications [85]. The signal is sinusoidal, and its frequency linearly increases from 3.563 to 12.184 kHz, within a pulse duration of $2^L/f_s$ seconds, i.e. the chip rate is $f_s(12184 - 3563)/2^L$ Hz/s. Here, 2^L is the number of samples, and f_s is the sampling frequency of the sound card, chosen as $2^L = 2^{13}$ and $f_s = 48$ kHz by default, respectively.

In the audio system, the ToA is obtained using the same Inverse Filtering plus Hanmming window technique as with the radio system. The difference is that the reference signal is now obtained from the reference loop. The distance between the transmitter and the receiver is also computed using (2.1), but c needs to be replaced with c_s here and the delay added by the operating system has to be subtracted from ToT, based on the reference signal. Since c_s is not exactly constant in practice, its value is calculated using the equation proposed in [86] to obtain more accurate results,

$$c_s = 331.4 + 0.6T + 0.012H, \tag{2.40}$$

where T in [°C] and H in [%] account for the temperature and the humidity of the air. Alternatively, it is possible to use a calibration phase for c_s, by measuring the ToA between two reference nodes with a known distance.

To perform full 3-D positioning, multiple loudspeakers use a TDMA scheme; each loud-

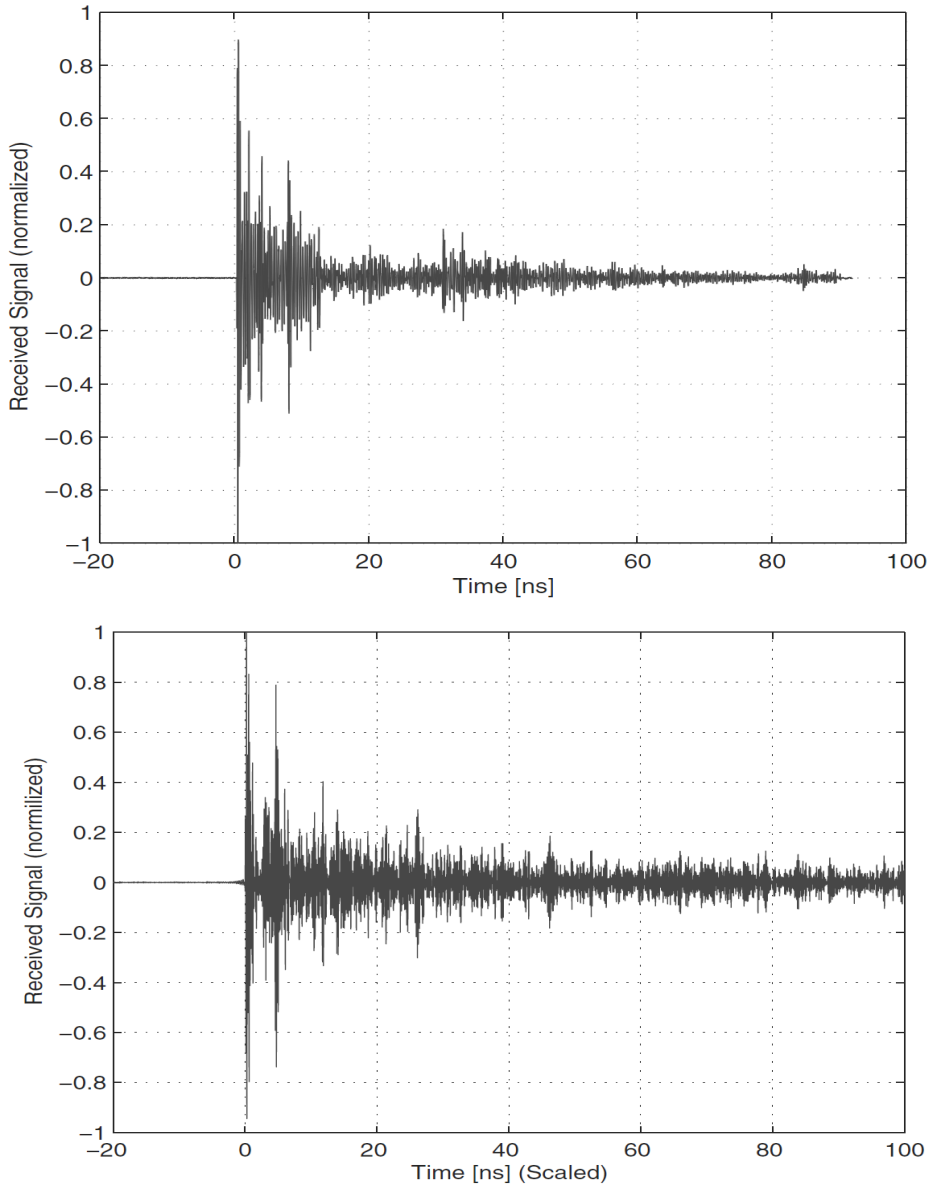

Figure 2.5: Estimated impulse response for the radio (upper) and audio (lower) UWB case.

speaker has been assigned a time slot in which the transmission is allowed and the microphone is stationary. In our system, the loudspeakers and the microphone are actually synchronized, since they are driven by the same PC. This is good for analysis of fully synchronized systems and RSS-based systems. For the study of TDoA systems with a clock offset, we simply assume that there exists a common clock offset in the system, with true value equal to 0, and it is treated as an additional unknown parameter. In this way, this system can fit in a more general framework for range-based positioning systems.

Another issue in the system setup is that the actual acoustic center of a loudspeaker is not easy to observe based on its physical dimensions. On the contrary, this issue is of much less importance for the much smaller microphone. So, for better accuracy in determining the positions of the loudspeakers prior to the measurements for the experiment, inverse positioning has been applied, i.e. the positions of the loudspeakers are determined through

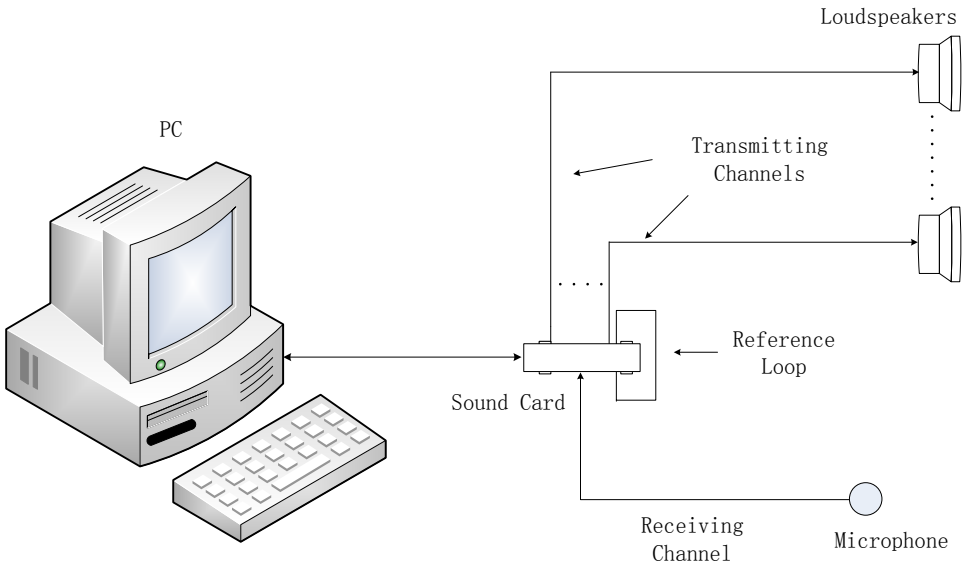

Figure 2.6: Setup for UWB audio system.

measurements at several accurately measured positions of the microphone. Under the full bandwidth (3.563 to 12.184 kHz), the STDs of the obtained position estimators of the loudspeakers are at cm-level.

2.2.2.3 Measurement Setup

The system is located in an office of the EEMCS building of Delft University of Technology. The range measurements are obtained in four different system geometries. Meanwhile, for each geometry, the measurements are obtained under a variety of signal bandwidths in order to have different range measurement accuracies. The tested bandwidths are the Full Bandwidth (FB) 8.5 kHz, 4.25 kHz (0.5FB), 1.7 kHz (0.2FB) and 0.567 kHz (0.0667FB). Note that a bandwidth of 0.567 kHz for audio signals corresponds to a bandwidth of 500 MHz for radio signals, which is the smallest bandwidth[1] for a signal to be recognized as UWB [11]. With one specific geometry and a certain bandwidth, 128 measurements (microphone positions) are collected. In Geometry 1, 2, and 3, four loudspeakers are installed and in Geometry 4, one additional loudspeaker is installed. The microphone is placed at a height (z) of 1.605 m above the ground. In the x-y plane, in Geometry 1, 2 and 4, the microphone is placed inside the square formed by the loudspeakers and in Geometry 3, the microphone is placed outside. The positions of the loudspeakers and the microphone in all the systems geometries are shown in Figure 2.7 and Figure 2.8, respectively.

2.2.2.4 Ranging Results

Figure 2.9 gives the CDF of the range error for all the measurements collected with the full bandwidth. From the figure, it can be seen that the largest error is about 6 cm, considering all positions. A Gaussian distribution with mean $\mu = 0$ cm and STD $\sigma = 1.15$ cm fits the

[1]The part of the definition of UWB signals regarding the fractional bandwidth is not considered here.

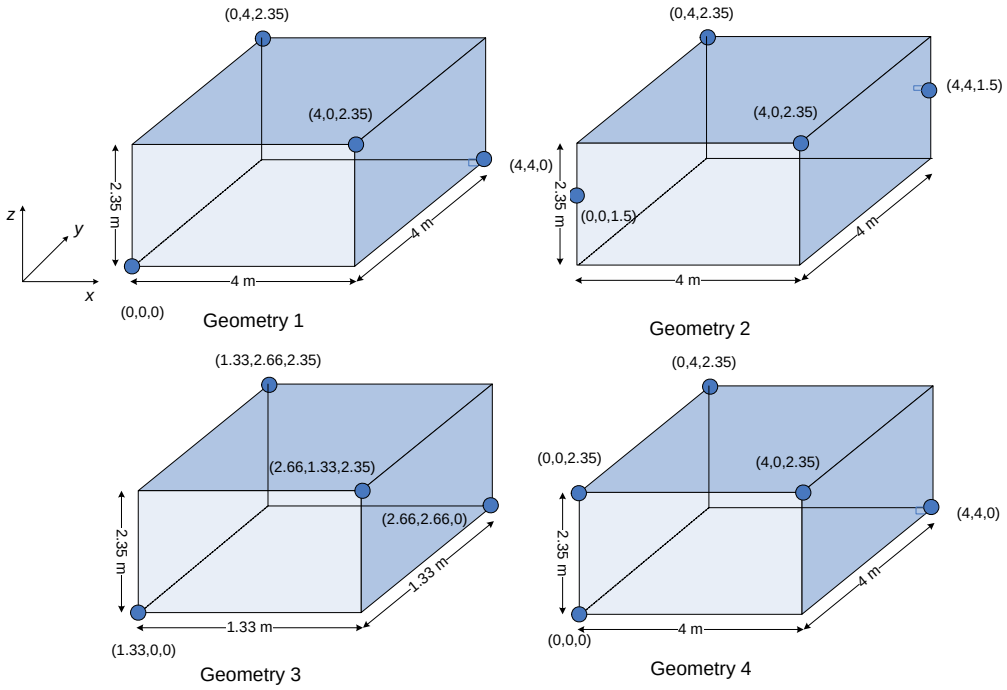

Figure 2.7: Positions of the loudspeakers in four tested system geometries.

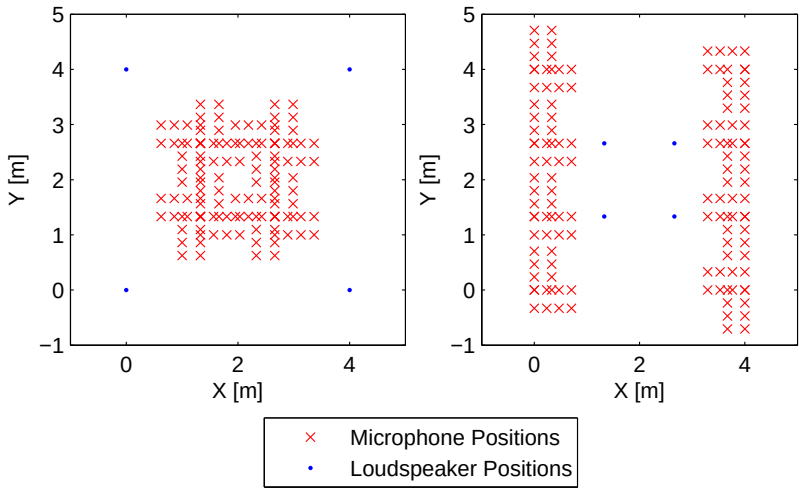

Figure 2.8: Positions of the microphone and the loudspeakers in four tested system geometries on the horizontal x-y plane, with Geometry 1, 2 and 4 on the left and Geometry 3 on the right.

Table 2.4: Mean and variance of the ToA range errors under different UWB audio bandwidths.

Bandwidth [kHz]	Mean [cm]	Standard Deviation [cm]
8.500	-0.03	1.15
4.250	-0.46	1.75
1.700	-0.49	3.41
0.567	1.37	11.86

empirical CDF well. The STD of audio measurements is less than the radio measurements since 1) the largest true distance of all the audio measurements is less than 8 m, while for radio measurements, the value is 15 m, and 2) ranging STDs are distance dependent.

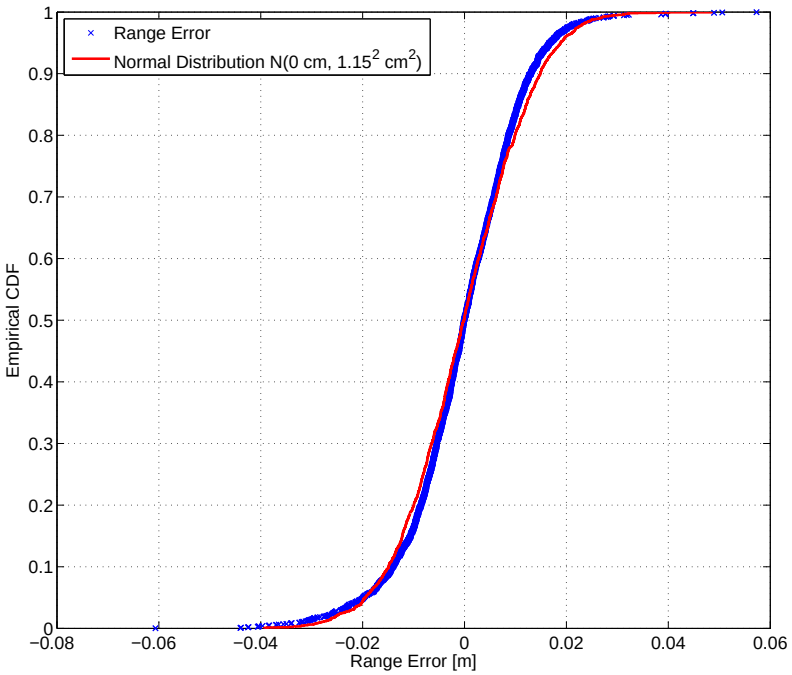

Figure 2.9: CDF of the UWB audio ToA range error with full bandwidth 8.5 kHz, including all geometries.

When the bandwidth is decreased, the quality of range estimation degrades. The values of the mean and the STD of the range error under different bandwidths are listed in Table 2.4. The reason of the degradation is simply that smaller bandwidths correspond to poorer resolution in the time domain. So, multipaths received at the microphone become less resolvable and even overlap with the first path. This agrees with the trend in the bandwidth dependent range model developed in [36] based on radio signals, where non-zero mean and large variance of the range errors are expected as the bandwidth is reduced.

2.3 Summary

In this chapter, the models for timing and RSS measurements are described. Two measurement systems, one based on radio and one based on audio, are introduced and they have been used to generate the measurements used in later chapters. The audio system can even perform full 3-D positioning. As will be shown in Chapter 3, the achieved positioning accuracy in terms of RMSE is at centimeter level.

The models for ToA and RSS measurements that will be used in later chapters are summarized in Table 2.5.

Table 2.5: Summary of the ToA and RSS models.

Measurement	Simple	Advanced
ToA-range/LoS	Unbiased Gaussian (2.2); variance obtained empirically; bias (zero) and variance treated as constants	Biased Gaussian; bias (2.5) and variance (2.6) dependent on bandwidth and distance
ToA-range/NLoS	Biased Gaussian (2.8); bias μ as unknown parameter; variance obtained empirically, and treated as constant	Biased Gaussian, bias (2.9) and variance (2.10) dependent on bandwidth and distance
TDoA-range/LoS	Unbiased Gaussian (2.14); variance obtained empirically; bias (zero) and variance treated as constants	Biased Gaussian; bias (2.16) and variance (2.17) dependent on bandwidth and distance
TDoA-range/NLoS	Biased Gaussian (2.24); bias μ as unknown parameter; variance obtained empirically, and treated as constant	Biased Gaussian, bias (2.26) and variance (2.27) dependent on bandwidth and distance
RSS/LoS	Unbiased Gaussian in dB (2.30)	
RSS/NLoS	Unbiased Gaussian in dB (2.30)	
RSS-range/LoS	Unbiased log-normal (2.35)	
RSS-range/NLoS	Unbiased log-normal (2.35)	

The relations between the listed models and later chapters are given in Table 2.6.

The relations between the different measurements and the later chapters are given in Table 2.7. Please note that the advanced TDoA-pseudorange model is never used, due to the practical problem of finding a good approximation for the true distance d_i. Meanwhile, the NLoS RSS-range model and measurements are not used in any chapter, due to the fact that in NLoS cases, the RSS measurement, rather than the RSS-range measurement, is

Table 2.6: Relations between the listed models in Table 2.5 and the later chapters.

Models	Chapter 3	Chapter 4	Chapter 5	Chapter 6
ToA-range/LoS Simple	✓	✓	✓	✓
ToA-range/LoS Advanced		✓	✓	✓
ToA-range/NLoS Simple				✓
ToA-range/NLoS Advanced				✓
TDoA-pseudorange/LoS Simple	✓	✓	✓	✓
TDoA-pseudorange/LoS Advanced				
TDoA-pseudorange/NLoS Simple				✓
TDoA-pseudorange/NLoS Advanced				
RSS/LoS Simple	✓	✓	✓	✓
RSS/NLoS Simple				✓
RSS-range/LoS Simple	✓	✓	✓	
RSS-range/NLoS Simple				

Table 2.7: Relations between the measurements and the later chapters.

Models	Chapter 3	Chapter 4	Chapter 5	Chapter 6
ToA-range/LoS	Audio	Audio and Radio	Radio	Radio
ToA-range/NLoS				Radio
TDoA-pseudorange/LoS	Audio	Audio and Radio	Radio	Radio
TDoA-pseudorange/NLoS				Radio
RSS/LoS				Radio
RSS/NLoS				Radio
RSS-range/LoS			Radio	
RSS-range/NLoS				

directly used for positioning purpose. The details can be found in Chapter 6.

3 Range-based Positioning and Least-squares

The range d_i between an unknown user position x_u and a base station x_i equals:

$$d_i = \sqrt{(x_u - x_i)^T (x_u - x_i)}. \tag{3.1}$$

Based on this simple but non-linear equation, together with the measurement models discussed in the previous chapter, the positioning models can be built.

We will divide positioning models into two major groups:

1. Models on original observables. These models directly use the original range measurements or their linear combinations.
2. Models on squared observables. These models use squared range measurements or their linear combinations.

Please note that in this chapter and in later Chapter 3 ,4 only LoS measurements and models are considered. The NLoS case will be addressed in Chapter 6.

After the positioning models are introduced, the estimation criteria, especially the least-squares theory, for positioning will be discussed, with focus given to indoor scenarios.

As a general rule, for any vector a, we will use the notation $\underline{\hat{a}}$ to represent its *estimator* and the notation $\hat{a}$ for its estimate. The estimator is a random (denoted with the underscore) vector while the estimate is a deterministic vector, which is a realization of the estimator.

3.1 Models on Original Observables

The models on original observables represent the intrinsic relation between the user position and a set of range measurements. They can be further divided into two groups, depending on the number of involved differencing operations.

3.1.1 No Differencing

Suppose that m measurements are collected, based on either timing or RSS information. Putting these measurements into a vector y, the original model with no differencing can be given as

$$\underline{y} = A(x) + \underline{e}, \tag{3.2}$$

Table 3.1: Unknown parameters for different systems.

unknowns	system
$x_{(n+1)\times 1} = [x_u^T,\ b_c]$	In case of TDoA with m pseudorange equations
$x_{n\times 1} = x_u$	In case of ToA/RSS with m range equations ($b_c = 0$), or TDoA with $m-1$ pseudorange differences (b_c is canceled)

where the measurement vector $\underline{y}$ is in the m-dimensional space of real numbers $\mathbb{R}^m$, $x \in \mathbb{R}^n$ is the unknown vector, cf. the details in Table 3.1, $A(.) : \mathbb{R}^n \to \mathbb{R}^m$ is a non-linear mapping, and $\underline{e} \in \mathbb{R}^m$ is the measurement error vector.

For each transmitter-receiver link, the details of the model can be obtained by replacing d_i with $\|x_u - x_i\|$ in (2.1), (2.12), (2.30), and (2.35), and the results are given as follows:

- For one single timing-based range measurement, either ToA (2.1) or TDoA (2.12), the equation reads

$$\underline{y}_i = A_i(x) + \underline{e}_i = \underbrace{\|x_u - x_i\|}_{d_i} + b_c + \underline{e}_i,\ i = 1 \dots m, \tag{3.3}$$

  where $A_i(x) = d_i + b_c$, x_u is the unknown position, x_i is the position vector of the i-th base station, b_c represents the clock offset, and $\|.\| = \sqrt{(.)^T(.)}$. Note that in ToA cases, $b_c = 0$.

- In case of a RSS measurement (2.30), it holds:

$$\underline{y}_i = A_i(x) + \underline{e}_i = \underline{P}_i = P_0 + 10n_i \log_{10} \frac{\|x_u - x_i\|}{d_0} + \underline{s}_i,\ i = 1 \dots m. \tag{3.4}$$

  where $A_i(x) = A_i(x_u) = P_0 + 10n_i \log_{10} \frac{\|x_u - x_i\|}{d_0}$ and $\underline{e}_i = \underline{s}_i$. Please note that here we have $x = x_u$ since there is no clock bias in a RSS scheme.

- For a RSS-based range measurement (2.35), we have:

$$\underline{y}_i = A_i(x) + \underline{e}_i = \|x_u - x_i\| e^{\frac{-s_i \ln 10}{10 n_i}} e^{-\frac{1}{2}(\frac{\sigma_{s_i} \ln 10}{10 n_i})^2},\ i = 1 \dots m, \tag{3.5}$$

  where $A_i(x) = A_i(x_u) = \|x_u - x_i\|$ and $\underline{e}_i = \|x_u - x_i\|(e^{\frac{-s_i \ln 10}{10 n_i}} e^{-\frac{1}{2}(\frac{\sigma_{s_i} \ln 10}{10 n_i})^2} - 1)$. Please note that in the above equation, the presence of the term $e^{-\frac{1}{2}(\frac{\sigma_{s_i} \ln 10}{10 n_i})^2}$ is to make $\underline{y}_i$ unbiased.

Please note that the term 'range-based' in the title of this chapter refers to the fact that the relation (3.1) is exploited. Hence, positioning techniques that directly use RSS measurements with model (3.4), not RSS-range measurements, are also range-based.

3.1.1.1 System Solvability

Position estimation can be seen as an inverse problem to find a mapping $M(.) : \mathbb{R}^m \to \mathbb{R}^n$, such that the estimator of x can be represented as $\hat{\underline{x}} = M(\underline{y})$. The solvability of the problem can be analyzed by first Taylorizing $A(x)$ at the estimate $\hat{x}$,

$$\underline{y} = A(\hat{x}) + \partial_{x^T} A(\hat{x})(x - \hat{x}) + o(x - \hat{x}) + \underline{e}, \tag{3.6}$$

where the Landau order term $o(x-\hat{x})$ indicates the remainder in the Taylor series, which can be ignored when $\hat{x}$ is close to x The linear part of (3.6) can be used as the approximation of the original equations in (3.2). With the linear approximation, the solvability of the inverse problem can be described as follows:

1. Underdetermined: $m < \text{rank}(\partial_{x^T} A(\hat{x}))$. There exists an infinite number of solutions.
2. Consistent: $m = \text{rank}(\partial_{x^T} A(\hat{x}))$ and/or $\underline{e} = 0$ (error free measurements). The solution can be used to reproduce y. The solution is unique, if and only if the column vectors of $\partial_{x^T} A(\hat{x})$ are linearly independent, i.e. $\text{rank}(\partial_{x^T} A(\hat{x})) = n$.
3. Overdetermined: $m > \text{rank}(\partial_{x^T} A(\hat{x}))$. In this case, there are more equations than strictly needed for determining the solution. The number $m - \text{rank}(\partial_{x^T} A(\hat{x}))$ is referred to as the redundancy of the system.

In Figure 3.1, the geometric interpretations of consistent ToA/RSS systems and TDoA systems are given with $x_u \in \mathbb{R}^2$ for a better demonstration, which can be described as:

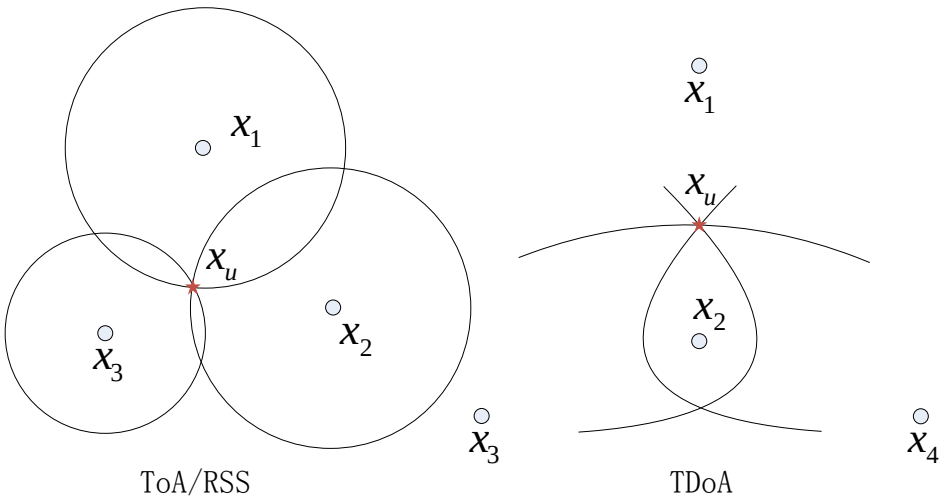

Figure 3.1: Geometric interpretation of a consistent positioning problem. Transmitters are at known positions x_1, x_2, x_3, and x_4, and the user position x_u is unknown.

- *ToA/RSS*: for each transmitter, the unknown user is located on the circle centered at the transmitter with radius equal to the corresponding range measurement. Positioning is done by determining the intersection of all the circles.

- *TDoA*: in this case, it is not possible to use a spherical interpretation since each pseudorange contains an additional unknown parameter b_c. As b_c is canceled in the pseudorange difference, the user is located on the hyperbola with two of the transmitters as the foci, and the difference of distances to the two foci is equal to the pseudorange difference between the same pair of transmitters. Hence, the user position can be found as the intersection of all the hyperbolas.

In an overdetermined system, redundant measurements are often taken to increase the position estimation accuracy. Due to the presence of $\underline{e}$, redundant measurements generally lead to an inconsistent system of equations. Least-squares, reviewed in Section 3.3, is one of the most widely applied theories to solve an inconsistent system of equations.

3.1.1.2 Measurement and Error Statistics

Recall from Chapter 2, the statistics of a single ToA, RSS, or RSS-based range measurement and the corresponding measurement error have been given as:

- Timing-based range: $\underline{y}_i \sim \mathcal{N}(d_i + b_c, \sigma_i^2)$, and $\underline{e}_i \sim \mathcal{N}(0, \sigma_i^2)$. Please note that $b_c = 0$ if it is a ToA-range.
- RSS: $\underline{y}_i = \underline{P}_i(d_i) \sim \mathcal{N}(P_0 + 10n_i \log_{10} \frac{d_i}{d_0}, \sigma_{s_i}^2)$, and $\underline{e}_i = \underline{s}_i \sim \mathcal{N}(0, \sigma_{s_i}^2)$.
- RSS-range: $\underline{y}_i \sim \text{Log}\mathcal{N}(d_i, \sigma_i^2)$, and $\underline{e}_i = \|x_u - x_i\|(e^{\frac{-\underline{s}_i \ln 10}{10n_i}} e^{-\frac{1}{2}(\frac{\sigma_{s_i} \ln 10}{10n_i})^2} - 1)$ is equal to a log-normal random variable plus a constant; its mean is 0 and its STD is σ_i.

Please note that the bandwidth and distance dependency effects do not change the distribution of the LoS ToA measurements, and therefore are not included here for simplicity. Moreover, the STD σ_i for Timing-based range and RSS-range measurements are not equal.

Assuming that the measurements collected from different base stations are independent, the statistics of the measurement vector $\underline{y}$ and the error term $\underline{e}$ in (3.2) can be determined as follows.

Timing-based Range and RSS The statistics for $\underline{y}$ and $\underline{e}$ given in this paragraph with general notations are valid for timing-based range and RSS measurements, since they both follow Gaussian distributions.

$$\begin{aligned} \underline{y} &\sim \mathcal{N}_m(y, Q_{yy}) \text{ with a diagonal } Q_{yy} \\ [Q_{yy}]_{i,i} &= \sigma_i^2, \ \ i = 1, ..., m, \end{aligned} \tag{3.7}$$

where y is the error free true value of $\underline{y}$. In timing-based cases, an element y_i is equal to the true distance d_i (plus the clock bias b_c in case of pseudoranges) and σ_i is the STD of the timing-based range measurement; in RSS cases, y_i is equal to the actual power received at true distance d_i, and $\sigma_i = \sigma_{s_i}$ is the STD of the corresponding shadowing effects, in dB. Moreover,

$$\begin{aligned} \underline{e} &\sim \mathcal{N}_m(0, Q_{ee}) \text{ with a diagonal } Q_{ee} \\ [Q_{ee}]_{i,i} &= \sigma_i^2, \ \ i = 1, ..., m. \end{aligned} \tag{3.8}$$

Please note that σ_i in the above equations is usually obtained empirically. To consider the bandwidth and distance dependencies in calculating the mean and variance for ToA-range measurements, one can use the expression given in (2.5)-(2.6).

RSS-range The distribution of m independent RSS-range measurements can be given as:

$$\begin{aligned} \underline{y} &\sim \text{Log}\mathcal{N}_m(y, Q_{yy}) \text{ with a diagonal } Q_{yy} \\ [Q_{yy}]_{i,i} &= \sigma_i^2 = d_i^2(e^{(\frac{\sigma_{s_i}\ln 10}{10n_i})^2} - 1), \;\; i = 1, ..., m, \end{aligned} \tag{3.9}$$

where $y_i = d_i$.

And the error term $\underline{e}_i = \|x_u - x_i\|(e^{\frac{-\underline{s}_i \ln 10}{10n_i}} e^{-\frac{1}{2}(\frac{\sigma_{s_i}\ln 10}{10n_i})^2} - 1)$ is a log-normal random variable plus a constant; its mean is 0 and its STD is equal to σ_i.

To this end, the original positioning model without differencing has been described. Based on (3.2)-(3.8), other models and the statistics of the corresponding error can be derived, which are introduced next.

3.1.2 Single Differencing

Recall from (3.3), b_c is a common term in all m measurements. Therefore, it can be canceled by subtracting one of the measurement (a reference) from all the other $m-1$ measurements, which result in $m-1$ range-difference equations:

$$\underline{y}_i - \underline{y}_r = \|x_u - x_i\| - \|x_u - x_r\| + \underline{e}_i - \underline{e}_r, \;\; i = 1 \ldots m-1, \tag{3.10}$$

where the subscript r stands for the reference, and in this case $r = m$.

Consuming one observation equation to get rid of the unknown b_c is the idea of the model introduced here. The matrix form can be easily generated by multiplying a transformation matrix T on both sides of (3.2):

$$T\underline{y} = TA(x) + T\underline{e}. \tag{3.11}$$

With the assumption $r = m$, the transformation matrix takes the form:

$$T = \begin{bmatrix} 1 & & 0 & -1 \\ & \ddots & & \vdots \\ 0 & & 1 & -1 \end{bmatrix}_{(m-1)\times m}. \tag{3.12}$$

Two issues about the single-differenced original model needs to be discussed:

1. One should only consider using the model when b_c is in presence, i.e. in a TDoA system. For a ToA/RSS system, there is no need to perform differencing operation between original measurements. Otherwise, one equation is consumed but the number of unknowns stays unchanged.

2. The choice for the transform matrix T is not limited to (3.12). The reference can be chosen as any one of, or as the mean of the m measurements. In the latter case, T takes the form:

$$T = \begin{bmatrix} 1-\frac{1}{m} & -\frac{1}{m} & \cdots & -\frac{1}{m} \\ -\frac{1}{m} & \ddots & & -\frac{1}{m} \\ \vdots & & \ddots & \vdots \\ -\frac{1}{m} & \cdots & \frac{1}{m} & 1-\frac{1}{m} \end{bmatrix}_{m\times m} . \tag{3.13}$$

Please note that T in (3.13) has only $m-1$ linearly independent rows, which means plugging (3.13) into (3.11) gives us only $m-1$ linearly independent equations.

3.1.2.1 Measurement and Error Statistics

The error statistics of the model can be easily derived from (3.7) and (3.8). Since the new measurement and error term are linearly related to $\underline{y}$ and $\underline{e}$, they are also Normally distributed, and can be characterized as

$$T\underline{y} \sim \mathcal{N}_m(Ty, TQ_{yy}T^T), \tag{3.14}$$
$$T\underline{e} \sim \mathcal{N}_m(0, TQ_{ee}T^T). \tag{3.15}$$

3.2 Models on Squared Observables

Recall from (3.3), the non-linear part takes the form $\sqrt{(.)^T(.)}$. The presence of the square-root makes the positioning procedure at a later stage an uneasy task and it disables the possibility of further transformations and simplifications. Therefore, models on squared observables, as the name suggests, apply a squaring operation to get rid of the square-root. These models share a common structure where, after squaring, the unknown vector is involved in both a linear and a non-linear term. The general form of these models can be written as:

$$\underline{y}_D = A_D x_u + B_D b_c + C_D \kappa + \underline{e}_D, \text{ with } f(x_u, b_c, \kappa) = 0, \tag{3.16}$$

where $\underline{y}_D$ is the new measurement vector obtained from $\underline{y}$, different for different direct methods, A_D, B_D, C_D are known matrices, κ is an unknown non-linear term involving x_u and b_c, with the relation indicated by $f(x_u, b_c, \kappa) = 0$, and $\underline{e}_D$ is the error term. Note that the subscript D stands for 'direct', since positioning models with this form are widely used by the direct methods.

In the rest of this Chapter, it is assumed that κ is treated as an extra unknown parameter in addition to and independent on x, i.e. the relation $f(x_u, b_c, \kappa) = 0$ is ignored. This assumption will be discussed later in the next chapter.

In case of TDoA systems, some of the known matrices may contain range measurements $\underline{y}$, and therefore are perturbed by measurement errors. How this fact effects corresponding positioning results will also be discussed in the next chapter.

Moreover, the models that involve squaring operations are only valid for timing-based range and RSS-range measurements. This is because RSS measurements involve $\|x_u - x_i\|$ in a non-linear logarithmic form, see (2.30), and a squaring operation on a RSS measurement does not give a linear equation of x_u. In the following content of this chapter and also in Chapter 4-5, the term ToA/RSS only refers to systems that use ToA/RSS range measurements when models with squaring operations are applied.

According to the number of differencing operations applied to (3.2), and therefore different choices of A_D, B_D, C_D in (3.16), the positioning models with squaring are further subcategorized into three groups, which are discussed as follows.

3.2.1 No Differencing

Squaring on both sides of (3.3) gives a new set of squared-range equations:

$$\underline{y}_i^2 = x_u^T x_u - b_c^2 - 2x_i^T x_u + 2\underline{y}_i b_c + x_i^T x_i + 2d_i\underline{e}_i + \underline{e}_i^2,\ i = 1 \dots m, \tag{3.17}$$

which clearly contains linear terms of x_u, b_c and a common non-linear term $\kappa = x_u^T x_u - b_c^2$.

In accordance with (3.16), the model on squared observables with no differencing is characterized by the following choices of $\underline{y}_{D,i} = \underline{y}_i^2 - x_i^T x_i$, $A_{D,i} = -2x_i^T$, $\underline{B}_{D,i} = 2\underline{y}_i$, $C_{D,i} = 1$, $\kappa = x_u^T x_u - b_c^2$ and $\underline{e}_{D,i} = 2d_i\underline{e}_i + \underline{e}_i^2$. Note that with ToA/RSS systems, one has $\underline{B}_D = 0$ and $\kappa = x_u^T x_u$.

3.2.1.1 Measurement and Error Statistics

Timing-based Range According to (3.7), $\underline{y}_i \sim \mathcal{N}(y_i, \sigma_i^2)$, and therefore the distribution of $\underline{y}_{D,i}$ can be derived as:

$$\underline{y}_{D,i} = \underline{y}_i^2 - x_i^T x_i = (\underline{e}_i + d_i)^2 - x_i^T x_i = \sigma_i^2 \underline{\chi}_{e_i}^2 - x_i^T x_i, \tag{3.18}$$

where $\underline{\chi}_{e_i}^2 \sim \chi^2(1, \frac{d_i^2}{\sigma_i^2})$ is a non-central chi-square distributed variable. Moreover, the mean and variance of the error term can be derived as:

$$\begin{aligned} \mathsf{E}\{\underline{y}_{D,i}\} &= \sigma_i^2 + y_i^2 - x_i^T x_i,\ i = 1, ..., m, \\ [Q_{y_D y_D}]_{ii} &= 4y_i^2\sigma_i^2 + 2\sigma_i^4, \\ [Q_{y_D y_D}]_{ij} &= 0,\ i, j = 1, ..., m, i \neq j. \end{aligned} \tag{3.19}$$

The distribution of $\underline{e}_{D,i}$ can be derived in a similar way:

$$\underline{e}_{D,i} = 2d_i\underline{e}_i + \underline{e}_i^2 = (\underline{e}_i + d_i)^2 - d_i^2 = \sigma_i^2\underline{\chi}_{e_i}^2 - d_i^2. \tag{3.20}$$

and, the mean and variance of the error term can be derived as:

$$\begin{aligned} \mathsf{E}\{\underline{e}_{D,i}\} &= \sigma_i^2,\ i = 1, ..., m, \\ [Q_{e_D e_D}]_{ii} &= 4d_i^2\sigma_i^2 + 2\sigma_i^4, \\ [Q_{e_D e_D}]_{ij} &= 0,\ i, j = 1, ..., m, i \neq j. \end{aligned} \tag{3.21}$$

Please note that in case of TDoA systems, $y_i = d_i + b_c$ is generally not equal to d_i, and therefore, $Q_{y_D y_D} \neq Q_{e_D e_D}$.

RSS-range According to (3.5) and (3.9),

$$\underline{y}_i = d_i e^{\frac{-\underline{s}_i \ln 10}{10 n_i}} e^{-\frac{1}{2}(\frac{\sigma_{s_i} \ln 10}{10 n_i})^2} \sim \text{Log}\mathcal{N}(d_i, \sigma_i^2), \tag{3.22}$$

which means $\underline{y}_i^2$ is also a log-normal random variable. Therefore,

$$\underline{y}_{D,i} = \underline{y}_i^2 - x_i^T x_i = d_i^2 e^{\frac{-2\underline{s}_i \ln 10}{10 n_i}} e^{-(\frac{\sigma_{s_i} \ln 10}{10 n_i})^2} - x_i^T x_i \tag{3.23}$$

is the sum of a log-normal distributed variable and a constant, with the following mean and vairance:

$$\begin{aligned} \text{E}\{\underline{y}_{D,i}\} &= d_i^2 e^{(\frac{\sigma_{s_i} \ln 10}{10 n_i})^2} - x_i^T x_i,\ i = 1, ..., m, \\ [Q_{y_D y_D}]_{ii} &= d_i^4 e^{2(\frac{\sigma_{s_i} \ln 10}{10 n_i})^2} (e^{4(\frac{\sigma_{s_i} \ln 10}{10 n_i})^2} - 1), \\ [Q_{y_D y_D}]_{ij} &= 0,\ i, j = 1, ..., m, i \neq j. \end{aligned} \tag{3.24}$$

The error term $\underline{e}_{D,i} = \underline{y}_{D,i} - y_{D,i} = \underline{y}_i^2 - d_i^2$ is also a log-normal variable plus a constant, and its mean and variance can be derived as:

$$\begin{aligned} \text{E}\{\underline{e}_{D,i}\} &= d_i^2 (e^{(\frac{\sigma_{s_i} \ln 10}{10 n_i})^2} - 1),\ i = 1, ..., m, \\ [Q_{e_D e_D}]_{ii} &= d_i^4 e^{2(\frac{\sigma_{s_i} \ln 10}{10 n_i})^2} (e^{4(\frac{\sigma_{s_i} \ln 10}{10 n_i})^2} - 1), \\ [Q_{e_D e_D}]_{ij} &= 0,\ i, j = 1, ..., m, i \neq j. \end{aligned} \tag{3.25}$$

Please note that the variance of $\underline{e}_{D,i}$ is equal to the variance of $\underline{y}_{D,i}$.

3.2.2 Single Differencing

Single differencing can be applied either to (3.17), ending up with a set of *squared-range differences*, or directly to the equations (3.3) then squaring, ending up with a set of *squared range-differences* (please note the different hyphenation).

3.2.2.1 Squared-Range Differences

The squared-range differences can be obtained from the squared-ranges (3.17) in a similar way as we obtain range-differences from (3.3).

Suppose $\underline{y}_r^2$ with $r = m$ is chosen as a reference, subtracting it from all other $m-1$ squared ranges gives:

$$\underline{y}_i^2 - \underline{y}_r^2 = -2(x_i - x_r)^T x_u + 2(\underline{y}_i - \underline{y}_r) b_c + x_i^T x_i - x_r^T x_r + 2 d_i \underline{e}_i - 2 d_r \underline{e}_r + \underline{e}_i^2 - \underline{e}_r^2, \tag{3.26}$$

with $i = 1 \ldots m-1$. Note that the common non-linear term $x_u^T x_u - b_c^2$ in (3.17) is canceled.

The model using squared-range differences then takes the form $\underline{y}_{D,i} = \underline{y}_i^2 - \underline{y}_r^2 - x_i^T x_i + x_r^T x_r$, $A_{D,i} = -2(x_i - x_r)^T$, $\underline{B}_{D,i} = 2(\underline{y}_i - \underline{y}_r)$, $C_{D,i} = 0$, $\underline{e}_{D,i} = 2 d_i \underline{e}_i + \underline{e}_i^2 - 2 d_r \underline{e}_r - \underline{e}_r^2$ and κ is canceled. In case of ToA/RSS systems, one can simply choose $\underline{B}_D = 0$. Compared to the case with no differencing, the number of equations is reduced by one, but the number of unknowns is also reduced by one, with κ canceled.

3.2.2.2 Measurement and Error Statistics

Timing-based Range The statistics of $\underline{y}_{D,i}$ in a squared-range difference equation can be derived as:

$$\underline{y}_{D,i} = \underline{y}_i^2 - \underline{y}_r^2 - x_i^T x_i + x_r^T x_r = \sigma_i^2 \underline{\chi}_{e_i}^2 - \sigma_r^2 \underline{\chi}_{e_r}^2 - x_i^T x_i + x_r^T x_r \tag{3.27}$$

with $i = 1, ..., m-1$. The mean and variance read:

$$\begin{aligned} \mathsf{E}\{\underline{y}_{D,i}\} &= \sigma_i^2 - \sigma_r^2 + y_i^2 - y_r^2 - x_i^T x_i + x_r^T x_r, \\ [Q_{y_D y_D}]_{ii} &= 4y_i^2\sigma_i^2 + 2\sigma_i^4 + 4y_r^2\sigma_r^2 + 2\sigma_r^4, \\ [Q_{y_D y_D}]_{ij} &= 4y_r^2\sigma_r^2 + 2\sigma_r^4,\ i,j = 1, ..., m-1, i \neq j. \end{aligned} \tag{3.28}$$

The statistics of the error term in a squared-range difference equation can be derived as:

$$\underline{e}_{D,i} = 2d_i\underline{e}_i - 2d_r\underline{e}_r + \underline{e}_i^2 - \underline{e}_r^2 = \sigma_i^2 \underline{\chi}_{e_i}^2 - d_i^2 - \sigma_r^2 \underline{\chi}_{e_r}^2 + d_r^2 \tag{3.29}$$

with $i = 1, ..., m-1$. The mean and variance read:

$$\begin{aligned} \mathsf{E}\{\underline{e}_{D,i}\} &= \sigma_i^2 - \sigma_r^2, \\ [Q_{e_D e_D}]_{ii} &= 4d_i^2\sigma_i^2 + 2\sigma_i^4 + 4d_r^2\sigma_r^2 + 2\sigma_r^4, \\ [Q_{e_D e_D}]_{ij} &= 4d_r^2\sigma_r^2 + 2\sigma_r^4,\ i,j = 1, ..., m-1, i \neq j. \end{aligned} \tag{3.30}$$

Please note the difference between $Q_{y_D y_D}$ and $Q_{e_D e_D}$.

RSS-range For RSS-range measurements, the a single squared-range difference is expressed as:

$$\underline{y}_{D,i} = \underline{y}_i^2 - \underline{y}_r^2 - x_i^T x_i + x_r^T x_r,\ i = 1, ..., m-1 \tag{3.31}$$

which is the difference between two independent log-normal variables plus a constant $-x_i^T x_i + x_r^T x_r$. The mean and variance of $\underline{y}_{D,i}$ are:

$$\begin{aligned} \mathsf{E}\{\underline{y}_{D,i}\} &= d_i^2 e^{(\frac{\sigma_{s_i}\ln 10}{10n_i})^2} - d_r^2 e^{(\frac{\sigma_{s_r}\ln 10}{10n_r})^2} - x_i^T x_i + x_r^T x_r, \\ [Q_{y_D y_D}]_{ii} &= d_i^4 e^{2(\frac{\sigma_{s_i}\ln 10}{10n_i})^2}(e^{4(\frac{\sigma_{s_i}\ln 10}{10n_i})^2} - 1) + d_r^4 e^{2(\frac{\sigma_{s_r}\ln 10}{10n_r})^2}(e^{4(\frac{\sigma_{s_r}\ln 10}{10n_r})^2} - 1), \\ [Q_{y_D y_D}]_{ij} &= d_r^4 e^{2(\frac{\sigma_{s_r}\ln 10}{10n_r})^2}(e^{4(\frac{\sigma_{s_r}\ln 10}{10n_r})^2} - 1),\ i,j = 1, ..., m-1, i \neq j. \end{aligned} \tag{3.32}$$

The error term $\underline{e}_{D,i} = \underline{y}_{D,i} - y_{D,i} = \underline{y}_i^2 - \underline{y}_r^2 - d_i^2 + d_r^2$, which is also the difference between two independent log-normal variables plus a constant $-d_i^2 + d_r^2$. Its mean and variance are given as:

$$\begin{aligned} \mathsf{E}\{\underline{e}_{D,i}\} &= d_i^2(e^{(\frac{\sigma_{s_i}\ln 10}{10n_i})^2} - 1) - d_r^2(e^{(\frac{\sigma_{s_r}\ln 10}{10n_r})^2} - 1), \\ [Q_{e_D e_D}]_{ii} &= d_i^4 e^{2(\frac{\sigma_{s_i}\ln 10}{10n_i})^2}(e^{4(\frac{\sigma_{s_i}\ln 10}{10n_i})^2} - 1) + d_r^4 e^{2(\frac{\sigma_{s_r}\ln 10}{10n_r})^2}(e^{4(\frac{\sigma_{s_r}\ln 10}{10n_r})^2} - 1), \\ [Q_{e_D e_D}]_{ij} &= d_r^4 e^{2(\frac{\sigma_{s_r}\ln 10}{10n_r})^2}(e^{4(\frac{\sigma_{s_r}\ln 10}{10n_r})^2} - 1),\ i,j = 1, ..., m-1, i \neq j. \end{aligned} \tag{3.33}$$

3.2.2.3 Squared Range-Differences

Other than dealing with squared-range differences, it is also possible to build a model using squared range-differences. Squaring on both sides of the range-difference equation (3.10) gives:

$$(\underline{y}_i-\underline{y}_r)^2 = -2(x_i-x_r)^T x_u - 2(\underline{y}_i-\underline{y}_r)d_r + x_i^T x_i - x_r^T x_r + 2d_i(\underline{e}_i-\underline{e}_r) + (\underline{e}_i-\underline{e}_r)^2, \quad (3.34)$$

for $i = 1 \ldots m-1$. In this case, the squaring operation is applied after the single differencing is performed. One measurement is consumed to get rid of the unknown clock offset b_c, instead of κ in the case of squared-range differences. Hence, this type of model should not be applied to ToA/RSS systems where no clock offset exists. The model using squared range-differences is then characterized by the following choices: $\underline{y}_{D,i} = (\underline{y}_i - \underline{y}_r)^2 - x_i^T x_i + x_r^T x_r$, $A_{D,i} = -2(x_i - x_r)^T$, $\underline{B}_{D,i} = 0$, $C_{D,i} = -2(\underline{y}_i - \underline{y}_r)$, $\kappa = d_r = \sqrt{(x_u - x_r)^T(x_u - x_r)}$ and $\underline{e}_{D,i} = 2d_i(\underline{e}_i - \underline{e}_r) + (\underline{e}_i - \underline{e}_r)^2$.

3.2.2.4 Measurement and Error Statistics

The statistics of $\underline{y}_{D,i}$ in a squared range-difference can be derived based on the fact that $\underline{y}_i - \underline{y}_r \sim \mathcal{N}(y_i - y_r, \sigma_i^2 + \sigma_r^2)$:

$$\underline{y}_{D,i} = (\underline{y}_i - \underline{y}_r)^2 - x_i^T x_i + x_r^T x_r = (\sigma_i^2 + \sigma_r^2)\underline{\chi}^2_{y_i - y_r} - x_i^T x_i + x_r^T x_r \quad (3.35)$$

where $\underline{\chi}^2_{y_i-y_r} \sim \chi^2(1, \frac{(y_i-y_r)^2}{\sigma_i^2+\sigma_r^2})$. The mean and variance of $\underline{y}_D$ then follows:

$$\begin{aligned} \mathsf{E}\{\underline{y}_{D,i}\} &= \sigma_i^2 + \sigma_r^2 + (y_i - y_r)^2 - x_i^T x_i + x_r^T x_r, \\ [Q_{y_D y_D}]_{ii} &= 4(y_i - y_r)^2(\sigma_i^2 + \sigma_r^2) + 2(\sigma_i^2 + \sigma_r^2)^2, \\ [Q_{y_D y_D}]_{ij} &= 4(y_i - y_r)(y_j - y_r)\sigma_r^2 + 2\sigma_r^4, \ i, j = 1, ..., m-1, i \neq j. \end{aligned} \quad (3.36)$$

The error term $\underline{e}_{D,i}$ has the same structure of the one in a squared range, which can be obtained by simply replacing $\underline{e}_i$ in(3.20) by $\underline{e}_i - \underline{e}_r$. Given the fact that $\underline{e}_i - \underline{e}_r \sim \mathcal{N}(0, \sigma_i^2 + \sigma_r^2)$, it holds:

$$\underline{e}_{D,i} = 2d_i(\underline{e}_i - \underline{e}_r) + (\underline{e}_i - \underline{e}_r)^2 = (\sigma_i^2 + \sigma_r^2)\underline{\chi}^2_{e_i - e_r} - d_i^2. \quad (3.37)$$

where $\underline{\chi}^2_{e_i-e_r} \sim \chi^2(1, \frac{d_i^2}{\sigma_i^2+\sigma_r^2})$. The mean and variance of $\underline{e}_D$ then follows:

$$\begin{aligned} \mathsf{E}\{\underline{e}_{D,i}\} &= \sigma_i^2 + \sigma_r^2, \\ [Q_{e_D e_D}]_{ii} &= 4d_i^2(\sigma_i^2 + \sigma_r^2) + 2(\sigma_i^2 + \sigma_r^2)^2, \\ [Q_{e_D e_D}]_{ij} &= 4d_i d_j \sigma_r^2 + 2\sigma_r^4, \ i, j = 1, ..., m-1, i \neq j. \end{aligned} \quad (3.38)$$

3.2.3 Double Differencing

Similar as the case of single differencing, double differencing can also be applied to two different types of equations, i.e. squared-range differences in (3.26) and squared range-differences in (3.34). However, the equations obtained by double differencing these two types of equations are identical.

Dividing both sides of (3.26) by $\underline{y}_i - \underline{y}_r$ gives:

$$\underline{y}_i + \underline{y}_r = \frac{-2(x_i - x_r)^T x_u + x_i^T x_i - x_r^T x_r + 2d_i\underline{e}_i - 2d_r\underline{e}_r + \underline{e}_i^2 - \underline{e}_r^2}{\underline{y}_i - \underline{y}_r} + 2b_c \quad (3.39)$$

with $i = 1 \dots m-1$.

The subscript rr will be used to denote references used in the second round of differencing. Suppose $rr = m-1$, subtracting $\underline{y}_{rr} + \underline{y}_r$ from the first $m-2$ equations in (3.39) gives:

$$\begin{aligned}\underline{y}_i - \underline{y}_{rr} = -2\left[\frac{(x_i - x_r)^T}{\underline{y}_i - \underline{y}_r} - \frac{(x_{rr} - x_r)^T}{\underline{y}_{rr} - \underline{y}_r}\right] x_u + \left[\frac{x_i^T x_i - x_r^T x_r}{\underline{y}_i - \underline{y}_r} - \frac{x_{rr}^T x_{rr} - x_r^T x_r}{\underline{y}_{rr} - \underline{y}_r}\right] \\ + \left[\frac{2d_i\underline{e}_i - 2d_r\underline{e}_r + \underline{e}_i^2 - \underline{e}_r^2}{\underline{y}_i - \underline{y}_r} - \frac{2d_{rr}\underline{e}_{rr} - 2d_r\underline{e}_r + \underline{e}_{rr}^2 - \underline{e}_r^2}{\underline{y}_{rr} - \underline{y}_r}\right]\end{aligned} \quad (3.40)$$

Although it seems that only one differencing operation $\underline{y}_i - \underline{y}_{rr}$ exists, the above equation is obtained from (3.26) where another differencing has already been applied to obtain squared-range differences and there are only $m-2$ equations now.

Equation (3.40) can also be obtained by first dividing both sides of (3.34) with $\underline{y}_i - \underline{y}_r$:

$$\underline{y}_i - \underline{y}_r = \frac{-2(x_i - x_r)^T x_u + x_i^T x_i - x_r^T x_r + 2d_i(\underline{e}_i - \underline{e}_r) + (\underline{e}_i - \underline{e}_r)^2}{\underline{y}_i - \underline{y}_r} - 2d_r \quad (3.41)$$

with $i = 1 \dots m-1$, and then choosing $\underline{y}_{rr} - \underline{y}_r$ as the reference and subtracting it from the first $m-2$ equations in (3.41), which gives exactly (3.40).

Clearly, in both cases of double differencing, one more equation is consumed to get rid of the clock offset b_c and another one for the non-linear term κ. Therefore, double differencing should not be applied to a ToA/RSS system which does not have a clock offset, i.e. single differencing is already enough to obtain a set of equations that contains only x_u as the unknown, see (3.26).

In contrast to (3.16), the model on double differenced squared observables is characterized with the following choices:

$$\underline{y}_{D,i} = \underline{y}_i - \underline{y}_{rr} - \left[\frac{x_i^T x_i - x_r^T x_r}{\underline{y}_i - \underline{y}_r} - \frac{x_{rr}^T x_{rr} - x_r^T x_r}{\underline{y}_{rr} - \underline{y}_r}\right] \quad (3.42)$$

$$\underline{A}_{D,i} = -2\left[\frac{(x_i - x_r)^T}{\underline{y}_i - \underline{y}_r} - \frac{(x_{rr} - x_r)^T}{\underline{y}_{rr} - \underline{y}_r}\right] \quad (3.43)$$

$$\underline{e}_{D,i} = \left[\frac{2d_i\underline{e}_i - 2d_r\underline{e}_r + \underline{e}_i^2 - \underline{e}_r^2}{\underline{y}_i - \underline{y}_r} - \frac{2d_{rr}\underline{e}_{rr} - 2d_r\underline{e}_r + \underline{e}_{rr}^2 - \underline{e}_r^2}{\underline{y}_{rr} - \underline{y}_r}\right] \quad (3.44)$$

and $B_D = C_D = 0$.

3.2.3.1 Measurement and Error Statistics

The statistics of in a double differenced squared range is quite simple:

The expression of $\underline{y}_{D,i}$ and $\underline{e}_{D,i}$ in the model on double differenced squared observables are quite complicated and we need to rely on approximations to evaluate the error.

Here we first address the mean and variance of $\underline{y}_D$. Apparently, $\underline{y}_D$ contains functions of $\underline{e}$, and expanding it at $\mathsf{E}\{\underline{e}\} = 0$ gives:

$$\underline{y}_D(\underline{e}) = y_D(0) + \partial_\alpha y_D(0)\underline{e}^\alpha + \frac{1}{2}\underline{e}^\alpha \partial^2_{\alpha\beta} e_D(0)\underline{e}^\beta + \cdots, \;\; \alpha, \beta = 1, ..., m \tag{3.45}$$

where the Einstein's summation convention is used, with $\partial_\alpha(.)$ denoting the partial derivative of $(.)$ w.r.t. the α-th element of $\underline{e}$, denoted as $\underline{e}^\alpha$.

Ignoring the terms with order higher than 2, the mean and variance of $\underline{y}_D$ can be approximated as:

$$\begin{aligned}
\mathsf{E}\{\underline{y}_D(\underline{e})\} &\approx y_D(0) + \frac{1}{2}\mathsf{E}\left\{\underline{e}^\alpha \partial^2_{\alpha\beta} y_D(0)\underline{e}^\beta\right\} \\
&= y_D(0) + \frac{1}{2}\partial^2_{\alpha^2} y_D(0)\mathsf{E}\{(\underline{e}^\alpha)^2\} = y_D(0) + \frac{1}{2}\partial^2_{\alpha^2} y_D(0)\sigma^2_\alpha \qquad (3.46) \\
Q_{y_D y_D} &\approx \mathsf{E}\left\{\partial_\alpha y_D(0)\underline{e}^\alpha \underline{e}^\beta \partial_\beta y_D^T(0)\right\} = \partial_\alpha y_D(0)\sigma^2_\alpha \partial_\alpha y_D^T(0), \qquad (3.47)
\end{aligned}$$

where the fact that $\mathsf{E}\{\underline{e}^\alpha \underline{e}^\beta\} = 0, \text{ for } \alpha \neq \beta$, has been used. The expressions of $\partial_\alpha y_{D,i}(0)$ and $\partial^2_{\alpha^2} y_{D,i}(0)$ are given in Appendix A.1.

The expressions in (3.46) and (3.47) can be directly used for the mean and variance of $\underline{e}_D$, and one only needs to change $\underline{y}_D$ to $\underline{e}_D$, which gives:

$$\begin{aligned}
\mathsf{E}\{\underline{e}_D(\underline{e})\} &\approx e_D(0) + \frac{1}{2}\mathsf{E}\left\{\underline{e}^\alpha \partial^2_{\alpha\beta} e_D(0)\underline{e}^\beta\right\} \\
&= e_D(0) + \frac{1}{2}\partial^2_{\alpha^2} e_D(0)\mathsf{E}\{(\underline{e}^\alpha)^2\} = e_D(0) + \frac{1}{2}\partial^2_{\alpha^2} e_D(0)\sigma^2_\alpha \qquad (3.48) \\
Q_{e_D e_D} &\approx \mathsf{E}\left\{\partial_\alpha e_D(0)\underline{e}^\alpha \underline{e}^\beta \partial_\beta e_D^T(0)\right\} = \partial_\alpha e_D(0)\sigma^2_\alpha \partial_\alpha e_D^T(0), \qquad (3.49)
\end{aligned}$$

and the details expressions of $\partial_\alpha e_{D,i}(0)$ and $\partial^2_{\alpha^2} e_{D,i}(0)$ are given in Appendix A.1.

To this end, the expressions of the first two moments of $\underline{y}_D$ and $\underline{e}_D$, in case of a model with double differenced squared observables, have been derived. The full distributions, however, require more elaboration. In fact, the knowledge of the first two moments is already sufficient for least-squares estimation, which is discussed in the sequel.

3.3 Non-linear Least-squares: A Brief Review

3.3.1 Principle

It seems reasonable to choose the estimate for x such that $A(\hat{x})$ is, in some sense, as close as possible to a measurement vector y. If the (weighted) sum of squares of the residual vector $y - A(x)$ is used to 'measure' the size of discrepancy, the estimate $\hat{x}$ will be the one minimizing

$$F(x) = [y - A(x)]^T W [y - A(x)] = \|y - A(x)\|^2_W, \tag{3.50}$$

where W is a weight matrix. This is the weighted least-squares principle. The sufficient and necessary conditions for a point $\hat{x}$ to be a minimizer of $F(x)$ in (3.50) are shown in [87]:

$$\begin{array}{ll} a) & \partial_x F(\hat{x}) = -\partial_x A^T(\hat{x}) W e(\hat{x}) = 0, \\ b) & \partial_{xx^T} F(\hat{x}) = \partial_x A^T(\hat{x}) W \partial_{x^T} A(\hat{x}) - e(\hat{x})^T W \partial^2_{xx^T} A(\hat{x}) > 0. \end{array} \tag{3.51}$$

where $e(\hat{x}) = y - A(\hat{x})$.

The geometric interpretation for the first condition can be described as follows. For all values of $x \in \mathbb{R}^n$, $A(x)$ describes an n-dimensional manifold, embedded in $\mathbb{R}^m$. If the metric of $\mathbb{R}^m$ is described by the positive definite matrix W, the scalar $\|y - A(x)\|_W$ equals the distance from y to the point $A(x)$ on the manifold. Hence, the problem of minimizing $F(x)$ corresponds to the problem of finding the point on the manifold $A(\hat{x})$, which has the least distance to y [41]. Hence, the first condition is equivalent to have $e(\hat{x})$ perpendicular to the tangent space of the manifold at the point $A(\hat{x})$, i.e. the first condition can be written as:

$$e(\hat{x}) \perp R(\partial_{x^T} A(\hat{x})), \text{ (W-orthogonal)}. \tag{3.52}$$

The geometry of the non-linear least-squares problem [41] is shown in Figure 3.2.

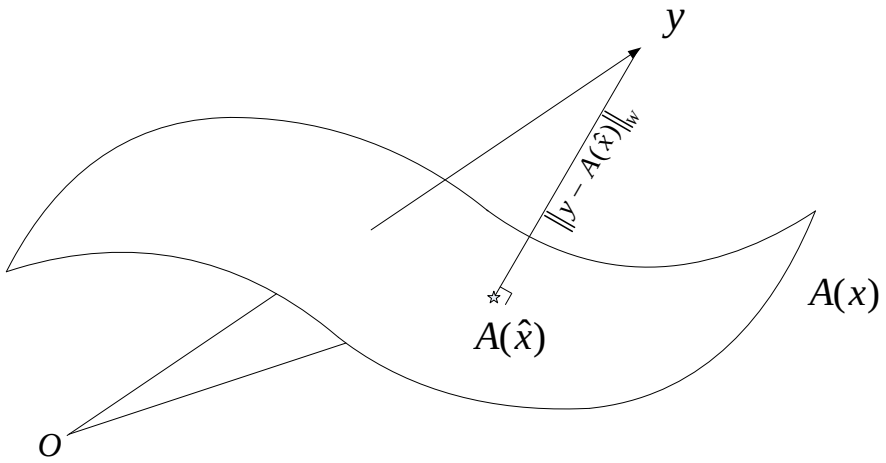

Figure 3.2: Geometry of non-linear least-squares.

The interpretation of the second condition is more complicated, which is given in [41] based on the concept of normal curvature. In the rest of Section 3.3.1, we assume that the weight matrix W is chosen as the inverse of the variance matrix of $\underline{y}$, i.e. $W = Q_{yy}^{-1}$.

To give the details about the normal curvature, it is convenient to first introduce a function:

$$d(x) = n^T W (A(x) - A(\hat{x})), \tag{3.53}$$

where n is a unit vector normal to the tangent space of the manifold at the point $A(\hat{x})$:

$$n^T W \partial_x A(\hat{x}) = 0, \text{and } n^T W n = 1. \tag{3.54}$$

For a point x close to $\hat{x}$, a Taylor expansion of (3.53) at $\hat{x}$ gives a good approximation of (3.53) by ignoring the terms of $(x - \hat{x})$ with order higher than two:

$$d(x) \approx \frac{1}{2}(x - \hat{x})^T (n^T W \partial^2_{xx^T} A(\hat{x}))(x - \hat{x}). \tag{3.55}$$

On the other hand, the expression of the normal curvature is defined in [88] as:

$$k_n(v) = \frac{v^T(n^T W \partial^2_{xx^T} A(\hat{x}))v}{v^T \partial_x A^T(\hat{x}) W \partial_{x^T} A(\hat{x}) v}. \tag{3.56}$$

Based on the resemblance of (3.55) and the numerator in (3.56), the latter is a good approximation of $2d(x)$ by choosing a small value of $v = (x - \hat{x})$. Moreover, since (3.53) is the projection of the distance from a point on the manifold $A(x)$ to the point of tangency $A(\hat{x})$ onto n, a study of its approximation gives information about the non-linearity of the manifold near the point of tangency. Therefore, the normal curvature is related to the non-linearity of the manifold.

Now the rest of the task is to relate the normal curvature the second condition, which can be further written as:

$$\begin{aligned} & \partial_x A^T(\hat{x}) W \partial_{x^T} A(\hat{x}) - e(\hat{x})^T W \partial^2_{xx^T} A(\hat{x}) > 0 \\ \Rightarrow \quad & v^T \partial_x A^T(\hat{x}) W \partial_{x^T} A(\hat{x}) v - v^T e(\hat{x})^T W \partial^2_{xx^T} A(\hat{x}) v > 0 \\ \Rightarrow \quad & \frac{v^T (e(\hat{x})^T W \partial^2_{xx^T} A(\hat{x}))v}{v^T \partial_x A^T(\hat{x}) W \partial_{x^T} A(\hat{x}) v} < 1, \end{aligned} \tag{3.57}$$

for any non-zero vector v.

With the resemblance of (3.56) and (3.57), the concept of normal curvature can be used to interpret the second condition by choosing as $n = \frac{e(\hat{x})}{\|e(\hat{x})\|_W}$, and the second condition (3.57) can be written as the product of the normal curvature and the length of the residual vector $e(\hat{x})$:

$$k_n(v)\|e(\hat{x})\|_W < 1, \forall v \in \mathbb{R}^n. \tag{3.58}$$

As mentioned before, the normal curvature is related to the nonlinearity of the manifold near $A(\hat{x})$, (3.58) then reveals an important fact that the second condition is affected by both the non-linearity of the system and the residual $\|e(\hat{x})\|_W$.

3.3.2 Local/Global Minima

In a non-linear system, $F(x)$ could have multiple minima. This can be seen based on the first condition (3.51a). With $A(x)$ being non-linear, $\partial_x F(x)$ could contain non-linear terms of x, and $\partial_x F(x) = 0$ has multiple solutions. These solutions are usually categorized as local minimizers and a global minimizer. A point $\hat{x}$ is said to be a local minimizer if $F(\hat{x}) < F(x)$ for all x near $\hat{x}$, e.g. for all x within a sphere (or spheroid for $\dim(x) > 3$) centered at $\hat{x}$ with a certain radius ϵ. A point is said to be a global minimizer if $F(\hat{x}) < F(x)$ for all $x \in R^n$.

Multiple minima can appear in both consistent and overdetermined systems. In consistent 3-D ToA/RSS systems with $m = n = 3$, the transmitters are always placed coplanar, and the two minimizers of $F(x)$ are of equal quality, with both satisfying $\partial_x F(\hat{x}) = 0$. Without additional information, it is not possible to pick the correct solution.

In redundant cases, local minima occur when the transmitters are not well separated along at least one dimension. For example, in a 3-D synchronized system, local minima appear when the transmitters are placed (close to) coplanar. The difference between minima depends not on only the system geometry, e.g. how close to coplanar the transmitters are placed, but also on measurement errors in e. This point is illustrated with the following example.

In Figure 3.3, the x-z portion of the least squares objective function of 3 fully synchronized overdetermined system with 4 transmitters is shown, where the range measurement errors are assumed to be independent and identically Gaussian distributed with zero mean. In Figure 3.3(a), the transmitters are placed almost in a plane and there is a clear sharp global peak pointing at the true position and a local peak corresponding to a incorrect solution. The global minimum in this case is much smaller than the local one, since the transmitters are still separated and the range measurements are of very good quality, with STD 0.02 m. In the same geometry, as the error STD gets larger, the local minimum in Figure 3.3(a) can become smaller than the global one with certain probability. This effect of measurement error can be seen in Figure 3.3(b) with STD 0.2 m, where the global peak is pointing at the incorrect solution. In case of 3.3(a) it is still possible to choose the correct solution by picking the one corresponding to the global minimum, while in Figure 3.3(b) it is not possible without additional information. In short, the multiple minima problem is mainly related to the system geometry, and coplanar type of geometries should be avoided. The local minimum is eliminated with a good geometry in Figure 3.3(c), where the transmitters are placed more separately, and the objective function has only one clear peak, pointing at the correct solution.

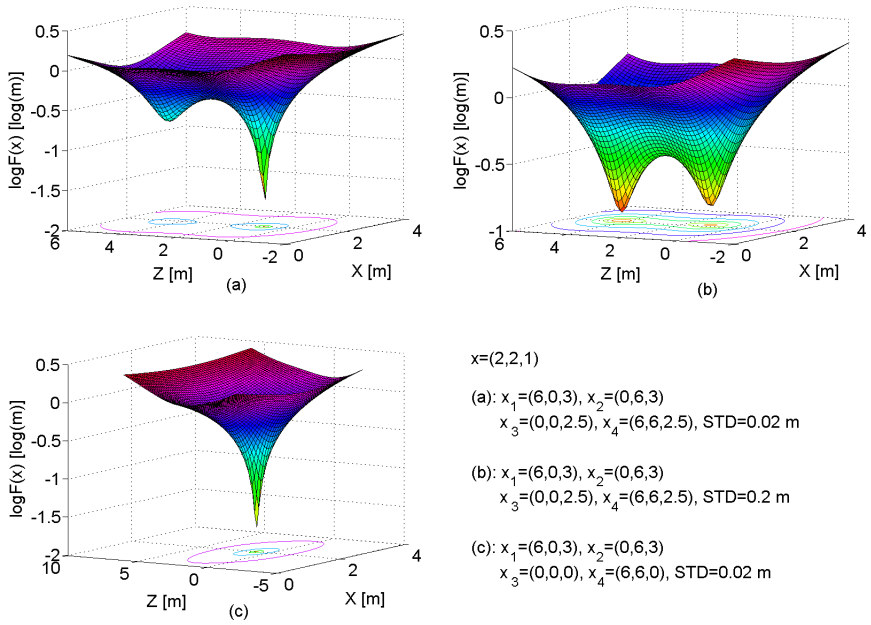

Figure 3.3: x-z portion of the least squares objective function in logarithm of three 3-D (x-y-z) positioning systems. The range measurement errors are assumed to be Gaussian independent and identically-distributed with zero mean.

3.3.3 Least-Squares, Maximum Likelihood and Cramer-Rao Lower Bound

ML estimation is only applicable when the PDF of $\underline{y} \in R^m$ is available, i.e. $f_{\underline{y}}(y|x)$ is known apart from unknown x. The ML estimate is the value which maximizes $f_{\underline{y}}(\underline{y}|x)$ as function of x, the so-called likelihood function [39], i.e.

$$\hat{x} = \arg\max_x f_{\underline{y}}(y|x). \tag{3.59}$$

In the mean time, based on the PDF of $\underline{y}$, the CRLB on the estimator of x in the model (3.2) can be calculated using the following equation [39, 89]:

$$\text{CRLB} = \left\{ \text{E}\left\{ [\partial_x \ln f_{\underline{y}}(\underline{y}|x)]^T [\partial_x \ln f_{\underline{y}}(\underline{y}|x)] \right\} \right\}^{-1}, \tag{3.60}$$

which indicates the lowest achievable variance of the unbiased estimator $\underline{\hat{x}}$, independent on the estimation method.

When $f_{\underline{y}}(y|x)$ is elliptically contoured, the properly weighted ($W = Q_{yy}^{-1}$) least-squares estimate and the ML estimate are identical and the variance of the estimator achieves the CRLB (approximately). For example, in case of normally distributed measurements, $f_{\underline{y}}(y|x) = (\det(2\pi Q_{yy}^{-1}))^{-1/2} \exp\{-\frac{1}{2}\|y - A(x)\|^2_{Q_{yy}^{-1}}\}$. The ML estimate can be expressed as $\arg\max_x f_{\underline{y}}(y|x) = \arg\max_x \|y - A(x)\|^2_{Q_{yy}^{-1}}$, which is the same as the weighted least-squares estimate with $W = Q_{yy}^{-1}$, and the CRLB in this case is calculated as $\left\{ \text{D}\left[\partial_x \|y - A(x)\|^2_{Q_{yy}^{-1}} \right] \right\}^{-1} = \left[\left(\partial_x A^T(x)\right)^T Q_{yy}^{-1} \partial_{x^T} A(x) \right]^{-1}$, which is identical to the expression of the least-squares estimator variance derived in [41], when $W = Q_{yy}^{-1}$.

3.3.4 Error Analysis and Bias due to Non-linearity

An important measure of the quality of a position estimator is the RMSE, which statistically reflects the mean distance from the estimator to the true position. The first and second moments of the estimator need to be evaluated for this purpose.

Here we review the general way [90], to derive the first and second moments of a (non-linear) least-squares estimator. Suppose the true position x can be obtained from $y = A(x)$ via a mapping $M(.)$, i.e. $x = M(y)$, and its estimator $\underline{\hat{x}}$ can be obtained as $M(\underline{y})$, where we do not specify function $M(y)$ explicitly. With $\underline{y}$ containing error $\underline{e}$, i.e. $\underline{y} = y + \underline{e}$, Taylorizing $\underline{\hat{x}}$ at y gives an expansion in $\underline{e}$:

$$\underline{\hat{x}} = M(y) + \partial_{y^T} M(y)\underline{e} + \frac{1}{2}\underline{e}^T \partial^2_{yy^T} M(y)\underline{e} + \cdots \tag{3.61}$$

Define the bias in $\underline{\hat{x}}$ as $\underline{\mu}_{\hat{x}} = \underline{\hat{x}} - x$, which is given, together with the variance of $\underline{\hat{x}}$, as

$$\mu_{\hat{x}} \approx \text{E}\{\underline{\mu}_{\hat{x}}\} = \frac{1}{2}\text{trace}\{\partial^2_{yy^T} M(y) Q_{ee}\}, \tag{3.62}$$

$$Q_{\hat{x}\hat{x}} \approx \partial_{y^T} M(y) Q_{ee} \partial_y M^T(y), \tag{3.63}$$

$$\text{RMSE} = \sqrt{\text{trace}\{Q_{\hat{x}\hat{x}}\} + \mu_{\hat{x}}^T \mu_{\hat{x}}}, \tag{3.64}$$

where the terms with order higher than 2 are ignored.

The rest of the task is to derive the expression of $M(.)$. This is in general non-trivial in non-linear least-squares problems. However, in (3.62) and (3.63), only the expressions of the $n \times m$ Jacobian $\partial_{y^T} M(y)$ and the $n \times m \times m$ Hessian $\partial^2_{yy^T} M(y)$ are required to evaluate the first and second moments of $\underline{\hat{x}}$, which can be obtained as follows. Taylorizing the left hand side of the first of the necessary and sufficient conditions (3.51a), $\partial_x A^T(\underline{\hat{x}}) W[\underline{y} - A(\underline{\hat{x}})] = 0$, at x gives an expansion in $\underline{\hat{x}} - x$. Using (3.61), an expansion in $\underline{e}$ can be obtained with unknown coefficients $\partial_y M(y)$ and $\partial^2_{yy^T} M(y)$. If $\underline{\hat{x}}$ is a least-squares solution, the condition (3.51a) holds for all $\underline{e}$ and one can collect the terms of the same order and set them to zero. In this way, the expressions of $\partial_{y^T} M(y)$ and $\partial^2_{yy^T} M(y)$ can be determined.

The full expressions of (3.62) and (3.63) are derived in [90] as:

$$\mu_{\hat{x}} = -Q(x)\partial_x A(x)^T Q_{yy}^{-1} \frac{1}{2} \text{trace}[\partial^2_{xx^T} A(x) Q(x)], \tag{3.65}$$

$$Q_{\hat{x}\hat{x}} = Q(x) = (\partial_x A(x)^T Q_{yy}^{-1} \partial_{x^T} A(x))^{-1}. \tag{3.66}$$

Equation (3.65) shows that non-linear least-squares estimators are inherently biased, even when the measurements are unbiased with $\text{E}(\underline{e}) = 0$. A rough check in [84] shows that the bias due to non-linearity can be a potential degrader for accurate indoor positioning. Ideally, this problem can be tackled by carefully arranging the positions of the base stations. In a clock offset free system, when the base stations are pair-wisely symmetric w.r.t. to the user position x, it can be proved that the bias due to non-linearity is zero. Without loss of generality, we place the unknown user x at the origin in a positioning system, surrounded by m base stations symmetrically. In such a case, we have

$$[Q(x)\partial_x A(x)^T Q_{yy}^{-1}]_{i,j} = -[Q(x)\partial_x A(x)^T Q_{yy}^{-1}]_{i,k}, \tag{3.67}$$

$$\text{trace}[\partial^2_{xx^T} A_j(x) Q(x)] = \text{trace}[\partial^2_{xx^T} A_k(x) Q(x)] \tag{3.68}$$

where the j-th and k-th base stations are symmetric w.r.t. x, and we assume a diagonal Q_{yy} with $[Q_{yy}]_{jj} = [Q_{yy}]_{kk}$.

Plugging the results into (3.65), the biases caused by the non-linearity in the symmetric measurement pairs are canceled.

As an example, we check the bias in two simple 2-D positioning setups as depicted in Figure 3.4. For simplicity, we assume that all the distances are equal to d, and $Q_{yy} = \sigma^2 I$. In case a) the bias is equal to $[-\frac{\sigma^2}{2d},\ \frac{\sigma^2}{2d}]$, and in b) the bias is zero.

The zero-bias case with symmetric topology of base stations gives insight for designing the geometry of a system where the movement of a target is limited in a space of relatively small scale compared to the whole system. In practice, however, pair-wise symmetry in full 3-D space of base stations around the user may still difficult to achieve, but it should be feasible to have pair-wise symmetry in the horizontal plane, and therefore, achieve zero-bias horizontally. For example, the base stations can be installed symmetrically in a big room to precisely monitor the small area body movements such as arms, legs, etc., of a dancer, being trained on a stage at the room center.

For more general situations where pair-wise symmetry is not possible, we propose to check the bias in the estimator by comparing it to the positioning accuracy in terms of variance

[84], more precisely by looking at the term $\|\mu_{\hat{x}}\|_{Q^{-1}_{\hat{x}\hat{x}}}$ as a measure of significance, where $\|\,.\,\|_{Q^{-1}_{\hat{x}\hat{x}}} = \sqrt{(.)^T Q^{-1}_{\hat{x}\hat{x}}(.)}$. The bias is considered large, if it overwhelms the variance of the estimator, i.e.

$$\|\mu_{\hat{x}}\|_{Q^{-1}_{\hat{x}\hat{x}}} > 1. \tag{3.69}$$

This idea is validated in Section 4.4.2, based on both simulations and actual audio UWB timing-based range measurements. Results show that the bias due to non-linearity is not considered significant for all tested geometries when the system is clock offset free; it can be a potential problem for systems with a clock offset when the measurement STD is large and the system geometry is not 'healthy', i.e. all the transmitters are on one side of the receiver. Design suggestions regarding system geometry have also been given based on the results of tests. The reason we do not describe the validation work here is because estimation of unknown position needs knowledge of positioning algorithms, which are discussed in the next chapter.

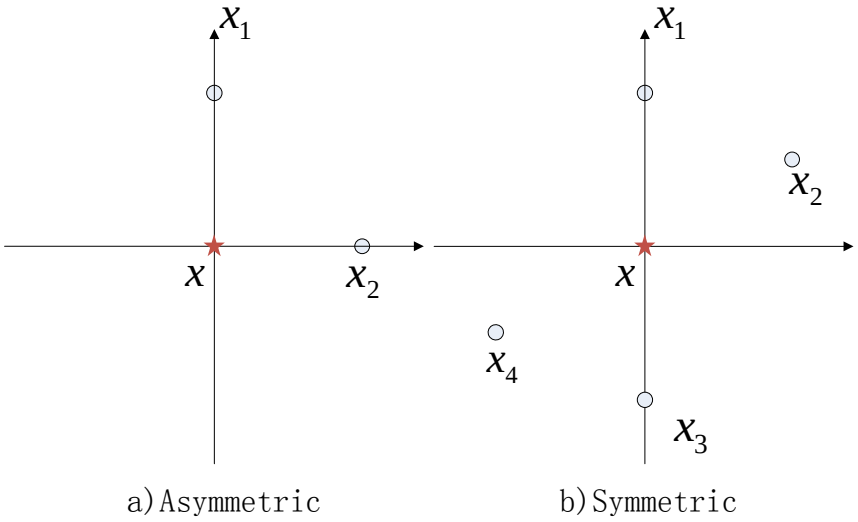

Figure 3.4: a) an asymmetric 2-D positioning system with 2 base stations, b) a symmetric system with 4 base stations, with two pairs (x_1, x_3) and (x_2, x_4) symmetric w.r.t. x, respectively.

3.4 Summary

In this chapter, the models for original and squared observables have been introduced. These models will be used by the positioning algorithms introduced in the next chapter. The corresponding properties and applicabilities are summarized in Table 3.2.

A brief review of the non-linear least-squares concept has been provided. This criteria is widely used because, when the ranging errors are distributed with an elliptically contoured PDF, the weighted ($W = Q^{-1}_{yy}$) least-squares estimator with and the ML estimator are equivalent, and its variance achieves the CRLB. Furthermore, compared to ML estimation, un-weighted least-squares estimation does not require any knowledge of the measurement PDF, and only first and second order moments of the measurement errors are required in the weighted case with $W = Q^{-1}_{yy}$.

Table 3.2: Positioning Models.

Model	Differencing	Equations	Unknowns	Applicability
Equation (3.2)	0	m	x_u, b_c	TDoA and ToA/RSS
Equation (3.10)	1	$m-1$	x_u	only TDoA
Equation (3.17)	0	m	x_u, b_c, κ	TDoA and ToA/RSS
Equation (3.26)	1	$m-1$	x_u,b_c	TDoA and ToA/RSS
Equation (3.34)	1	$m-1$	x_u,κ	only TDoA
Equation (3.40)	2	$m-2$	x_u	only TDoA

The sufficient and necessary conditions for a point $\hat{x}$ to be a least-squares solution to (3.2) can be found in (3.51a) and (3.51b). The geometric interpretation indicates that a) the least-squares residual vector $e(\hat{x})$ should be orthogonal to the tangent space of the manifold at the point $A(\hat{x})$, and b) whether a stationary $\hat{x}$ is a minimizer depends on the non-linearity of the manifold and the uncertainty in the measurement, i.e. $\hat{x}$ is a minimizer if the product of the normal curvature and the norm of the residual vector is less than 1. These two conditions only guarantee $\hat{x}$ to be a local minimizer. In non-linear least-squares problems, e.g. positioning, multiple local minimizers may exist and system designers should avoid a co-planar type of geometry to make the global minimizer easier to locate.

The general expressions for the mean and variance of a non-linear least-squares estimator have been given by (3.65) and (3.66). The estimator is inherently biased, even when the measurements are unbiased. A criteria to decide whether the bias is significant or not is given in (3.69), and the corresponding validation work will follow in Section 4.4.2.

Range-based Least-Squares Positioning Algorithms

4

As mentioned in Chapter 1, the range-based least-squares positioning algorithms can be grouped into Iterative Descent (ID) algorithms, direct algorithms and constrained iterative algorithms.

The purpose of this Chapter is to give a structured and detailed review of the three groups of range-based least-squares positioning methods, and give special attention to the direct methods. The basic principles, assumptions and approximations of these methods are discussed, their relations are pointed out, and the properties of different types of methods are summarized. Positioning results obtained with actual UWB range measurements are used to illustrate and support the review. The content is applicable to various applications and systems, no matter the ranging signal is optical, radio or acoustic, and it aims to address the following issues:

- The iterative methods all require an initial guess to start, which is much more difficult to choose in indoor applications than in traditional long range maritime navigation and on-earth satellite navigation, due to the small scale of indoor environments. The low-complexity nature of the direct methods makes them good candidates for this purpose. A study and assessment of different direct methods is therefore necessary.

- There is a large number of direct methods documented in literature, scattered across the fields of radar, aerospace engineering, oceanic engineering, (acoustic) signal processing and wireless communications. Some of the documented methods are, surprisingly, identical, though the derivations are often greatly different. With a proper classification of the direct methods, a better understanding can be achieved and overlapping research can be avoided.

- With the underlying principles of the direct methods revealed, and thereby their inherent strong and weak points, the content of this chapter can be used to assist researchers and developers in the field to find the right choice for their applications. The level of technical detail of this review should allow readers to straightforwardly implement the method of their choice into an algorithm.

4.1 Iterative Descent Methods

In this section, we will review the ID methods based on [41], including the Steepest Descent method, the Newton method, the Trust Region method, the Gauss-Newton method, and

the Levenberg Marquardt method. These methods are widely used to solve non-linear least-squares problems. The mechanics of the ID methods can be described as:

1. Set $k = 0$, and choose an initial guess x_0,
2. update the estimate x_{k+1} from x_k based on the predefined rule,
3. if a termination criteria is met, choose x_{k+1} as the final solution, otherwise increase k by 1 and return to step 2.

According to [41], the estimates of the ID methods in each iteration are updated with a general form:

$$x_{k+1} = x_k - t_k Q(x_k) \partial_x F(x_k), \tag{4.1}$$

where t_k is a positive scalar and $Q(x_k)$ is an arbitrary but positive definite matrix.

An important performance measure of an iterative method is its rate of convergence, which can be derived by Taylorizing (4.1) at the final estimate $\hat{x}$:

$$\begin{aligned} x_{k+1} - \hat{x} &= [I - t_k Q(\hat{x}) \partial^2_{xx^T} F(\hat{x})](x_k - \hat{x}) \\ &- \frac{1}{2} t_k (x_k - \hat{x}) \left[2\partial_x Q(\hat{x}) \partial^2_{xx^T} F(\hat{x}) + Q(\hat{x}) \partial^3_{xx^T x} F(\hat{x}) \right] (x_k - \hat{x})^T + \cdots \end{aligned} \tag{4.2}$$

where we use the fact that for a least-squares estimate $\hat{x}$, $\partial_x F(\hat{x}) = 0$. For points sufficiently close to the solution, the rate of convergence is dominated by the term with lowest order of $x_k - \hat{x}$.

The choices of t_k and $Q(x_k)$ vary with different ID methods, which are briefly reviewed as follows.

4.1.1 Steepest Descent Method

The choice of $Q(x_k)$ of the steepest descent method is

$$Q(x_k) = I, \tag{4.3}$$

and (4.1) can be further written as

$$x_{k+1} = x_k - t_k \partial_x F(x_k). \tag{4.4}$$

The motivation of this choice is that the opposite direction of the gradient vector, $-\partial_x F(x_k)$, represents a direction of maximum rate of decrease for the function $F(x)$ at x_k. Plugging (4.3) into (4.2) gives the rate of convergence for the steepest descent method,

$$x_{k+1} - \hat{x} = [I - t_k \partial^2_{xx^T} F(\hat{x})](x_k - \hat{x}) + o(x_k - \hat{x}). \tag{4.5}$$

which is *linear* for points close to the solution $\hat{x}$.

The scalar t_k is generally chosen such that $F(x_{k+1}) = F(x_k - t_k \partial_x F(x_k))$ is minimized, which is called the *exact line search strategy*. In this case,

$$\partial_{t_k} F(x_{k+1}) = -\partial_x F(x_{k+1})^T \partial_x F(x_k) = 0. \tag{4.6}$$

With the simple choice of $Q(x_k) = I$, the advantage of the steepest descent method is clearly its low-complexity (in terms of computation load). The involved highest order partial derivative of $F(x)$ is only one, and no matrix needs to be inversed. However, the steepest descent method has the following two drawbacks [41]: 1) With the exact line search strategy, $\partial_x F(x_{k+1})$ is orthogonal to $\partial_x F(x_k)$, which means the steepest descent method has the tendency to zigzag when the contours of $F(x)$ are very elongated. 2) the performance of the steepest descent method is *variant* under reparametrization.

4.1.2 Newton Method

The Newton method is characterized by the choice of

$$t_k Q(x_k) = [\partial^2_{xx^T} F(x_k)]^{-1}, \tag{4.7}$$

and the update equation then becomes

$$x_{k+1} = x_k - [\partial^2_{xx^T} F(x_k)]^{-1} \partial_x F(x_k) \tag{4.8}$$

The motivation of the Newton method is that the root searching of the non-linear function $\partial_x F(x) = 0$ can be approximated by iteratively searching the root of its tangent, evaluated at the intermediate points x_k. With this choice, the rate of convergence for the Newton method can be evaluated based on (4.2)

$$x_{k+1} - \hat{x} = -\frac{1}{2}(x_k - \hat{x}) \left[\partial_x [\partial^2_{xx^T} F(\hat{x})]^{-1} \partial^2_{xx^T} F(\hat{x})\right] (x_k - \hat{x})^T + o[(x_k - \hat{x})^2]. \tag{4.9}$$

The above equation shows that the Newton method has a *quadratic* rate of convergence for points close to the solution $\hat{x}$. This makes it very effective in dealing with general non-linear minimization problems. However, one drawback of the Newton method is the heavy computational load required for the evaluation and inversion of the Hessian matrix $\partial^2_{xx^T} F(x)$. The related calculations of $\partial^2_{xx^T} A(x)$, a $m \times n \times n$ tensor, are usually quite involved. Another drawback of the Newton method is that it fails to work when $\partial^2_{xx^T} F(x)$ is close to singular or fails to be positive definite. The Trust Region method is then developed to cope with these situations, and it is reviewed in the next subsection.

4.1.3 Trust Region Method

In case of a poor conditioned $\partial^2_{xx^T} F(x_k)$, the approximation used by the Newton method:

$$f(x) = F(x_k) + \partial_x F(x_k)(x - x_k) + \frac{1}{2}(x - x_k)^T \partial^2_{xx^T} F(x_k)(x - x_k) \tag{4.10}$$

of $F(x)$ may become inadequate, with $o[(x - x_k)^2]$ too large to ignore.

The approximation can be improved by restricting the region for x, which is the idea of the trust region method, and it works as follows. Once the estimate x_{k+1} is updated, $F(x_{k+1}) - F(x_k)$ is compared to its approximation $f(x_{k+1}) - f(x_k)$. If the approximation is good, x_{k+1} is accepted as the new iteration point. Otherwise, a badly conditioned Hessian matrix is possibly indicated, and the search region should be contracted to obtain another x_{k+1}. This is realized by choosing $t_k Q(x_k)$ as:

$$t_k Q(x_k) = [\partial^2_{xx^T} F(x_k) + \alpha_k R]^{-1}, \tag{4.11}$$

where α_k is a non-negative scalar and R is a positive definite matrix. The trust region method then takes the form

$$x_{k+1} = x_k - [\partial^2_{xx^T} F(x_k) + \alpha_k R]^{-1} \partial_x F(x_k). \tag{4.12}$$

In this way, increasing the value of α_k results in a smaller step size $\|x_{k+1} - x_k\|$, and the trust region is contracted. Actually, the trust region method is capable of alternating between the Newton method and the steepest descent method. That is, by adjusting the scalar α_k, the trust region method switches between the quadratic approximation of $F(x)$ used by the Newton method, and the linear approximation used by the steepest descent method. This point can be well explained as we assume that R is chosen as the identity matrix. In case of $\alpha_k = 0$, (4.12) becomes

$$x_{k+1} = x_k - [\partial^2_{xx^T} F(x_k)]^{-1} \partial_x F(x_k), \tag{4.13}$$

which is identical to the Newton method. And for a large α_k, (4.12) can be approximated as

$$x_{k+1} = x_k - \alpha_k^{-1} \partial_x F(x_k), \tag{4.14}$$

which is identical to the steepest descent method with α_k^{-1} playing the role of t_k.

4.1.4 Gauss-Newton Method

Unlike the steepest descent method and the Newton method that use direct approximations of $F(x)$, the Gauss-Newton method completes the linearization *within* the norm of $F(x)$, which is done by replacing $A(x)$ with its linearized version [41]:

$$A(x) \approx A(x_k) + \partial_{x^T} A(x_k)(x - x_k). \tag{4.15}$$

The objective function is then approximated as $\|y - A(x_k) - \partial_{x^T} A(x_k)(x - x_k)\|^2_W$. This motivation gives the choice of $Q(x_k)$ as:

$$t_k Q(x_k) = \left(\partial_x A^T(x_k) W \partial_{x^T} A(x_k)\right)^{-1}, \tag{4.16}$$

and the update equation then becomes

$$x_{k+1} = x_k - \left(\partial_x A^T(x_k) W \partial_{x^T} A(x_k)\right)^{-1} \partial_x F(x_k) \tag{4.17}$$

Table 4.1: Summary of the reviewed ID algorithms.

Methods	Complexity	Convergence Rate	Comment
Steepest Descent	Simple with $Q(x_k) = I$	Linear	Tendency to zigzag when elliptically contoured $F(x)$
Newton	Heavy computations to evaluate $\partial^2_{xx^T} F(x_k)$	Quadratic	Convergence problem when $\partial^2_{xx^T} F(x_k)$ close to singular
Trust Region	Similar as Newton	Linear with big α_k Quadratic with small α_k	Several α_k tested per iteration with $\partial^2_{xx^T} F(x_k)$ close to singular
Gauss-Newton	Moderate computations to evaluate $\partial_x A^T(x_k) W \partial_{x^T} A(x_k)$	Linear	Convergence problem when $\partial_x A^T(x_k) W \partial_{x^T} A(x_k)$ close to singular
Levenberg-Marquardt	Similar as Gauss-Newton	Linear	Several α_k tested per iteration when $\partial_x A^T(x_k) W \partial_{x^T} A(x_k)$ close to singular

The Gauss-Newton method does not have a zigzag problem of the steepest descent method, and only the first order derivative needs to be calculated. But the convergence rate is *linear*, which can be seen by plugging (4.16) into (4.2):

$$x_{k+1} - \hat{x} = \left[I - \left(\partial_x A^T(x_k) W \partial_{x^T} A(x_k) \right)^{-1} \partial^2_{xx^T} F(\hat{x}) \right] (x_k - \hat{x}) + o(x_k - \hat{x}). \quad (4.18)$$

In the mean time, the Gauss-Newton method faces difficulties with a non-invertible $\partial_x A(x_k)^T W \partial_x A(x_k)$. These difficulties can be overcome by an adaptive technique called the Levenberg-Marquardt method, which is reviewed next.

4.1.5 Levenberg-Marquardt Method

The Levenberg-Marquardt method [91] can be looked on as a trust region modification of the Gauss-Newton method, since it takes the form

$$t_k Q(x_k) = \left[\partial_x A^T(x_k) W \partial_{x^T} A(x_k) + \alpha_k R \right]^{-1}, \quad (4.19)$$

where α_k is a non-negative scalar and R is a positive definite matrix. The motivation and the mechanics of the Levenberg-Marquardt method can be explained in a similar way as the trust region method, the difference is that the Levenberg-Marquardt alternates between the Gauss-Newton method and the steepest descent method.

The properties of the ID methods are summarized in Table 4.1.

4.1.6 Error Analysis

In the error analysis of a specific ID method, one might be inclined to use the update equation (4.1) in the form

$$\underline{\hat{x}} = x - tQ(x)\partial_x F(x), \quad (4.20)$$

to derive the moments of the estimator obtained with a specific ID method. The result is obviously variant under a change of choice for $tQ(x)$, thus variant with different methods, even when the objective function to minimize is unique. However, the ID methods are nothing else but techniques to minimize the objective function, which uniquely determines the least-squares estimator. A least-squares solution should always fulfill the first necessary and sufficient condition $\partial_x F(\hat{x}) = 0$, i.e. the final (local) least-squares solutions obtained with different ID methods are identical, provided that the corresponding methods converge successfully. Hence, the correct way to derive the first, second moments and the RMSE of a least-squares estimator is to rely on the condition $\partial_x F(\hat{x}) = 0$, as reviewed previously in Section 3.3.4.

4.1.7 Initial Guess

All iterative methods require an initial guess x_0 as the start point to gradually converge to the final solution. Intuitively, the closer the initial guess is to the final solution, the smaller is the required number of iterations. Moreover, in case the objective function has multiple minima, the method will converge to the peak of the cone (see Figure 3.3) where the $F(x_0)$ is located. Hence, obtaining a good initial guess is one of the crucial problems that should be addressed when applying iterative techniques.

In long range maritime navigation and on-earth satellite navigation, a good initial guess is easy to obtain, e.g. even the center of the earth works. It gets tricky however for indoor applications, where the scale of the system is much smaller than a satellite system and the prior information on whereabouts of the target is more limited. In the mean time, indoor positioning systems should be built with preferably (very) low-cost and low-power equipments. One promising way to obtain a good initial guess is to use so called direct methods, which complete position estimation in one go. This idea will be verified with the simulations and actual audio UWB ToA-range measurements in Section 4.4.1 and these non-iterative methods are discussed in the next section.

4.2 Direct Methods

There is a large number of direct methods documented in literature, scattered across the fields of radar, aerospace engineering, oceanic engineering, (acoustic) signal processing and wireless communications. In this section, these direct methods are classified based on the least-squares objective function to minimize, which is the soul of the least-squares estimation and determines the accuracy (in terms of RMSE) of the corresponding estimator.

Recall from (3.16), the models with squared observables used by the direct methods share the general form:

$$\underline{y}_D = A_D x_u + B_D b_c + C_D \kappa + \underline{e}_D, \text{ with } f(x_u, b_c, \kappa) = 0. \tag{4.21}$$

During least-squares estimation, the constraint $f(x_u, b_c, \kappa) = 0$ should be taken into consideration since it provides information in addition to the model $\underline{y}_D = A_D x_u + B_D b_c +$

$C_D\kappa + \underline{e}_D$. One of the possible ways is to use a Lagrange Multiplier λ to account for $f(x_u, b_c, \kappa) = 0$. The rigorous least-squares solution of (4.21) then can be found by:

$$\begin{aligned} [\hat{\underline{x}}_u^T,\ \hat{\underline{b}}_c,\ \hat{\underline{\chi}}]^T &= \arg\min_{x_u, b_c, \kappa} \|\underline{y}_D - A_D x_u - B_D b_c - C_D \kappa\|^2_{W_D} \\ \text{subject to:} &\quad f(x_u, b_c, \kappa) = 0, \end{aligned} \tag{4.22}$$

which can be obtained by searching the stationary point of the Lagrange function:

$$\|\underline{y}_D - A_D x_u - B_D b_c - C_D \kappa\|^2_{W_D} + \lambda f(x_u, b_c, \kappa). \tag{4.23}$$

which satisfies both (3.51a) and (3.51b), where $F(x)$ is now the Lagrange function and $x = [x_u^T, b_c, \kappa, \lambda]^T$. In the rest of this thesis, it is assumed by default that (3.51b) is satisfied when referring to a stationary point of a Lagrange function.

Here the weight matrix W_D is often chosen as the inverse of the variance matrix of $\underline{y}_D$, i.e. $W_D = Q^{-1}_{\underline{y}_D \underline{y}_D}$. Please note that here we refer (4.22) as a *rigorous*, not strict, least-squares solution, since the objective function to minimize is different from the one in (3.50) for the original positioning problem.

In general, the above rigorous least-squares solution (4.22) can only be approximated *iteratively*, due to the non-linear constraint $f(x_u, b_c, \kappa) = 0$. Simplifications are therefore required, in order to obtain a solution directly in one go. In literature, two major types of simplifications exist:

1. Single Objective Function. With the first type of simplification, the constraint is completely ignored. In this case, (4.21) becomes just a linear model

   $$\underline{y}_D = A_D x_u + B_D b_c + C_D \kappa + \underline{e}_D, \tag{4.24}$$

   and the corresponding least-squares solution reads:

   $$[\hat{\underline{x}}_u^T,\ \hat{\underline{b}}_c,\ \hat{\underline{\kappa}}]^T = \arg\min_{x_u, b_c, \kappa} \|\underline{y}_D - A_D x_u - B_D b_c - C_D \kappa\|^2_{W_D} \tag{4.25}$$

   Denote $F_D = [A_D,\ B_D,\ C_D]$ and $x_D = [x_u^T,\ b_c,\ \kappa]^T$, the final expression of the estimator reads:

   $$\hat{\underline{x}}_D = (F_D^T W_D F_D)^{-1} F_D^T W_D \underline{y}_D. \tag{4.26}$$

   The direct methods that use this type of simplification will be referred to as the Single Objective Function (SOF) methods, since only one objective function is involved for estimation.

   The solution (4.26) however is not valid if $\dim(\underline{y}_D) < \dim(x_u) + 1$ for ToA/RSS and $\dim(\underline{y}_D) < \dim(x_u) + 2$ for TDoA. Compared to the original problem (3.3), one extra measurement is required to solve the problem with one additional unknown κ. Hence, some of the SOF methods do not work with $m = n$.

2. Multiple Objective Function. Instead of completely discarding the information $\kappa = f(x_u, b_c)$, the second type of simplification partially exploits the constraint. The final solution can be seen as a combination of solutions obtained with multiple objective

functions, and the expression is different for different methods. In contrast to the SOF methods, the direct methods applying the second type of simplifications will be referred to as the Multiple Objective Function (MOF) methods. Moreover, the MOF methods are capable of solving a system of (3.2) with $m = n$.

In the following sections, the SOF and MOF methods are reviewed in detail, and the results show that although the derivations of the documented direct methods can be greatly different, the SOF methods are, surprisingly, identical, in the sense that they can be seen as a unique problem solved by just applying different weighting schemes.

4.2.1 Single Objective Function Methods

The first SOF method was introduced in [42] to solve range difference equations of TDoA systems. Later on, other SOF methods are found in [46, 45, 47, 51, 52] for TDoA systems, and [55, 56, 57, 58, 59] for ToA/RSS systems. The SOF methods are further sub-categorized into three groups, according to the applied positioning models.

4.2.1.1 No Differencing

The first group of SOF methods use the squared model with no differencing (3.17), and are characterized by the following choices of $\underline{y}_{D,i} = \underline{y}_i^2 - x_i^T x_i$, $A_{D,i} = -2x_i^T$, $\underline{B}_{D,i} = 2\underline{y}_i$, $C_{D,i} = 1$, $\kappa = x_u^T x_u - b_c^2$ and $\underline{e}_{D,i} = 2d_i\underline{e}_i + \underline{e}_i^2$. Note that with ToA/RSS systems, one has $\underline{B}_D = 0$ and $\kappa = x_u^T x_u$. The method introduced in [45, 46], falls in this category. Although the authors introduced the method with two steps, the expression of the estimator can be given in one single equation, as shown with (4.26). For this type of SOF methods, the system is determined when $\dim(y_D) = m = \dim(x_u) + \dim b_c + \dim(\kappa)$.

4.2.1.2 Single Differencing

The second group of SOF methods use the squared models with single differencing. The used equations can be either squared-range differences (3.26) or squared range-differences (3.34) (please note the different hyphenation).

In the first case, the non-linear term κ is cancelled, and the corresponding SOF methods choose $\underline{y}_{D,i} = \underline{y}_i^2 - \underline{y}_r^2 - x_i^T x_i + x_r^T x_r$, $A_{D,i} = -2(x_i - x_r)^T$, $\underline{B}_{D,i} = 2(\underline{y}_i - \underline{y}_r)$, $C_{D,i} = 0$. In case of ToA/RSS systems, $\underline{B}_D = 0$.

The single differencing SOF methods using squared-range differences include the ones in [55, 56, 57, 58, 59]. The methods in [55, 56] follow exactly the formulation of (3.26). Slight differences exist in [57], where, instead of $\underline{y}_r : \ r = m$, the reference is chosen as the average of all m available squared ranges: $\underline{y}_r = \frac{1}{m}\sum_{i=1}^{m} \underline{y}_i^2$; in [58], the differencing operation is repeated for m times for $r = 1, ..., m$; and in [59], the authors suggest to pick the shortest estimated range as the reference. However, the number of linearly independent squared-range difference equations in [55, 56, 57, 58, 59] is all $m - 1$. As shown later in

Section 4.2.1.4, these methods are equivalent. Moreover, for this type of SOF methods, the system is determined when $\dim(y_D) = m - 1 = \dim(x_u) + \dim b_c$.

In the second case with the squared range-differences, the canceled term is b_c, and the corresponding SOF methods are then characterized by the following choices: $\underline{y}_{D,i} = (\underline{y}_i - \underline{y}_r)^2 - x_i^T x_i + x_r^T x_r$, $A_{D,i} = -2(x_i - x_r)^T$, $\underline{B}_{D,i} = 0$, $C_{D,i} = -2(\underline{y}_i - \underline{y}_r)$ and $\kappa = d_r = \sqrt{(x_u - x_r)^T(x_u - x_r)}$. The SOF method that uses (3.34) is the one introduced in [51, 52]. For this type of SOF methods, the system is determined when $\dim(y_D) = m - 1 = \dim(x_u) + \dim(\kappa)$.

4.2.1.3 Double Differencing

No matter the second round of differencing is applied to (3.26) or (3.34), the squared model with double differencing has a unique expression as shown in (3.40).

In contrast to (4.24), the SOF methods with double differencing are characterized with the following choices:

$$\underline{y}_{D,i} = \underline{y}_i - \underline{y}_{rr} - \left[\frac{x_i^T x_i - x_r^T x_r}{\underline{y}_i - \underline{y}_r} - \frac{x_{rr}^T x_{rr} - x_r^T x_r}{\underline{y}_{rr} - \underline{y}_r} \right] \tag{4.27}$$

$$\underline{A}_{D,i} = -2 \left[\frac{(x_i - x_r)^T}{\underline{y}_i - \underline{y}_r} - \frac{(x_{rr} - x_r)^T}{\underline{y}_{rr} - \underline{y}_r} \right] \tag{4.28}$$

and $B_D = C_D = 0$. The methods that belong to this subcategory include the ones in [42, 47]. The system in this case is determined when $\dim(y_D) = m - 2 = \dim(x_u)$. Actually, in [42] the author first introduced the method for a set of equations with $\dim(\underline{y}) = \dim(x_u) + \dim(b_c)$ and then dealt with the case with $\dim(y_D) = \dim(y) - 2 = \dim(x_u)$. Note that only in the second case, the method belongs to SOF methods, while in the first case, it belongs to MOF methods, which will be covered later in Section 4.2.2.

4.2.1.4 Relating SOF Methods

If two linear sets of equations can be linearly related with a transformation matrix, the corresponding solutions can be seen as obtained using different measurement weighting schemes that reflect the propagation of transformation relation, and therefore are equivalent.

Proof: Constructing another set of equations, in contrast to (4.24):

$$\underline{y}'_D = A'_D x_u + B'_D b_c + C'_D \kappa + \underline{e}'_D, \tag{4.29}$$

and $\underline{y}'_D$ can be related to $\underline{y}_D$ in (4.24) with an arbitrary transformation matrix T: $\underline{y}'_D = T\underline{y}_D$. Hence, it holds that $(.)'_D = T(.)_D$ for all coefficients on the right hand sides of (4.24) and (4.29). Using the notation $F'_D = [A'_D,\ B'_D,\ C'_D]$ and $x_D = [x_u^T,\ b_c,\ \kappa]^T$, the least-squares estimator of (4.29) can be written as:

$$\hat{\underline{x}}'_D = \arg\min_{x_D} \|\underline{y}'_D - F'_D x_D\|^2_{W'_D}. \tag{4.30}$$

With a specific choice of

$$W'_D = (TW_D^{-1}T^T)^{-1}, \tag{4.31}$$

Eq. (4.30) yields:

$$\begin{aligned} \underline{\hat{x}}'_D &= (F_D'^T W'_D F'_D)^{-1} F_D'^T W'_D \underline{y}'_D & (4.32)\\ &= (F_D^T W_D F_D)^{-1} F_D^T W_D \underline{y}_D = \underline{\hat{x}}_D, & (4.33) \end{aligned}$$

where T has to be invertible, and $(TW_D^{-1}T^T)^{-1} = T^{-T}W_D T^{-1}$. Hence, two sets of equations, that can be linearly related by any invertible transformation matrix, provide *equivalent* solutions, with the specific choice of (4.31).

Our case is more complicated, since T is not invertible. For example, in (3.26), the numbers of equations and unknowns are both reduced by one, as compared to (3.17). The matrix, transforming (3.17) to (3.26), then takes the form:

$$T = \begin{bmatrix} 1 & & 0 & -1 \\ & \ddots & & \vdots \\ 0 & & 1 & -1 \end{bmatrix}_{(m-1)\times m}, \tag{4.34}$$

which is non-invertible.

However, it can still be proved that when W'_D, the weight matrix for (3.26), is related to W_D, the weight matrix for (3.17), as $W'_D = (TW_D^{-1}T^T)^{-1}$, these two sets of equations provide identical least-squares solutions.

Proof: Construct an auxiliary set of equations:

$$\begin{aligned} \underline{y}_A &= F_A x_D + \underline{e}_A, & (4.35)\\ \underline{\hat{x}}_{D_A} &= (F_A^T W_A F_A)^{-1} F_A^T W_A \underline{y}_A & (4.36) \end{aligned}$$

which is obtained from (4.24) via $(.)_A = T_A(.)_D$, with the following transformation matrix:

$$T_A = \begin{bmatrix} 1 & & 0 & -1 \\ & \ddots & & \vdots \\ 0 & & 1 & -1 \\ 0 & \cdots & 0 & 1 \end{bmatrix}_{m\times m} = \begin{bmatrix} T \\ 0\cdots 0 \;\; 1 \end{bmatrix}. \tag{4.37}$$

Thus, we have:

$$y_A = \begin{bmatrix} \underline{y}'_D \\ \underline{y}_{D,m} \end{bmatrix}, \;\; F_A = \begin{bmatrix} F'_D & 0 \\ A_{D,m} & 1 \end{bmatrix}, \;\; x_A = \begin{bmatrix} x'_D \\ \kappa \end{bmatrix} \tag{4.38}$$

with $A_{D,m}$ the m-th row of A_D and $\underline{y}_{D,m}$ the m-th element of $\underline{y}_D$.

Suppose that the weight matrix in (4.36) is chosen as $W_A = (T_A W_D^{-1} T_A^T)^{-1}$, it holds $\underline{\hat{x}}_{D_A} = \underline{\hat{x}}_D$, since T_A in (4.37) is invertible.

Furthermore, with $W'_D = (TW_D^{-1}T^T)^{-1}$, we have

$$\begin{aligned} W_A &= \begin{bmatrix} W_D'^{-1} & c \\ c^T & d \end{bmatrix}^{-1} \\ &= \begin{bmatrix} W'_D + W'_D c(d - c^T W'_D c)^{-1} c^T W'_D & \underbrace{-W'_D c(d - c^T W'_D c)^{-1}}_{\alpha} \\ -(d - c^T W'_D c)^{-1} c^T W'_D & \underbrace{(d - c^T W'_D c)^{-1}}_{\beta} \end{bmatrix} \\ &= \begin{bmatrix} W'_D + \alpha\beta^{-1}\alpha^T & \alpha \\ \alpha^T & \beta \end{bmatrix} \end{aligned} \tag{4.39}$$

where $c_{1\times(m-1)}$ and $d_{1\times 1}$ are some constants.

Combining (4.38) and (4.39) gives:

$$(F_A^T W_A F_A)^{-1} = \left(\begin{bmatrix} F_D'^T & A_{D,m}^T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} W'_D + \alpha\beta^{-1}\alpha^T & \alpha \\ \alpha^T & \beta \end{bmatrix} \begin{bmatrix} F'_D & 0 \\ A_{D,m} & 1 \end{bmatrix} \right)^{-1} \tag{4.40}$$

$$= \begin{bmatrix} (F_D'^T W'_D F'_D)^{-1} & -(F_D'^T W'_D F'_D)^{-1}(F_D'^T \alpha\beta^{-1} + A_{D,m}^T) \\ -(A_{D,m} + \beta^{-1}\alpha^T F'_D)(F_D'^T W'_D F'_D)^{-1} & \beta + (A_{D,m} + \beta^{-1}\alpha^T F'_D)(F_D'^T W'_D F'_D)^{-1}(F_D'^T \alpha\beta^{-1} + A_{D,m}^T) \end{bmatrix}$$

$$F_A^T W_A \underline{y}_A = \begin{bmatrix} F_D'^T W'_D \underline{y}'_D + F_D'^T \alpha\beta^{-1}\alpha^T \underline{y}'_D + F_D'^T \alpha \underline{y}_{D,m} + A_{D,m}^T \alpha^T \underline{y}_D + A_{D,m}^T \beta \underline{y}_{D,m} \\ \alpha^T \underline{y}'_D + \beta \underline{y}_{D,m} \end{bmatrix} \tag{4.41}$$

$$\Rightarrow \hat{\underline{x}}_{D_A} = (F_A^T W_A F_A)^{-1} F_A^T W_A \underline{y}_A$$

$$= \begin{bmatrix} (F_D'^T W'_D F'_D)^{-1} F_D'^T W'_D \underline{y}'_D \\ -(A_{D,m} + \beta^{-1}\alpha^T F'_D)(F_D'^T W'_D F'_D)^{-1} F_D'^T W'_D \underline{y}'_D + \beta(\alpha^T \underline{y}'_D + \beta \underline{y}_{D,m}) \end{bmatrix} \tag{4.42}$$

The first $m-1$ row of (4.42) is exactly the expression of $\hat{\underline{x}}'_D$. Note that (4.39) and (4.40) is calculated based on the following two key equations [92]:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix} \tag{4.43}$$

$$= \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{bmatrix} \tag{4.44}$$

Other linear sets of equations obtained from (3.17) using a transformation matrix different from (4.34) can be addressed in a similar way.

To this end, it is clear that all the mentioned SOF methods are linearly related, and provide identical solutions with special choices of weight matrices, see (4.31). In the literature, SOF methods are presented as different methods in different academic areas, but in general, these methods can be seen as realizations of a unique method with just different weighting schemes.

4.2.2 Multiple Objective Function Methods

The MOF methods employ more than one different objective functions, and the corresponding estimator is in general not a minimizer for any of the employed objective functions, but a 'combination' of them.

Recall from (4.25) and (4.22), the simplified and rigorous least-squares solutions to the squared range equations are obtained by minimizing two different objective functions: $F_1(x_D) = \|\underline{y}_D - F_D x_D\|^2_{W_D}$ and $F_2(x_D) = \|\underline{y}_D - F_D x_D\|^2_{W_D}$ subject to: $f(x_D) = 0$, with $x_D = [x_u^T,\ b_c,\ \kappa]^T$. The simplified solution is then obtained as:

$$\hat{\underline{x}}_D \ = \ \arg\min_{x_D} F_1(x_D), \tag{4.45}$$

and the rigorous solution is obtained by searching the stationary point of

$$F_2(x_D, \lambda) = \|\underline{y}_D - F_D x_D\|^2_{W_D} + \lambda f(x_D) \tag{4.46}$$

To obtain a rigorous least-squares solution non-iteratively without simplification is not possible, due to the presence of a non-linear constraint $f(x_D) = 0$. Other than directly using (4.45) as the final solution like SOF methods, the MOF methods derive the estimators by combining the following equations:

$$\partial_{[x_{D,1},\ldots x_{D,k-1},x_{D,k+1},\ldots]} F_1(x_D) = 0 \rightarrow \hat{\underline{x}}'_D = f_1(\underline{y}_D, x_{D,k}), \tag{4.47}$$

$$\partial_\lambda F_2(x_D, \lambda) = 0 \rightarrow f(x_D) = 0, \tag{4.48}$$

where in (4.47) one of the unknowns in x_D, e.g. $x_{D,k}$, is used to represent all other unknowns, ending up with an intermediate solution $\hat{\underline{x}}'_D = [\hat{\underline{x}}_{D,1}, \ldots \hat{\underline{x}}_{D,k-1}, x_{D,k}, \hat{\underline{x}}_{D,k+1}, \ldots]^T$. Since the non-linear constraint is absent in $F_1(x_D)$, the minimization of which should give a linear relation between $\hat{\underline{x}}'_D$ and $x_{D,k}$, i.e. (4.47) can be further written as $f_1(\underline{y}_D, x_{D,k}) = \underline{a} + \underline{b} x_{D,k}$. Combining the above two equations gives:

$$f(\underline{a} + \underline{b} x_{D,k}) = 0. \tag{4.49}$$

In positioning problems, the above equation is quadratic [42, 43, 44, 48, 49, 50, 53, 54], and two possible solutions of $\hat{\underline{x}}_{D,k}$ can be obtained. Together with (4.47), two sets of $[\hat{\underline{x}}_u^T,\ \hat{\underline{b}}_c]^T$ (part of x_D) can be calculated, and the final solution can be chosen as:

$$[\hat{\underline{x}}_u^T,\ \hat{\underline{b}}_c]^T = \arg \min_{\hat{\underline{x}}_{u,j},\ \hat{\underline{b}}_{c,j}} F(x_u, b_c),\ j = 1, 2, \tag{4.50}$$

where $F(x_u, b_c)$ is the objective function for non-linear least-squares defined in (3.50).

In case of an overdetermined system (3.2) with $m > n$, the MOF estimator (4.50) is not equivalent to the rigorous solution (4.46). Nevertheless, with a consistent set of equations, (4.50) and (4.46) are identical, since in this case:

$$\|\underline{y}_D - F_D x_D\|^2_{W_D} \ = \ 0 \tag{4.51}$$

$$f(x_D) \ = \ 0 \tag{4.52}$$

$$\lambda \ = \ 0 \tag{4.53}$$

In [43], Bancroft used the squared range equations (3.17) for estimation, and the constraint takes the form $f(x_D) = x_u^T x_u - b_c^2 - \kappa = 0$. This method was introduced for both overdetermined and consistent cases, i.e. $m \geq n$. In [44], the authors only covered the consistent case $m = n$, and the squared range-differences (3.34) are used instead with

$\kappa = d_r^2$. The constraint becomes $f(x_D) = \|x_u - x_r\|^2 - d_r^2 = 0$. Note that b_c is canceled in this case, and the author made the assumption that the reference transmitter is placed at the origin, i.e. $x_r = 0$. Later double range differences are used in [53] to form a *consistent* set of equations, and the constraint in use is equivalent to $f(x_D) = \|x_u - x_{rr}\| - \|x_u - x_r\| - (\underline{y}_{rr} - \underline{y}_r) = 0$. Similar methods can also be found in [42, 48, 49, 50, 54].

The mechanics of these MOF methods can be summarized as a process with two major steps:

1. Obtain a linear relation $\hat{x}'_D = a + bx_{D,k}$ by means of minimizing $\|y_D - F_D x_D\|^2_{W_D}$,
2. form a quadratic function $f(x_{D,k}) = 0$ based on the linear relation and the corresponding constraint to obtain $\hat{x}_{D,k}$, and then, $\hat{x}_D$.

The linear relations in the first step of the MOF methods that follow the above mechanics are equivalent, no matter they are obtained from squared ranges, squared-range differences or squared range-differences. In the second step, there exists an unlimited number of choices for the quadratic constraint.

To see this point, recall from (3.3), the range measurements can be expressed as,

$$\underline{y}_i = \underbrace{\|x_u - x_i\|}_{d_i} + b_c + \underline{e}_i,\ i = 1 \dots m, \tag{4.54}$$

which is equivalent to

$$\underline{y}_i - b_s = \|R(x_u - x_s) - R(x_i - x_s)\| + (b_c - b_s) + \underline{e}_i,\ i = 1 \dots m, \tag{4.55}$$

where x_s and b_s are arbitrary constant shifts added to the unknown x_u and b_c, and R is a rotation matrix, satisfying $R^T R = I$.

After squaring and rearranging, a set of equations with squared measurements can be obtained as:

$$\begin{aligned} & \underline{y}_i^2 - b_s^2 - x_i^T R^T R x_i + x_s^T R^T R x_s \\ = \ & (x_u - x_s)^T R^T R (x_u - x_s) - (b_c - b_s)^2 - 2(x_i - x_s)^T R^T R x_u + 2(\underline{y}_i - b_s) b_c \\ + \ & 2 d_i \underline{e}_i + \underline{e}_i^2, \end{aligned} \tag{4.56}$$

which is equivalent to the squared range equations in (3.17). The matrix form can be written in accordance with (4.24), where $\underline{y}_{D,i} = \underline{y}_i^2 - b_s^2 - x_i^T x_i + x_s^T x_s$, $A_{D,i} = -2(x_i - x_s)^T$, $\underline{B}_{D,i} = 2(\underline{y}_i - b_s)$, $C_{D,i} = 1$, $\underline{e}_{D,i} = 2 d_i \underline{e}_i + \underline{e}_i^2$, and

$$\kappa = (x_u - x_s)^T R^T R (x_u - x_s) - (b_c - b_s)^2. \tag{4.57}$$

Please note that the rotation matrix is always involved in the form $R^T R$, and therefore has no effect on the final output. On the other hand, since x_s and b_s can be set with any arbitrary values, there is an unlimited number of choices for κ.

In the consistent case, MOF methods with different constraints are equivalent. However when the system is overdetermined, MOF methods with different choices of the constraint

are not equivalent in general. To derive the expression of the best choice for the shifts $[x_s^T,\ b_s]^T$ in (4.57) is an non-trivial task, but $[x_s^T,\ b_s]^T = [x_u^T,\ b_c]^T$ seem a reasonable option since, in this case, the problem is linear with the non-linear term $\kappa = 0$, and a MOF method gives rigorous least-squares estimator to (4.24) as:

$$\underline{\hat{x}} = [\underline{\hat{x}}_u^T,\ \underline{\hat{b}}_c]^T = (\underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D \underline{y}_D, \tag{4.58}$$

where $\underline{D}_D = [A_D,\ \underline{B}_D]$.

Another way to interpret the choice of $[x_s^T,\ b_s]^T = [x_u^T,\ b_c]^T$ can be given as follows:

1. The Gauss-Newton method is chosen to solve squared range equations (3.17),
2. a Taylor expansion of (3.17) in the initial guesses x_0 and b_0:

$$\underline{y}_i^2 - b_0^2 - x_i^T x_i + x_0^T x_0 \approx -2(x_i - x_0)^T x_u + 2(\underline{y}_i - b_0) b_c + 2 d_i \underline{e}_i + \underline{e}_i^2 \tag{4.59}$$

   is used as the update equation,
3. the initial guess is somehow chosen as the true values of $[x_u^T,\ b_c]^T$, so $\approx$ in (4.59) becomes $=$, and the algorithm is terminated with only one iteration. The final solution is equivalent to (4.58) with $[x_0^T,\ b_0]^T = [x_s^T,\ b_s]^T = [x_u^T,\ b_c]^T$.

In practice, however, it is not possible to realize the above choice, since the true values are not known. A practical way to exploit the fact that MOF methods achieves rigorous least-squares solution when $[x_s^T,\ b_s]^T = [x_u^T,\ b_c]^T$, is to first use a SOF or MOF method to obtain an intermediate solution, then use it as the shift values in (4.56) to calculate (4.58) as the final estimate.

4.2.3 Relating SOF and MOF Methods

Although, with an overdetermined set of equations, MOF methods with different choices of $[x_s^T,\ b_s]^T$ are not equivalent in general, there is a special case when the MOF methods are equivalent, and the solution is also equivalent to a SOF solution.

Based on (4.56), and following the first step of the MOF mechanics (4.47) by choosing $x_{D,k} = \kappa$, we have:

$$\underline{\hat{x}} = [\underline{\hat{x}}_u^T,\ \underline{\hat{b}}_c]^T = (\underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D (\underline{y}_D - C_D \kappa). \tag{4.60}$$

In the second step of MOF methods, the above expression (4.60) should be plugged into $f(x_D) = (x_u - x_s)^T (x_u - x_s) - (b_c - b_s)^2 - \kappa = 0$ to work out two candidates for κ, since it is an intermediate solution which involves the unknown κ. However, with a specific choice of $\underline{D}_D^T W_D C_D = 0$, (4.60) is already the final solution and the second step of MOF methods can be avoided. Based on the fact that $C_{D,i} = 1$, this choice can be realized by choosing x_s and b_s as the weighted means of the base station positions and of the measurements,

respectively, i.e.

$$x_s = \frac{\sum_{i=1}^{m}\left[x_i \sum_{j=1}^{m} W_{D,i,j}\right]}{\sum_{i=1}^{m}\sum_{j=1}^{m} W_{D,i,j}}, \text{ and } \underline{b}_s = \frac{\sum_{i=1}^{m}\left[\underline{y}_i \sum_{j=1}^{m} W_{D,i,j}\right]}{\sum_{i=1}^{m}\sum_{j=1}^{m} W_{D,i,j}} \tag{4.61}$$

where $W_{D,i,j}$ stands for the entry of the weight matrix W_D on i-th row and j-th column.

Denoting $\underline{F}_D = [\underline{D}_D,\ C_D]$, the full solution for x_D with the choice in (4.61) can be represented as:

$$\hat{\underline{x}}_D = (\underline{F}_D^T W_D \underline{F}_D)^{-1} \underline{F}_D^T W_D \underline{y}_D,$$

which is equivalent to

$$\hat{\underline{x}}_D = \left(\underline{F}_D^T T_A^T \left(T_A W_D^{-1} T_A^T\right)^{-1} T_A \underline{F}_D\right)^{-1} \underline{F}_D^T T_A^T (T_A W_D^{-1} T_A^T)^{-1} T_A \underline{y}_D, \tag{4.62}$$

with T_A an invertible transformation matrix in (4.37).

Furthermore, since

$$T_A \underline{y}_D = \begin{bmatrix} \underline{y}'_D \\ \underline{y}_{D,m} \end{bmatrix},\ T_A \underline{F}_D = \begin{bmatrix} \underline{F}'_D & 0 \\ A_{D,m} & 1 \end{bmatrix},\ x_D = \begin{bmatrix} x'_D \\ \kappa \end{bmatrix} \tag{4.63}$$

with $(.)'_D$ the terms corresponding to squared-range difference equations, it can be proved in a similar way as in (4.38)-(4.42) that the solution for x_u and b_c in (4.62) is equivalent to the SOF solution obtained with squared-range difference equations.

To this end, it can be concluded that the SOF solution is actually a special case of the MOF solution with a specific choice of x_s and b_s in (4.61).

4.2.4 Geometric Interpretation

4.2.4.1 SOF Methods

The geometric interpretation of the SOF methods is probably best revealed with squared-range differences for ToA/RSS systems and with double differences for TDoA systems, where x_u is the only unknown. Since in both cases, the equations are linear in x_u, they represent *lines* in 2-D cases or *planes* in 3-D cases. Here, the interpretation is only given for 2-D for a clearer demonstration.

In ToA/RSS cases, each line is defined by the intersections of two circles, see Figure 4.1. Line i can be expressed as,

$$\left|\frac{(x_i - x_r)^T (x_u - x_r)}{\|x_i - x_r\|}\right| = l_i, \tag{4.64}$$

where l_i is the length of the segment between the reference transmitter x_r and the intersection point of the vector $x_r - x_i$ and line i. The unknown position then can be located

as the intersection of all these lines, when the system is consistent or measurements are error-free. In an overdetermined system with measurement errors, these lines generally do not intersect at a unique point. The geometric interpretation should then consider Figure 3.2, with a slight difference that the manifold $A(.)$ should be linear.

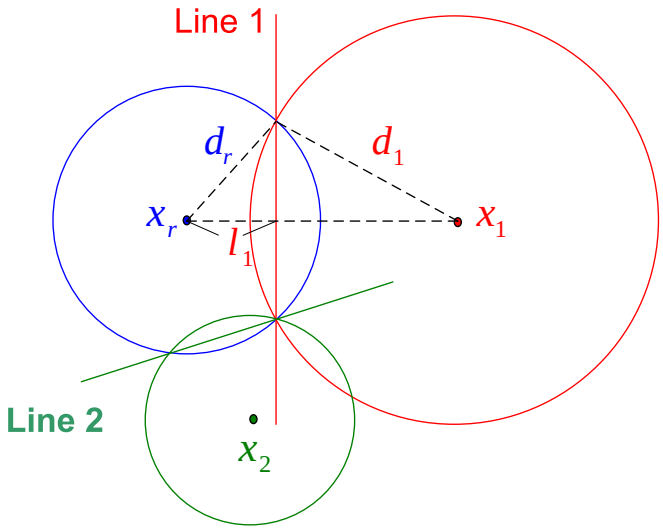

Figure 4.1: 2D geometric view of the intersections of two lines, obtained from two pairs of circles.

The calculation of l_i is simple, but should be done considering two different situations,

1. $(x_i - x_r)^T(x_u - x_r) \geq 0$, $d_r^2 - l_i^2 = d_i^2 - (\|x_i - x_r\| - l_i)^2 \Rightarrow l_i = \frac{d_r^2 - d_i^2 + \|x_i - x_r\|^2}{2\|x_i - x_r\|}$
2. $(x_i - x_r)^T(x_u - x_r) < 0$, $d_r^2 - l_i^2 = d_i^2 - (\|x_i - x_r\| + l_i)^2 \Rightarrow l_i = -\frac{d_r^2 - d_i^2 + \|x_i - x_r\|^2}{2\|x_i - x_r\|}$

Thus, together with (4.64), the following equation can be obtained

$$d_i^2 - d_r^2 = -2(x_i - x_r)^T x_u + x_i^T x_i - x_r^T x_r, \ i = 1, ..., m-1, \tag{4.65}$$

which is identical to the error free version of (3.26), with $b_c = 0$ for ToA/RSS cases.

In TDoA cases, each line is defined by the intersections of two hyperbolas. In [53], the equations linear in x_u that correspond to these lines, are found based on the fact that, for any point x_u on the hyperbola i, the ratio between the distance from x_u to x_r (the common focal point for all i) and the distance from x_u to the directrix (plane) is the eccentricity of the surface (see Figure 4.2). This gives

$$\|x_u - x_r\| = e_{c,i}\left[d_{x,i} - n_i^T(x_u - x_r)\right], \ i = 1, ...m-1 \tag{4.66}$$

where $e_{c,i} = \frac{\|x_i - x_r\|}{d_i - d_r}$ is the eccentricity of hyperbola i, $d_{x,i} = \frac{1}{2}\left[\|x_i - x_r\| - \frac{(d_i - d_r)^2}{\|x_i - x_r\|}\right]$ represents the (perpendicular) distance from x_r to the directrix, and $n_i = \frac{x_i - x_r}{\|x_i - x_r\|}$ is a unit vector pointing from x_r to the directrix.

Any two of the equations defined by (4.66) can be used to cancel the left hand side term $\|x_u - x_r\|$. Picking $rr = m - 1$ as the reference, subtracting it from all other equations gives a set of $m - 2$ linear equations:

$$(e_{c,i}n_i - e_{c,rr}n_{rr})^T(x_u - x_r) = e_{c,i}d_{x,i} - e_{c,rr}d_{x,rr}, \ i = 1, ...m-2, \tag{4.67}$$

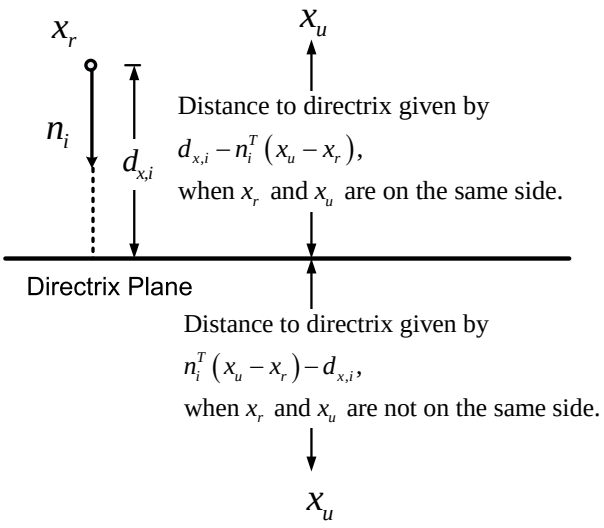

Figure 4.2: The directrix plane.

which, after rearranging, is identical to the error free version of (3.39).

4.2.4.2 MOF Methods

Here the geometric interpretation for the MOF method in [43], is shown for a 2-D ToA/RSS system with $\kappa = x_u^T x_u$. In this case, (4.47) can be written in a form as:

$$\begin{cases} x_{u,1} = a_1(x_{u,1}^2 + x_{u,2}^2) + b_1 \Rightarrow (x_{u,1} - \frac{1}{2a_1})^2 + x_{u,2}^2 = \frac{1-4a_1b_1}{4a_1^2} \\ x_{u,2} = a_2(x_{u,1}^2 + x_{u,2}^2) + b_2 \Rightarrow x_{u,1}^2 + (x_{u,2} - \frac{1}{2a_2})^2 = \frac{1-4a_2b_2}{4a_2^2} \end{cases} \tag{4.68}$$

where $x_{u,1}, x_{u,2}$ are the Cartesian coordinates of the user position, i.e. $x_u = [x_{u,1},\ x_{u,2}]^T$ and a_i, b_i are known constants.

Clearly, (4.68) represents two circles on x_{u1}-x_{u2} plane, centered at $(\frac{1}{2a_1}, 0)$ and $(0, \frac{1}{2a_2})$, with radius $\sqrt{\frac{1-4a_1b_1}{4a_1^2}}$ and $\sqrt{\frac{1-4a_2b_2}{4a_2^2}}$, respectively. The two intersections represent two possible solutions and the final estimate is chosen based on (4.50).

In a fully determined system, with $m = n = 2$, these two circles intersect at two points which are exactly the two intersections of the two measurement circles, centered at x_1 and x_2 with elements in $\underline{y}$ as radius, see Fig 4.3. Without additional information about the unknown user position x_u, it is not possible to tell which intersection point corresponds to the true position of the user, since they both satisfies $F(x) = 0$, with $F(x)$ defined in (3.50).

When $m = 3 > n = 2$ and $\underline{e} = 0$, it is much easier to pick the final solution, since in this case, the five circles, including two circles of (4.68) and three measurement circles, intersect at a unique point, which is the true user position, see Fig 4.4.

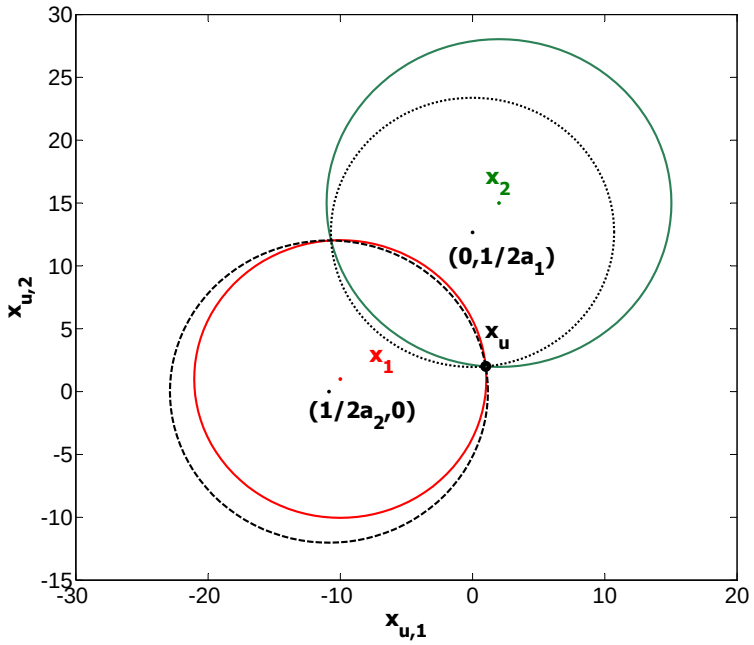

Figure 4.3: 2D geometric view of the MOF solution in determined case with 2 measurements. Solid: measurement circles; Dashed/Dotted: two circles represented by (4.68).

4.2.5 Error Analysis

4.2.5.1 SOF

ToA/RSS For the SOF methods, the expressions of the first and second moments of a linear least-squares estimator are rather simple. In case of an error-free design matrix F_D:

$$\mu_{\hat{x}_D} = (F_D^T W_D F_D)^{-1} F_D^T W_D \mathrm{E}\{\underline{y}_D - y_D\}, \tag{4.69}$$

$$Q_{\hat{x}_D \hat{x}_D} = (F_D^T W_D F_D)^{-1} F_D^T W_D Q_{y_D y_D} W_D F_D (F_D^T W_D F_D)^{-1}. \tag{4.70}$$

TDoA In this case, however, the design matrix F_D contains measurements $\underline{y}$. The derivations for the first and second moments of $\underline{\hat{x}}_D$ then become much more complicated and some approximations are required. As we already proved that SOF estimators can be seen as estimators obtained with a unique method applying different weighting matrices, here we show the error analysis for the SOF estimator using squared-range differences (3.26). The system contains $m-1$ equations with $n+1$ unknowns. The expression of $\underline{\hat{x}}_D$ (4.26) is equivalent to:

$$\underline{\hat{x}}_u = (A_D^T W_D A_D)^{-1} A_D^T W_D (\underline{y}_D - \underline{B}_D \hat{\underline{b}}_c) \tag{4.71}$$

$$\hat{\underline{b}}_c = \frac{\underline{B}_D^T P_A^{\perp T} W_D P_A^{\perp} \underline{y}_D}{\underline{B}_D^T P_A^{\perp T} W_D P_A^{\perp} \underline{B}_D} \tag{4.72}$$

$$\underline{\hat{x}}_D = G\underline{y}_D - G\underline{B}_D \hat{\underline{b}}_c + F\hat{\underline{b}}_c \tag{4.73}$$

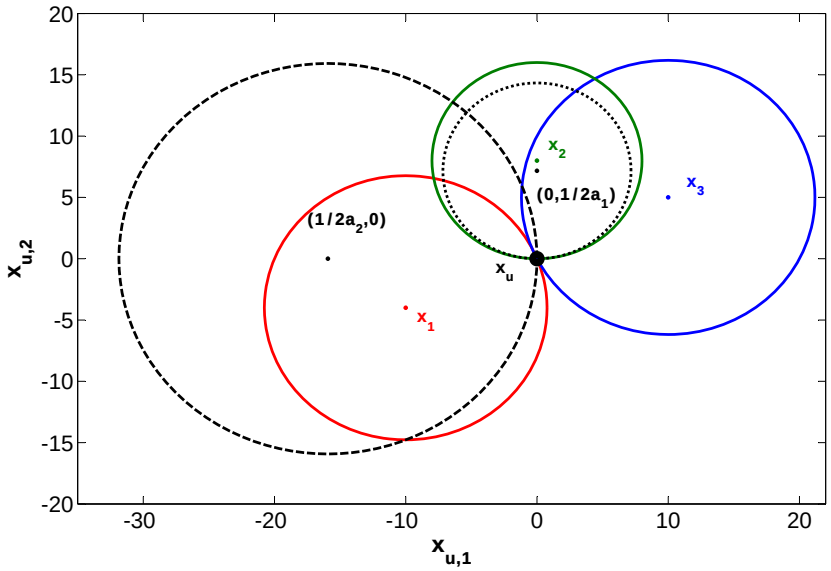

Figure 4.4: 2D geometric view of the MOF solution in consistent case with 3 measurements. Solid: measurement circles; Dashed/Dotted: two circles represented by (4.68).

with

$$\begin{aligned} P_A^{\perp} &= [I - A_D(A_D^T W_D A_D)^{-1} A_D^T W_D]_{(m-1)\times(m-1)} \\ G &= \begin{bmatrix} (A_D^T W_D A_D)^{-1} A_D^T W_D \\ 0 \end{bmatrix}_{m\times(n+1)} \\ F &= \begin{bmatrix} 0_{n\times 1} \\ 1 \end{bmatrix}_{(n+1)\times 1} \end{aligned}$$

where the expressions for A_D and $\underline{B}_D$ can be found in (3.26).

The expression of $\underline{\hat{b}}_c$ in (4.72) contains $\underline{y}$ and therefore contains error $\underline{e}$. Taylorizing $\underline{\hat{b}}_c(\underline{e})$ at $e = E\{\underline{e}\} = 0$ gives:

$$\underline{\hat{b}}_c(\underline{e}) = b_c(0) + \partial_{e^T} b_c(0)\underline{e} + \frac{1}{2}\underline{e}^T \partial_{ee^T} b_c(0)\underline{e} + \cdots \tag{4.74}$$

$$\tag{4.75}$$

and the expressions for $\partial_{e^T} b_c(0)$ and $\partial_{ee^T} b_c(0)$ are given in Appendix A.2.1.

Define the bias in $\underline{\hat{b}}_c$ as $\underline{\mu}_{\hat{b}_c} = \underline{\hat{b}}_c - b_c$, which is given, together with the variance of $\underline{\hat{b}}_c$, as

$$\mu_{\hat{b}_c} = \mathsf{E}\{\underline{\mu}_{\hat{b}_c}\} = \frac{1}{2} trace\{\partial_{ee^T} b_c(0) Q_{ee}\}, \tag{4.76}$$

$$Q_{\hat{b}_c\hat{b}_c} = Q_{\mu_{\hat{b}_c}\mu_{\hat{b}_c}} = \mathsf{E}\{(\underline{\mu}_{\hat{b}_c} - \mu_{\hat{b}_c})(\underline{\mu}_{\hat{b}_c} - \mu_{\hat{b}_c})^T\} = \partial_{e^T} b_c(0) Q_{ee} \partial_e b_c(0), \tag{4.77}$$

where the terms with order higher than 2 are ignored.

The bias and variance of $\underline{\hat{x}}_D$ can be given based on (4.76) and (4.77) as,

$$\mu_{\hat{x}_D} = \mathsf{E}\{\underline{\hat{x}}_D - x_D\} = G\mu_{e_D} - GK\partial_e b_c(0) + (F - GB_D)\tfrac{1}{2} trace\{\partial_{ee^T} b_c(0) Q_{ee}\}, \tag{4.78}$$

$$\begin{aligned} Q_{\hat{x}_D\hat{x}_D} &= \mathsf{E}\{(\underline{\hat{x}}_D - x_D - \mu_{\hat{x}_D})(\underline{\hat{x}}_D - x_D - \mu_{\hat{x}_D})^T\} \\ &= G\mathsf{E}\{\underline{e}_D \underline{e}_D^T\} G^T + J + J^T + (F - GB_D)\partial_{e^T} b_c(0) Q_{ee} \partial_e b_c(0)(F - GB_D)^T - \mu_{\hat{x}_D}\mu_{\hat{x}_D}^T, \end{aligned} \tag{4.79}$$

where the expressions of K and J can be found in Appendix A.2.1.

4.2.5.2 MOF

Here, we use the MOF method in [43] as an example, i.e. $\underline{y}_{D,i} = \underline{y}_i^2 - x_i^T x_i$, $A_{D,i} = -2x_i^T$, $\underline{B}_{D,i} = 2\underline{y}_i$, $C_{D,i} = 1$, $\underline{e}_{D,i} = 2d_i\underline{e}_i + \underline{e}_i^2$, and $\kappa = x_u^T x_u - b_c^2$. The analysis for other MOF methods can be done in a similar way.

Recall from (4.60), the expression for $\hat{\underline{x}}_D$ is:

$$\hat{\underline{x}}(\underline{e}) = (\underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D (\underline{y}_D - C_D \hat{\underline{\kappa}}). \tag{4.80}$$

and the expressions for two possible values of $\hat{\underline{\kappa}}$ can be obtained by plugging (4.80) into $\kappa = x_u^T x_u - b_c^2$, which gives:

$$\hat{\underline{\kappa}}^{\alpha,\beta} = \frac{-\underline{b} \pm \sqrt{-2\underline{b} - 1}}{2\underline{a}} \tag{4.81}$$

with

$$\underline{a} = C_D^T W_D \underline{D}_D (\underline{D}_D^T W_D \underline{D}_D)^{-1} (\underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D C_D \tag{4.82}$$

$$\underline{b} = -2C_D^T W_D \underline{D}_D (\underline{D}_D^T W_D \underline{D}_D)^{-1} (\underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D \underline{y}_D - 1 \tag{4.83}$$

TDoA Obviously, (4.80) involves the original measurement $\underline{e}$ in a complicated non-linear way. Simplifications are required to evaluate the bias and variance of $\hat{\underline{x}}(\underline{e})$. The analysis can be done by following the general way described in Section 3.3.4, but our case here is simpler since the expressions for $\hat{\underline{x}}$, $\hat{\underline{\kappa}}$ are available in closed-form.

A Taylor expansion of $\hat{\underline{x}}(\underline{e})$ at $\mathsf{E}\{\underline{e}\} = 0$ gives the following expansion in $\underline{e} = \underline{e} - \mathsf{E}\{\underline{e}\}$:

$$\hat{\underline{x}}(\underline{e}) = x(0) + \partial_i x(0)\underline{e}^i + \frac{1}{2}\partial_{ij}^2 x(0)\underline{e}^i\underline{e}^j + \cdots, \quad i,j = 1,...,m \tag{4.84}$$

where the Einstein's summation convention is used, with $\partial_i(.)$ denoting the partial derivative of $(.)$ w.r.t. the i-th element of $\underline{e}$, denoted as $\underline{e}^i$.

Ignoring the terms of $\underline{e}$ with order higher than 2, together with (4.84), the bias in the estimator, defined as: $\mu_{\hat{x}} = \mathsf{E}\{\hat{\underline{x}}(\underline{e}) - x(0)\}$, and the variance matrix, defined as: $Q_{\hat{x}\hat{x}} = \mathsf{E}\left\{(\hat{\underline{x}} - \mathsf{E}\{\hat{\underline{x}}\})(\hat{\underline{x}} - \mathsf{E}\{\hat{\underline{x}}\})^T\right\}$, can be approximated as:

$$\mu_{\hat{x}} = \frac{1}{2}\mathsf{E}\left\{\underline{e}^i\partial_{ij}^2 x(0)\underline{e}^j\right\} = \frac{1}{2}\partial_{i^2}^2 x(0)\mathsf{E}\{(\underline{e}^i)^2\} = \frac{1}{2}\partial_{i^2}^2 x(0)\sigma_i^2, \tag{4.85}$$

$$Q_{\hat{x}\hat{x}} = \mathsf{E}\left\{\partial_i x(0)\underline{e}^i\underline{e}^j\partial_j x^T(0)\right\} = \partial_i x(0)\sigma_i^2\partial_i x^T(0). \tag{4.86}$$

where the fact that $\mathsf{E}\{\underline{e}^i\underline{e}^j\} = 0,$ for $i \neq j$, has been used.

The rest of the analysis is to derive the expressions for $\partial_i x(0)$ and $\partial_{i^2}^2 x(0)$, which is done in Appendix A.2.2.

ToA/RSS The error analysis for ToA/RSS cases can be derived in a similar way as what has been done for TDoA cases. In fact, the whole procedure is much easier with D_D free of measurement errors. In fact, the results derived for TDoA cases can be directly adopted by setting $\partial_i D_D = 0$, $\partial_i(D_D^T W_D D_D)^{-1} = 0$, and using proper dimensions.

4.2.6 Comparison

A comparison of the reviewed direct methods can be done based on their basic principles.

The SOF methods are widely documented in literature [42, 45, 46, 47, 51, 52, 55, 56, 57, 58, 59]. It has been shown that these methods are essentially equivalent. There methods are mutually related through their measurement weighting schemes, and they share the following strong and weak points:

- Strong Point: The final solution is uniquely obtained from a set of *linear* equations. As compared to the MOF methods, no extra quadratic equation needs to be solved and there is no need to check certain criteria to pick the final estimate out of several candidates.
- Weak Point: The non-linear constraint is neglected (loss of information). Thus, the SOF methods work only with a redundant set of equations (3.2) where $m > n$. Moreover, the SOF methods may face numerical problems in TDoA systems, which will be shown later in Section 4.4.

The reviewed MOF methods include: [42, 43, 44, 48, 49, 50, 53], which are in general not equivalent to each other. The MOF methods have the following characteristics:

- Strong Point: The constraint is (partially) exploited.
- Weak Point: In contrast to SOF methods, one extra step of solving a quadratic function is required to obtain two candidate estimates, from which the final estimate is chosen with certain criteria.

4.3 Constrained Iterative Methods

Based on the discussions in the previous section, the SOF methods and the MOF methods are capable of finding the position estimate non-iteratively with some simplifications. However, with the employed simplifications, neither SOF methods nor MOF methods provide rigorous least-squares solutions to the squared range equations, with or without differencing.

In this section, another type of iterative methods, that bring improvements on estimation accuracy against the direct methods, is reviewed. They estimate the user position by fully exploiting the constraint $f(x_D) = 0$. In literature, through EVD, the solution is found by searching the roots of a high order polynomial of the Lagrange multiplier λ. The interesting feature of this type of methods is that in each iteration, only one single parameter needs to be calculated.

The methods to be reviewed include [61, 62], where the solution is exactly as (4.22):

$$\underline{\hat{x}}_D = \arg\min_{x_D,\lambda} \|\underline{y}_D - F_D x_D\|^2_{W_D} \text{ subject to: } f(x_D) = 0. \tag{4.87}$$

which is obtained by searching the stationary point of the Lagrange function:

$$\|\underline{y}_D - F_D x_D\|^2_{W_D} + \lambda f(x_D). \tag{4.88}$$

The mechanics can be described as follows,

1. Taking the partial derivative of (4.88) w.r.t. x_D and setting it to zero, gives an expression of $\hat{\underline{x}}_D$ in λ. By means of EVD (explained later in this section), one can transform x_D to x'_D, by $x'_D = U^{-1}x_D$ (U obtained with EVD), and each element of $\hat{\underline{x}}'_D$ can be further written as an explicit expression in λ, i.e. $\hat{\underline{x}}'_{D,k} = g_k(\lambda)$, $k = 1, ..., \dim(x'_D)$.
2. Taking the partial derivative of (4.88) w.r.t. λ and setting it to zero, gives $f(x_D) = f'(x'_D) = 0$.
3. Plugging $\hat{\underline{x}}'_{D,k} = g_k(\lambda)$ into $f'(x'_D) = 0$ gives an explicit high order polynomial in λ, which can be solved with the reviewed iterative search techniques. With the single unknown λ in this step, the iteration is performed in 1-D.
4. Once the estimator $\hat{\underline{\lambda}}$ is available, $\hat{\underline{x}}_D$ can be obtained via $\hat{\underline{x}}_D = U\hat{\underline{x}}'_D$ with $\hat{\underline{x}}'_{D,k} = g_k(\hat{\underline{\lambda}})$.

As an example, we show the full mechanics of the constrained iterative approach applied to solve a set of squared range measurements (3.17) with a clock offset b_c. With the specific choice of $\underline{y}_{D,i} = \underline{y}_i^2 - x_i^T x_i$, $A_{D,i} = -2x_i^T$, $\underline{B}_{D,i} = 2\underline{y}_i$, $C_{D,i} = 1$, $\kappa = x_u^T x_u - b_c^2$, and $f(x_D) = x_u^T x_u - b_c^2 - \kappa$, Equation (4.88) can be further elaborated as:

$$\|\underline{y}_D - F_D x_D\|^2_{W_D} + \lambda(x_D^T L x_D + s^T x_D), \text{ with} \tag{4.89}$$

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{(n+2)\times(n+2)}$$

$$s = [0 \ \cdots \ 0 \ 0 \ -1]^T_{(n+2)\times 1}$$

Following Step 1 of the mechanics described above, taking the partial derivative of (4.89) w.r.t. x_D and setting it to zero, we obtain:

$$\begin{aligned} (\underline{F}_D^T W_D \underline{F}_D + L\lambda)\hat{\underline{x}}_D &= \underline{F}_D^T W_D \underline{y}_D - \frac{1}{2}\lambda s \\ \Rightarrow (I + (\underline{F}_D^T W_D \underline{F}_D)^{-1} L\lambda)\hat{\underline{x}}_D &= (\underline{F}_D^T W_D \underline{F}_D)^{-1}(\underline{F}_D^T W_D \underline{y}_D - \frac{1}{2}\lambda s) \end{aligned} \tag{4.90}$$

Decompose $(\underline{F}_D^T W_D \underline{F}_D)^{-1} L$ as

$$(\underline{F}_D^T W_D \underline{F}_D)^{-1} L = U \Lambda U^{-1}, \tag{4.91}$$

where Λ is an diagonal matrix, which contains the eigenvalues of $(\underline{F}_D^T W_D \underline{F}_D)^{-1} L$.

Eq (4.90) can be further elaborate as:

$$\begin{aligned} \underline{\hat{x}}'_D = U^{-1}\underline{\hat{x}}_D &= (I+\Lambda\lambda)^{-1}U^{-1}(\underline{F}_D^T W_D \underline{F}_D)^{-1}(\underline{F}_D^T W_D \underline{y}_D - \frac{1}{2}\lambda s) \\ \Rightarrow \underline{\hat{x}}'_{D,k} &= \frac{\breve{\underline{y}}_k - \lambda \breve{\underline{s}}_k}{1+\lambda\Lambda_{kk}}, \quad k = 1, ..., n+2 \end{aligned} \tag{4.92}$$

where $\breve{\underline{y}} = U^{-1}(\underline{F}_D^T W_D \underline{F}_D)^{-1}\underline{F}_D^T W_D \underline{y}_D$, and $\breve{\underline{s}} = \frac{1}{2}(\underline{F}_D^T W_D \underline{F}_D)^{-1} s$.

To this end, Step 1 of the mechanics is finished. In Step 2, taking the partial derivative of (4.88) w.r.t. λ and setting it to zero, yields:

$$\begin{aligned} 0 &= \underline{\hat{x}}_D^T L \underline{\hat{x}}_D + s^T \underline{\hat{x}}_D \\ &= \underline{\hat{x}}_D'^T U^T L U \underline{\hat{x}}'_D + s^T U \underline{\hat{x}}'_D \end{aligned} \tag{4.93}$$

Plugging (4.92) into this equation gives a polynomial $p(\lambda)$. When the order of $p(\lambda)$ is higher than two, one usually needs to resort to iterative manner such as Newton's method to search the root. The equation to update estimate of $\hat{\lambda}$ in k-th iteration then takes the form:

$$\hat{\lambda}_k = \hat{\lambda}_{k-1} - \frac{f(\hat{\lambda}_{k-1})}{\partial_\lambda f(\hat{\lambda}_{k-1})}. \tag{4.94}$$

Once $\hat{\lambda}$ is estimated, $\hat{x}'_D$ can be obtained using (4.92), and the final estimate is calculated as $\hat{x}_D = U\hat{x}'_D$.

4.3.1 Discussion

Advantages The advantages of the constrained iterative method are:

1. The method provides estimates that closely approximate the *rigorous* least-squares solution to (4.21) from iteration to iteration. With the constraint $f(x_D) = 0$ fully exploited, the method achieves better accuracy than the direct methods which apply simplifications. However, it is hard to compare the solution (4.88) of the constrained iterative method with the traditional non-linear least-squares position estimator $\hat{x} = \arg\min_x \|y - A(x)\|_W^2$, in terms of accuracy, since they are obtained by minimizing different objective functions.

2. The heavy computations required by the evaluation of $\partial^2_{xx^T} F(x)$ in the (iterative) Newton method is greatly relieved with $\dim(x) = 1$. Thus one can enjoy the quadratic rate of convergence brought by the Newton method with much less computations than in the case of $\dim(x) = n$.

3. It is very easy to find the initial guess of λ, which is $\lambda_0 = 0$.

Eigen Value Decomposition We should also note that the EVD applied in both [61, 62] is itself a non-linear process, which usually needs to be solved iteratively. Actually, the reason of applying the EVD is to obtain (4.92), where each element of $\hat{\underline{x}}'_D$ is explicitly written as a function of λ. This, however, is helpful only when the roots of the polynomial $f(\lambda)$ can be found in closed forms. Otherwise, the EVD can be skipped and one can still solve the problem with only 1 unknown parameter, λ, in each iteration.

Combining (4.90), (4.94) and (4.94), we have

$$\hat{\underline{x}}_D(\lambda) = (\underline{F}_D^T W_D \underline{F}_D + L\lambda)^{-1}(\underline{F}_D^T W_D \underline{y}_D - \frac{1}{2}\lambda s) \tag{4.95}$$

$$f(\lambda) = \hat{x}_D \lambda^T L \hat{x}_D(\lambda) + s^T \hat{x}_D(\lambda) \tag{4.96}$$

$$\hat{\lambda}_k = \hat{\lambda}_{k-1} - \frac{\hat{x}_D(\lambda_{k-1})^T L \hat{x}_D(\lambda_{k-1}) - s^T \hat{x}_D(\lambda_{k-1})}{2\hat{x}_D(\lambda_{k-1})^T L \partial_\lambda \hat{x}_D(\lambda_{k-1}) - s^T \partial_\lambda \hat{x}_D(\lambda_{k-1})}. \tag{4.97}$$

The derivation of $\partial_\lambda \hat{x}_D(\lambda_{k-1}$ is given in Appendix A.3.1.

One should however also note that even if the EVD is skipped, the evaluation of (4.97) in each iteration is still quite complicated, since one needs to inverse a $(n+2)\times(n+2)$ matrix $(\underline{F}_D^T W_D \underline{F}_D + L\lambda)$, and perform matrix multiplications afterwards.

Error Analysis The error analysis can be done by following the general way as described in Section 3.3, and the results can be found in [62] which is not repeated here.

4.4 Validation and Results

In this section, three validation works are carried out:

1. The idea of using a direct method to obtain an initial guess is verified with actual audio UWB timing-based range measurements and simulations. The chosen direct method is the MOF method in [43] and the iterative method is the Gauss-Newton method.

2. The criteria (3.69) to check the bias due to non-linearity has been tested. Both actual audio UWB timing-based range measurements and simulations are used for this purpose. The Gauss-Newton method is used in positioning.

3. The positioning performance in terms of RMSE of the reviewed methods are compared based on actual radio UWB timing-based range measurements. The tested methods include the SOF method, the MOF method with $\kappa = x_u^T x_u - b_c^2$ [43], the constrained iterative method and the Gauss-Newton method.

The audio and radio UWB measurements are obtained using systems described in Section 2.2, where the details about ranging accuracy and positioning geometries can be found.

The implementation for direct methods are straightforward, but for iterative methods, several issues need to be covered:

- *Initial guess*. For the first validation work, both the true value and the outcome of the MOF method [43] are tested. For the bias check and the comparison of reviewed methods, we always use the true value as the initial guess. In this way, we exclude the errors in the final position estimates caused by a bad choice of the initial guess.

- *Successful convergence*. A successful convergence occurs and the algorithm is stopped when the update in the numerical estimate becomes negligible within 50 iterations, e.g.:

$$\|x_k - x_{k-1}\|_{Q^{-1}_{\hat{x}_k\hat{x}_k}} < 10^{-7}. \tag{4.98}$$

- *Convergence problem*. If the above inequality is not met after 50 iterations, the algorithm will be terminated, the estimate will not be considered, but the occurrence is counted.

The transmitters and the receiver in our audio and radio systems are fully synchronized, which enables us to perform the validation work for both ToA and TDoA systems. For a TDoA system, we simply assume that there exists a common clock offset (with true value 0) and it is treated as an additional unknown parameter. The verification using RSS measurements is not done here, but it can be easily performed in the same way as for ToA measurements.

4.4.1 Initial Guess

First of all, the idea of using the MOF method for the initial guess of the Gauss-Newton method is positively supported by both the experimental results and the simulations in general. For all four tested geometries under different signal bandwidths with actual data, using the true position (and true b_c if considered) as the initial guess and using the initial guess obtained with the MOF method make no difference in the final estimates of the Gauss-Newton method. They were the same in 100% of the tests with Geometry 1-4.

In Figure 4.5, we show the convergence procedures of two scenarios with Geometry 3 using actual audio measurements, no clock offset, bandwidth 0.567 kHz and STD 11.86 cm. In Scenario A, the initial guess is identical to the true position, $x_0 = x_{\text{true}}$ and in Scenario B, the initial guess is obtained using the MOF method. The final position estimates of the Gauss-Newton method in two scenarios are identical. Meanwhile, we notice that in Scenario B, the initial guess is closer to the final estimate than in Scenario A in this case with considerable ranging STD of 11.86 cm, which helps the Gauss-Newton to converge with fewer iterations.

Secondly, in some particular simulated cases where a clear local 'valley' exists, e.g. with near coplanar configuration such as the system in Figure 3.3(a) and Figure 3.3(b), the situation is more complicated. If the measurement error is small, not large enough to make the bottom value of the local 'valley' smaller than that of the global 'valley' (Figure 3.3(a)), the MOF method is still a very good choice for the initial guess. If the measurement error is large, then it can happen that the global minimum does not correspond to the correct position (Figure 3.3(b)) and the MOF method will give a wrong initial guess.

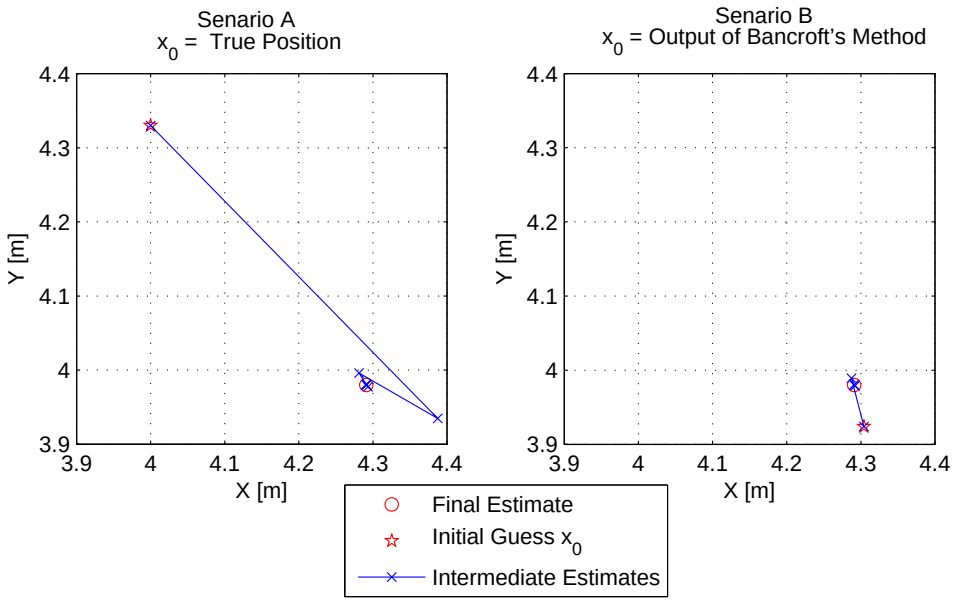

Figure 4.5: The convergence procedure on x-y plane of Geometry 3 (actual audio measurements) with no clock offset with a bandwidth of 0.567 kHz and STD 11.86 cm. On the left, the true position is chosen as the initial guess and on the right, MOF method is used for that purpose.

In the latter case, however, the MOF method should not be blamed for giving a 'wrong' initial guess, since the obtained solution is better than the true position in the sense of least-squares, which globally minimizes the objective function. So as we suggested in Section 3.3.2, coplanar and near coplanar configurations of the system should be avoided. One simple way to quantify 'healthiness' of the system configuration is to check the ratio d_{rt}/d_{ss} defined in the coming Section 4.4.2, where a smaller value indicates better geometry.

Nevertheless, if we somehow manage to choose an initial guess within the 'valley' pointing at the true position, the Gauss-Newton method will still converge to the correct solution. In practice, this requires additional information. In Figure 4.6 we show an example where the convergence procedures of the two scenarios where A) the initial guess is the true value and B) the initial guess is obtained using the MOF method. The system configuration is close to coplanar, similar as in Figure 3.3(a). Clearly, different choices of initial guess can lead the Gauss-Newton method converging to different least-squares solutions.

4.4.2 Bias Check

4.4.2.1 Bias Check for Systems with Clock Offset

Here, we present the obtained results when a clock offset is assumed to exist in the system, i.e., the unknown parameters are x_u and b_c. The verification of (3.65) is accomplished with simulation results and other results are obtained experimentally.

Verification of (3.65) Due to the limited number of measurements, we resort to simulations for the verification. In the simulation, the range error is assumed to be normally

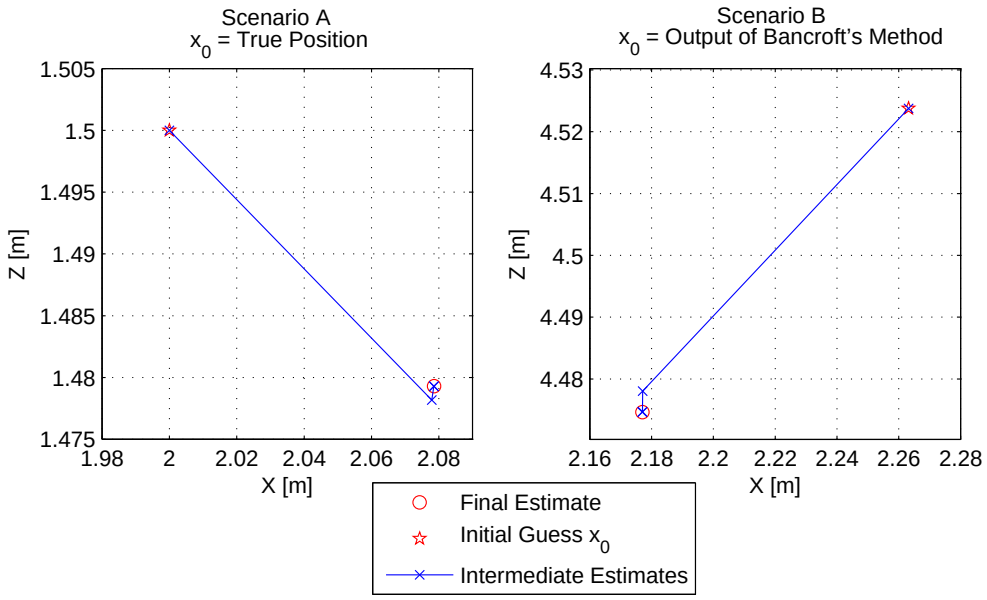

Figure 4.6: The convergence procedure on x-z plane. The 4 transmitters and the receiver are placed at (0, 0, 3), (0, 6, 3), (6, 6, 3), (12, -6, 1.5) and (2, 2, 1.5), respectively. The system has a clock offset equal to zero and the range measurement errors are sampled from a Gaussian distribution with zero mean and standard deviation 0.1 m. On the left, the true position and the true clock offset is chosen as the initial guess and on the right, MOF method is used for that purpose.

distributed with zero mean, and the tested error STDs range from 0.01 to 0.3 m. For each STD, 100000 sets of estimates are collected to obtain the empirical bias vector by taking the mean. A number of system geometries have been simulated and similar behaviors have been observed. As one example, the obtained straight norm of $\mu_{\hat{x}}$ in (3.65) for one of the geometries is shown in Figure 4.7. In this geometry, the loudspeakers are placed at (0, 0, 3), (6, 6, 3), (6, 0, 0), (0, 6, 0) and the microphone at (7, 7, 1.5), a bit similar to geometry 3 in Figure 2.7, but scaled up.

As a general expectation, (3.65) should be a very good approximation to the non-linearity bias when the range error STD is small. However, as the STD increases (so does the STD of position estimator), the quality of the approximation become worse. The reason is twofold:

1. Equation (3.65) considers only the second order remainder in (3.61). This means for the approximation to be good, the accuracy (in terms of STD on three Cartesian axes) of the estimators should be high, which requires a moderate to low noise STD in the range measurements. As the measurement STD goes larger, the position estimators are less accurate, and the theoretical curve obtained using (3.65) becomes less than the empirical curve since the terms with order higher than two become non-negligible. The non-negligible part appears at the measurement STD range about 0.1 to 0.15 m.

2. In Figure 4.7, the empirical curve gets below the theoretical curve when measurement STD becomes larger than about 0.15 m. This is due to the fact that with large

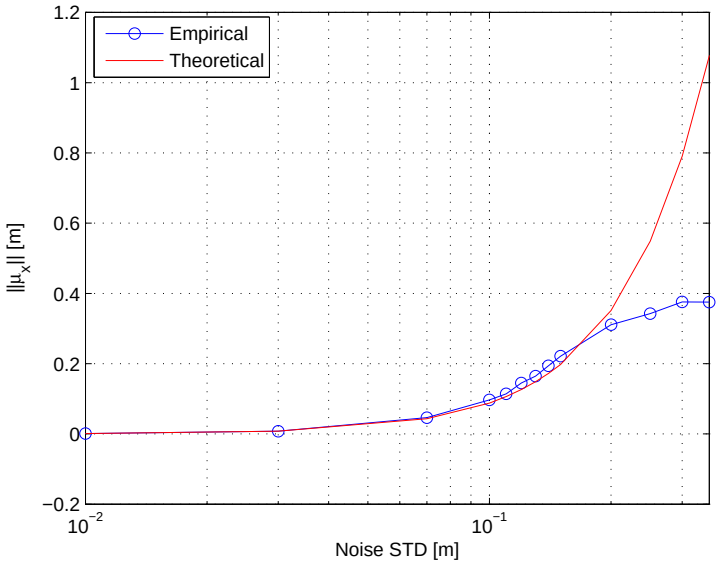

Figure 4.7: Empirical value through simulation and theoretical value (3.65) of non-linearity bias magnitude (case with clock offset).

measurement STD, the Gauss-Newton method may need more than 50 iterations to converge. Such an occurrence is consided as a convergence problem and the measurements are discarded. This means we manually changed the statistics of position estimates obtained with large STDs, which caused the deviation between the theoretical curve and the empirical curve.

This conclusion holds for all the results we obtained with simulations.

Positioning Results First of all, the positioning accuracy we achieved with the audio system using full bandwidth is very encouraging. With the assumption that the position estimators errors follow a Gaussian distribution, the region in the x-y plane with 95% confidence is shown in Figure 4.8.

As we can see, the semi-major axis of the ellipse is less than 5 cm.

Also, the empirical CDF for the errors in estimated clock offset in cm (the true value is 0) is shown in Figure 4.9, and the error is within 1.5 cm.

Furthermore, the analysis of the bias due to non-linearity has been carried out based on (3.69), and the results are summarized in Table 4.2. The parameters of interest are:

- Mean of $\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}}$.
- The 95% value of $\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}}$: 95% of the cases are within $[0, \|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}95\%}]$.
- Occurrence of $\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}} > 1$, in percentage.
- System scale. The system scale is defined as the mean value of the distances between the transmitters and the geometric center of the transmitters, denoted as d_{ss}.

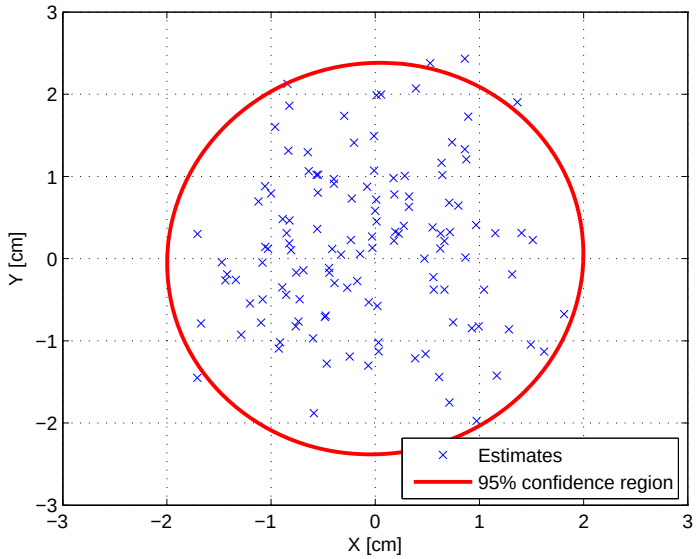

Figure 4.8: Region in the x-y plane with 95% confidence for the estimated positions of Geometry 4 using full bandwidth and the origin represents the true postion.

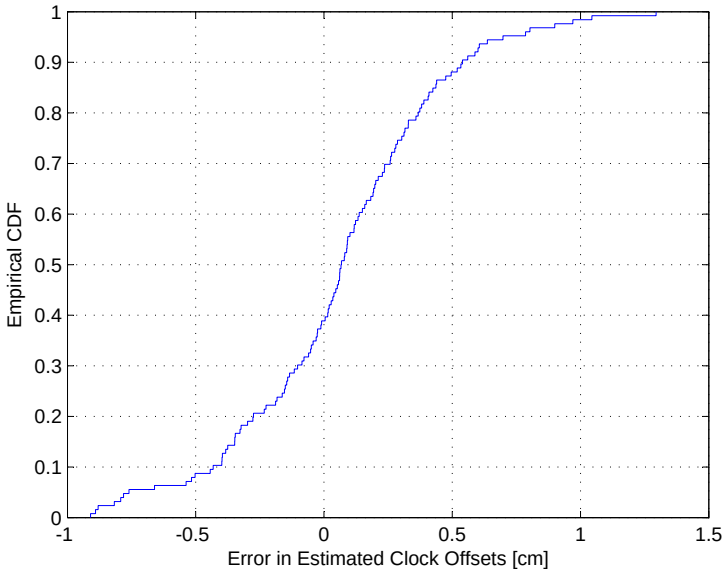

Figure 4.9: Empirical CDF for the errors in estimated clock offset when testing Geometry 4 using 128 full bandwidth measurements. The true value is equal to 0.

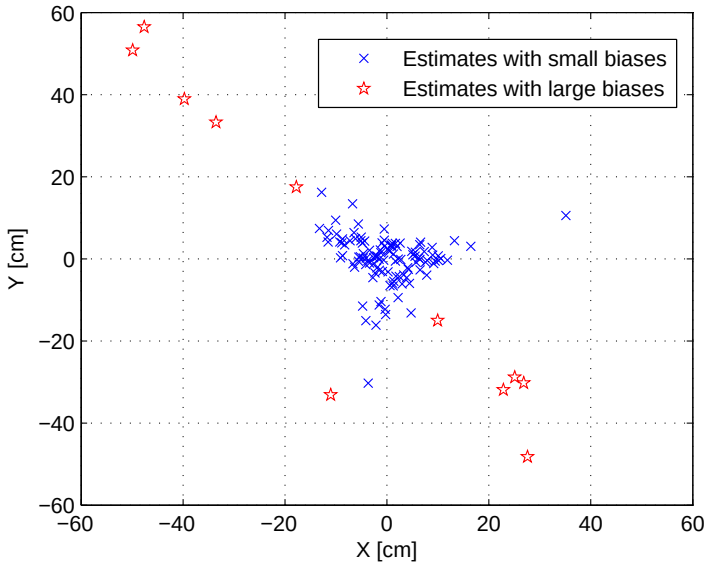

Figure 4.10: The estimates in the x-y plane for Geometry 3 with range error STD of 1.75 cm. In 9% of the cases, the bias due to non-linearity is significant as indicated by measure (3.69), and the corresponding estimates are plotted as red stars.

- The distance between the receiver and the geometric center of the transmitters, denoted as d_{rt}. For our measurements, the mean value is used, since for each geometry we tested more than one receiver position.
- the value of d_{rt}/d_{ss}.

The system performances can be ranked as Geometry 4 $>$ Geometry 1 $>$ Geometry 2 $>$ Geometry 3, from good to poor; the convergence problem happens when Geometry 3 is tested with an error STD 11.86 cm.

As mentioned before, the bias due to non-linearity in the system is related to the system geometry and the quality of the estimator (also the quality of the range estimation). The parameters d_{rt}, d_{ss} and their ratio can be used as a first simple measure of the 'health' of the geometry. For a system with non-coplanar transmitters, a large value of d_{ss} simply means that the space, where good positioning service can be obtained, is large. Meanwhile, the parameter d_{rt} describes how far the receiver is from the 'good service' space. In Table 4.2, it is clear that the geometries with larger d_{ss} and smaller d_{rt}/d_{ss} have better performance. Also, a system with larger d_{ss} and smaller d_{rt}/d_{ss} tends to have less convergence problems (this conclusion is mainly based on the simulation results which are not shown here), with the same level of error STD. So, we recommend to design systems with largest d_{ss} values, if possible, and keep the values of d_{rt}/d_{ss} at least smaller than 1 at the same time.

Geometry 3 has the smallest d_{ss} and the ratio d_{rt}/d_{ss} is larger than 1 as well. The biases in the estimators are large, since there are quite a lot of occurrences of $\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}} > 1$, indicating a significant effect of non-linearity. We can take a close look on the biases by plotting the position estimates for Geometry 3 with a range error STD of 1.75 cm in Figure 4.10, where positioning STD along x-axis is 11.28 cm and along y-axis is 12.43 cm.

In such cases where $\mu_{\hat{x}}$ is considered large, one can simply reject the position estimates

with $\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}} > 1$.

4.4.2.2 Bias Check for Systems without Clock Offset

If the system is fully synchronized, the system is clock offset free and the number of unknown parameters is reduced to 3. First, we verify (3.65) with simulation results and then we show the experimental positioning results obtained with the acoustic system.

Verification of (3.65) as we did in Section 4.4.2.1, the verification is done by simulations. Figure 4.11 shows the results obtained with the same system configuration as in Figure 4.7.

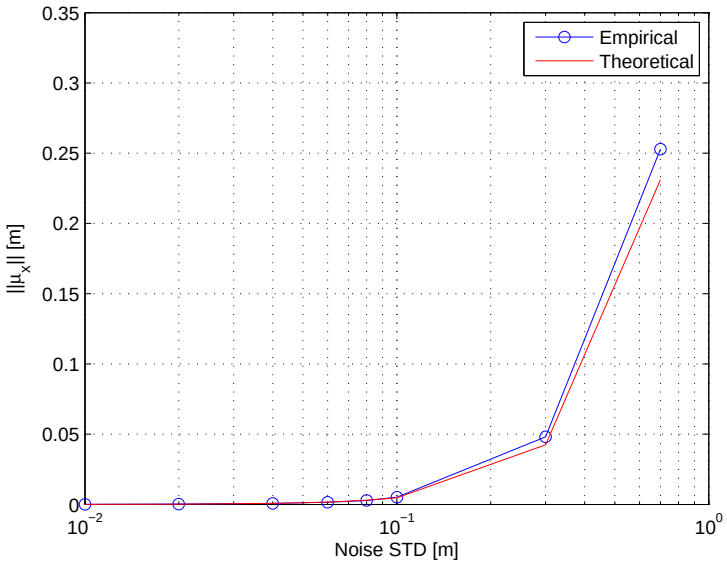

Figure 4.11: Empirical value through simulation and theoretical value (3.65) of non-linearity bias magnitude (case without clock offset).

As we can see, the theoretical curve approximates the empirical data very well with measurement STD below 0.1 m, and the theoretical curve gets slightly smaller than the empirical one at measurement STDs larger than 0.1 m. The reason of this small deviation is the same as the first reason given for the case with a clock offset. Please note that, unlike in Figure 4.7, here the empirical curve in Figure 4.11 does not go below the theoretical one at large measurement STDs, since the convergence problem never happened in our tests for the clock offset free cases.

Meanwhile, we can already get the impression that the significance of the non-linearity bias is reduced when the system is clock offset free. In Figure 4.11, even with noise STD $= 0.1$ m, the amplitude of the bias is negligible. This point can also be seen from the following results.

Positioning Results With the same assumption as we made previously, a 95% confidence region for Geometry 4 is plotted in Figure 4.12. With one less parameter to estimate, the obtained positioning accuracy is better. The ellipse is within a square of 2×2 cm^2.

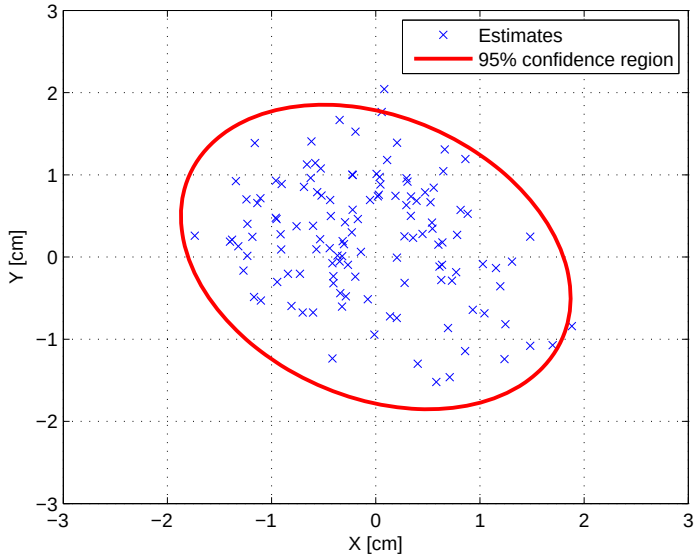

Figure 4.12: Region in the x-y plane with 95% confidence for the estimated positions of Geometry 4 using full bandwidth and the origin represents the true postion.

The results for the bias check, with the same parameters of interest, are summarized in Table 4.3.

The convergence problem never happened, and is not included in Table 4.3.

The system performances are also ranked as Geometry 4 $>$ Geometry 1 $>$ Geometry 2 $>$ Geometry 3, from good to poor; the relations between the performances and the parameters of interest are similar.

The great difference is that the bias due to non-linearity is never considered significant for all the geometries when the system is clock offset free. With a clock offset, $\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}} > 1$ happens a lot with Geometry 3 in Table 4.2 and it never happened with the absence of a clock offset in Table 4.3. This agrees with the impression we obtained from Figure 4.11. The main reason is that clock offset free systems are more robust to ill-conditioned geometric configurations such as Geometry 3 where the majority of the transmitters are on the same side of the receiver in one or more dimensions.

4.4.3 Comparing Different Methods

Here we show the positioning results obtained with the actual UWB radio timing-based measurements under LoS. With a fixed signal bandwidth, a total number of 50000 tests are performed. The performance of the positioning algorithms is evaluated based on the RMSE of the corresponding *position* estimates, which quantifies how far (in an average point of view) the estimates are from the true position. The RMSE of the estimator is obtained empirically as RMSE $= \sqrt{\frac{1}{N-1}\sum_{l=1}^{N}\|\hat{x}_u^l - x_u\|^2}$, with x_u the true position, $N = 50000$, and $\hat{x}_u^l$ the estimates obtained in the l-th test. Note that $\hat{\underline{b}}_c$ is not included in the RMSE in TDoA. No convergence problem has been observed during the comparison test for both

the constrained iterative method and the Gauss-Newton method.

The empirical RMSE curves are plotted in Figure 4.13 for the ToA case and in Figure 4.14 for the TDoA case. On the horizontal axis, the range error STD is given.

In the ToA case, the Gauss-Newton method and the constrained iterative method give best performances, since they closely approximate rigorous least-squares solutions to the corresponding models. The MOF method performs very good with x_s specially chosen as the origin (exactly as x_u), since in this case the MOF solution is the equivalent to the rigorous solution. On the contrary, the MOF method performs worst with x_s chosen as the position of the transmitter which corresponds to the largest range measurement (just to give a bad example). Again, we should note that in practice, it is not possible to chose x_s equal to x_u.

In the TDoA case, the design matrices of the the SOF method is close to singular. The RMSE of the estimate is about hundreds of meters, and is not shown in the figure. Other tested methods behave in a similar way as in the ToA case, only that the differences between various estimators are very small. Moreover, for the MOF method, different choices of constraints affect hardly the final estimates. In the figure, we show the result obtained with default choice $\kappa = x_u^T x_u - b_c^2$.

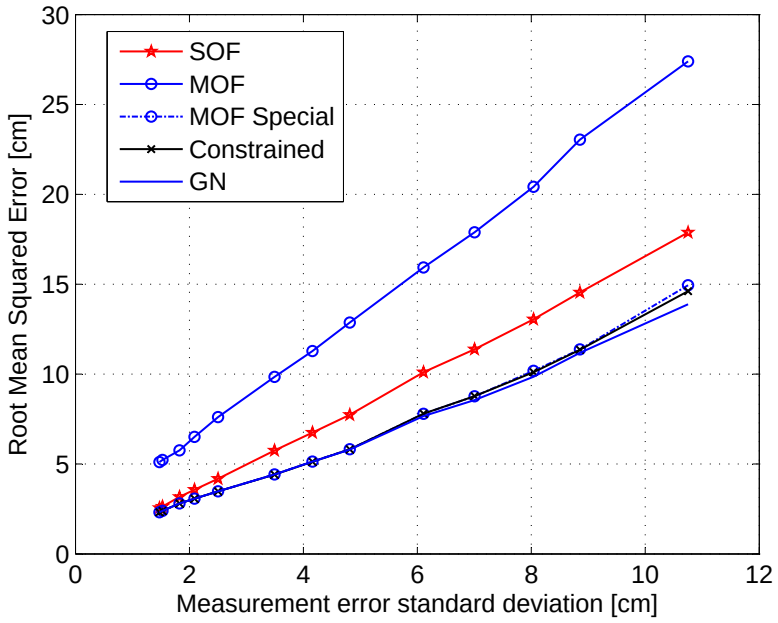

Figure 4.13: RMSE of the estimates obtained using the actual ToA UWB data, for different methods.

4.5 Summary

In this Chapter, range-based positioning algorithms are reviewed, including iterative descent methods, direct methods and constrained iterative methods.

The traditional iterative descent methods are widely used in non-linear estimation problems. They iteratively approximate a rigorous least-squares solution with accuracy achieving the

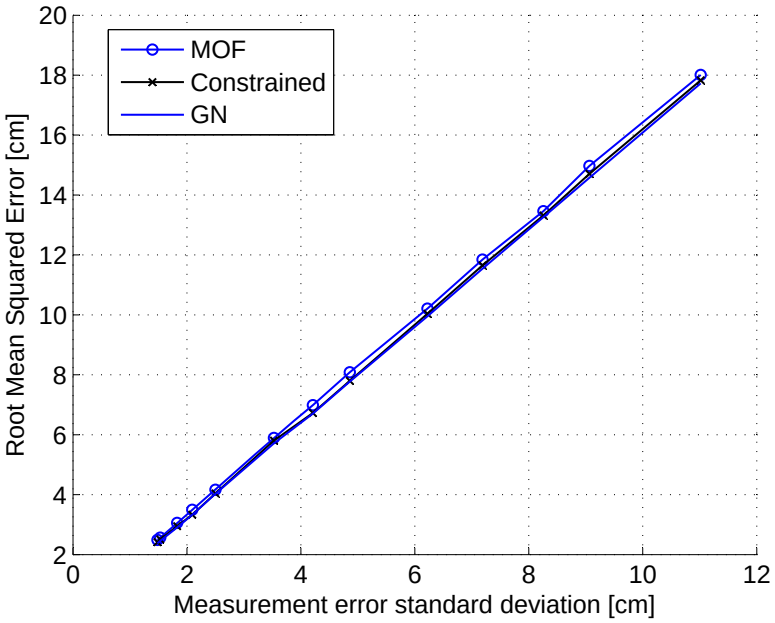

Figure 4.14: RMSE of the estimates obtained using the actual TDoA UWB data, for different methods.

CRLB [62], when successfully converged. The drawbacks of these methods are that 1) the final estimates are obtained iteratively, i.e. no closed-form analytical expressions are available; 2) these methods all require an initial guess to start, which can be quite tricky for some applications, e.g. indoor localizations.

One of the possible ways to provide an initial guess for the iterative descent methods is to use a direct method, which provides a closed-form solution without any iteration. In this Chapter, two major types of the direct methods, namely the SOF and MOF methods are reviewed. It has been shown that the SOF methods are strictly identical with a specific choice of weight matrices. The advantage of the direct methods, as compared to the iterative descent methods, is clearly their simplicity, which however is achieved at the price of a simplified solution with reduced accuracy. Validation work using actual audio LoS timing-based range measurements has been performed to show that the idea of using a direct method to obtain a initial guess for iterative methods works fine, when the system is not configured (close to) coplanar.

There is also a third group of methods, referred in this Chapter as constrained iterative methods, which solve the problem with only one parameter to estimate in each iteration. This type of methods provide better accuracy than the direct solutions by approximating rigorous least-squares solutions to squared range equations. Concerning the original publications [61, 62], a step of EVD is not necessary, which is itself a non-linear problem that needs to be solved iteratively in general.

A full summary of all reviewed methods can be found in Table 4.4, concerning computational load, positioning accuracy and requirements on redundancy. It should be pointed out that for a consistent system, there is no need to resort to iterative techniques and one should always use the MOF methods, since the least-squares residual is zero.

This review intends to help system designers and researchers to have a better understanding of the existing range based positioning algorithms, so that a balance between computational load and positioning accuracy can be found and the choice for a method can be made according to the requirements set by the application at hand. In the mean time, it helps to develop the ideas of improving existing methods, which are discussed in the next Chapter.

Table 4.2: Results of the bias check for all the geometries when the system includes a clock offset.

Positioning Results				
Range Error STD [cm]	1.15	1.75	3.41	11.86
Geometry 1				
Mean of $\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}}$	0.01	0.01	0.02	0.08
$\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}95\%}$	0.02	0.01	0.03	0.11
$\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}} > 1$ (percentage)	0	0	0	0
Convergence Problem (percentage)	0	0	0	0
d_{ss} [m]	3.06			
d_{rt} [m]	1.1			
d_{rt}/d_{ss}	0.36			
Geometry 2				
Mean of $\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}}$	0.02	0.03	0.04	0.17
$\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}95\%}$	0.04	0.04	0.06	0.26
$\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}} > 1$ (percentage)	0	0	0	0
Convergence Problem (percentage)	0	0	0	0
d_{ss} [m]	2.83			
d_{rt} [m]	1.05			
d_{rt}/d_{ss}	0.37			
Geometry 3				
Mean of $\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}}$	0.14	0.17	2.07	6.56
$\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}95\%}$	0.43	0.26	3.12	19.63
$\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}} > 1$ (percentage)	0	8.73	40	94.25
Convergence Problem (percentage)	0	0	0	0.8
d_{ss} [m]	1.47			
d_{rt} [m]	2.33			
d_{rt}/d_{ss}	1.59			
Geometry 4				
Mean of $\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}}$	0.01	0.01	0.05	0.06
$\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}95\%}$	0.01	0.01	0.06	0.08
$\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}} > 1$ (percentage)	0	0	0	0
Convergence Problem (percentage)	0	0	0	0
d_{ss} [m]	2.91			
d_{rt} [m]	1.12			
d_{rt}/d_{ss}	0.38			

Table 4.3: Results of the bias check for all the geometries when the system is clock offset free.

Positioning Results				
Range Error STD [cm]	1.15	1.75	3.41	11.86
Geometry 1				
Mean of $\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}}$	0.00	0.01	0.01	0.03
$\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}95\%}$	0.01	0.01	0.02	0.06
$\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}} > 1$ (percentage)	0	0	0	0
d_{ss} [m]	3.06			
d_{rt} [m]	1.1			
d_{rt}/d_{ss}	0.36			
Geometry 2				
Mean of $\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}}$	0.01	0.01	0.01	0.06
$\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}95\%}$	0.01	0.02	0.02	0.10
$\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}} > 1$ (percentage)	0	0	0	0
d_{ss} [m]	2.83			
d_{rt} [m]	1.05			
d_{rt}/d_{ss}	0.37			
Geometry 3				
Mean of $\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}}$	0.02	0.08	0.20	0.56
$\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}95\%}$	0.03	0.16	0.40	0.93
$\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}} > 1$ (percentage)	0	0	0	0
d_{ss} [m]	1.47			
d_{rt} [m]	2.33			
d_{rt}/d_{ss}	1.59			
Geometry 4				
Mean of $\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}}$	0.00	0.01	0.03	0.03
$\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}95\%}$	0.01	0.01	0.04	0.04
$\|\mu_{\hat{x}}\|_{Q(\hat{x})^{-1}} > 1$ (percentage)	0	0	0	0
d_{ss} [m]	2.91			
d_{rt} [m]	1.12			
d_{rt}/d_{ss}	0.38			

Table 4.4: Summary of the reviewed range based positioning algorithms.

Methods	Complexity	Accuracy	Determined
ID	Iterative with full dimensions	Closely approximates strict solution	$\dim(\underline{y}) = m = \dim(x_u) + \dim(b_c)$
SOF	Non-iterative with unique solution	Simplified solution	$\dim(\underline{y}_D) = \dim(x_u) + \dim(b_c) + \dim(\kappa)$
MOF	Non-iterative with two candidate solutions	Simplified solution	$dim(\underline{y}_D) = \dim(x_u) + \dim(b_c) + \dim(\kappa) - 1$ $+ \dim(b_c) + \dim(\kappa) - 1$
Constrained iterative	Iterative with 1-D parameter	Closely approximates rigorous solution	$dim(\underline{y}_D) = \dim(x_u) + \dim(b_c) + \dim(\kappa) - 1$

5 Improvements

The existing range-based positioning algorithms have been studied in detail in the previous chapter, including the traditional iterative descent methods, the direct methods including the SOF and MOF methods, and the constrained iterative methods. The relation between different methods regarding estimation accuracy and computational load is given in Figure 5.1, which aims at showing a general tendency on the relations between the different methods. The shown relations may not hold in some special situations. For example, the estimation accuracy of the MOF method is dependent on the choice of x_s in (4.55), a bad/good choice can make the method perform worse/better than the SOF methods, see Figure 4.13.

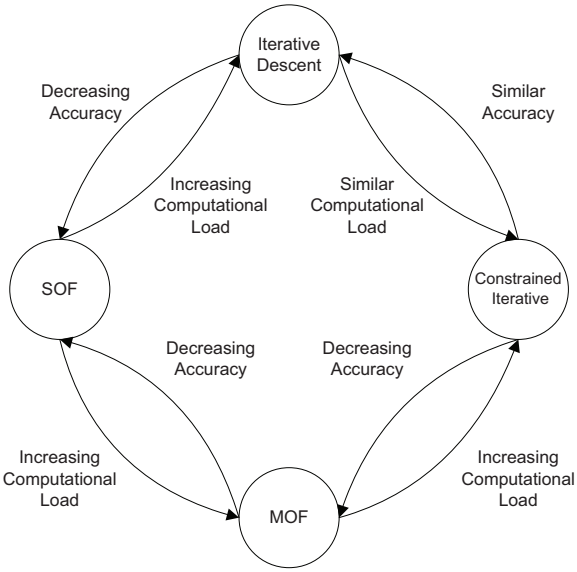

Figure 5.1: A general trend of the relation between the methods analyzed in Chapter 4.

The general trend in Figure 5.1 suggests that high positioning accuracy generally requires heavier computational load, while simple methods that require less computing power sacrifice on positioning accuracy. The objective of this chapter is to introduce some possible improvements based on the previous review, with the goal of:

1. Increasing the position estimation accuracy, but keeping the computational load as low as possible,
2. Reducing the computational load, but keeping the positioning accuracy as high as possible.

5.1 Direct: Multiple Objective Function Methods: Type II

In this section, a new method is proposed to improve the SOF methods by exploiting multiple, but similar, fully constrained models. The solution is better, in the sense that it corresponds to an equal or smaller value of the original least-squares objective function (3.50). Moreover, the method remains direct and the solution is given in a closed form. This method belongs to the group of MOF methods, but it exploits the constraint $f(x_D) = 0$ in a different way from any of the reviewed MOF methods. In the rest of the thesis, this method will be referred as Multiple Objective Function Method Type II (MOF II).

5.1.1 Mechanics

5.1.1.1 TDoA

Recall from (3.17), the squared measurement equations can be expressed as:

$$\underline{y}_i^2 = x_u^T x_u - b_c^2 - 2x_i^T x_u + 2\underline{y}_i b_c + x_i^T x_i + 2d_i\underline{e}_i + \underline{e}_i^2,\ i = 1 \dots m. \tag{5.1}$$

Multiplying both sides of (5.1) with the transformation matrix in (4.37):

$$T_A = \begin{bmatrix} 1 & & 0 & -1 \\ & \ddots & & \vdots \\ 0 & & 1 & -1 \\ 0 & \cdots & 0 & 1 \end{bmatrix}_{m\times m} \tag{5.2}$$

generates the first model used by the MOF II method:

$$\begin{aligned} y_A = \begin{bmatrix} \underline{y}_D \\ \underline{y}_r^2 \end{bmatrix} &= \begin{bmatrix} \underline{F}_D \\ 0 \end{bmatrix} x_D + C_A y_r^2 + \begin{bmatrix} \underline{e}_D \\ 2d_r\underline{e}_r + \underline{e}_r^2 \end{bmatrix} \\ \text{constraint:} \qquad & \|x_u - x_r\|^2 = (y_r - b_c)^2, \end{aligned} \tag{5.3}$$

where $[\underline{F}_D]_{(m-1)\times(n+1)} = [A_D\ \ \underline{B}_D]$, $x_D = [x_u^T,\ b_c]^T$ and $C_A = [0, ..., 0,\ 1]_{m\times 1}^T$. Note that the first $m-1$ rows of the above model are nothing else but the squared-range difference equations, meaning that the terms $(.)_D$ are characterized by $\underline{y}_{D,i} = \underline{y}_i^2 - \underline{y}_r^2 - x_i^T x_i + x_r^T x_r$, $A_{D,i} = -2(x_i - x_r)^T$, $\underline{B}_{D,i} = 2(\underline{y}_i - \underline{y}_r)$, $x_D = [x_u^T,\ b_c]^T$, and $\underline{e}_{D,i} = 2d_i\underline{e}_i + \underline{e}_i^2 - 2d_r\underline{e}_r - \underline{e}_r^2$.

The rigorous solution $(\hat{\underline{x}}_D, \hat{\underline{y}}_r d)$ to the model (5.3) can be given as:

$$\begin{aligned} [\hat{\underline{x}}_D^T, \hat{\underline{y}}_r]^T &= \arg\min_{x_D, y_r} \left\| \underline{y}_A - \begin{bmatrix} \underline{F}_D \\ 0 \end{bmatrix} x_D - C_A y_r^2 \right\|_{W_A}^2 \\ \text{subject to:} \qquad & \|x_D - \breve{x}_r\|_L^2 - (y_r - s^T x_D)^2 = 0, \end{aligned} \tag{5.4}$$

which is obtained by finding a stationary point of the following Lagrange function:

$$F_A(x_D, y_r^2, \lambda) = \left\| \underline{y}_A - \begin{bmatrix} \underline{F}_D \\ 0 \end{bmatrix} x_D - C_A y_r^2 \right\|_{W_A}^2 + \lambda\left(\|x_D - \breve{x}_r\|_L^2 - (y_r - s^T x_D)^2\right)$$

(5.5)

where W_A is a weight matrix, $\breve{x}_r = [x_r^T,\ 0]^T_{(n+1)\times 1}$ and $s = [0\ \cdots\ 0\ 0\ 1]^T_{(n+1)\times 1}$ and

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}_{(n+1)\times(n+1)} . \tag{5.6}$$

The above function can be simplified by assuming that the weight matrix W_A pertains a special form as

$$W_A = \begin{bmatrix} W_D & 0 \\ 0 & w \end{bmatrix}. \tag{5.7}$$

This assumption will be further discussed in Section 5.1.3.

Based on (5.7), it holds that:

$$\left\| \underline{y}_A - \begin{bmatrix} \underline{F}_D \\ 0 \end{bmatrix} x_D - C_A y_r^2 \right\|^2_{W_A} = \|\underline{y}_D - \underline{F}_D x_D\|^2_{W_D} + \|\underline{y}_r^2 - y_r^2\|^2_w \tag{5.8}$$

and (5.5) can be rewritten as

$$F_A(x_D, y_r^2, \lambda) = \|\underline{y}_D - \underline{F}_D x_D\|^2_{W_D} + \|\underline{y}_r^2 - y_r^2\|^2_w + \lambda \left(\|x_D - \breve{x}_r\|^2_L - (y_r - s^T x_D)^2 \right) \tag{5.9}$$

Taking first order derivatives of (5.9) w.r.t. x_D and setting the results to zero, one obtains

$$\hat{\underline{x}}_D = \left[\underline{F}_D^T W_D \underline{F}_D + \lambda(L - ss^T)\right]^{-1} \left[\underline{F}_D^T W_D \underline{y}_D + \lambda(L\breve{x}_r + s y_r)\right] \tag{5.10}$$

Now the rest of the task is to derive the closed-form expressions for λ and y_r.

The second model used here is:

$$\underline{y}_A = \begin{bmatrix} \underline{F}_D \\ 0 \end{bmatrix} x_D + C_A y_r^2 + \begin{bmatrix} \underline{e}_D \\ 2d_r \underline{e}_r + \underline{e}_r^2 \end{bmatrix} \text{ constraint: } (\|x_u - x_r\| + b_c)^2 = y_r^2, \tag{5.11}$$

which is very similar to (5.3), and the only difference is in the constraint.

The corresponding solution $(\hat{\underline{x}}'_D, \hat{\underline{y}}'_r)$ can be calculated as:

$$\begin{aligned} [\hat{\underline{x}}'_D{}^T, \hat{\underline{y}}'_r]^T &= \arg\min_{x_D, y_r} \|\underline{y}_D - \underline{F}_D x_D\|^2_{W_D} + \|\underline{y}_r^2 - y_r^2\|^2_w \\ \text{subject to:} &\quad (\|x_D - \breve{x}_r\|_L + s^T x_D)^2 - y_r^2 = 0, \end{aligned} \tag{5.12}$$

which is then obtained by finding a stationary point of the following Lagrange function:

$$F'_A(x_D, y_r^2, \lambda) = \|\underline{y}_D - \underline{F}_D x_D\|^2_{W_D} + \|\underline{y}_r^2 - y_r^2\|^2_w + \lambda \left((\|x_D - \breve{x}_r\|_L + s^T x_D)^2 - y_r^2 \right) \tag{5.13}$$

The third model used here is generated by replacing $\underline{y}_r^2$ and the unknown y_r^2 with $\underline{y}_r$ and y_r:

$$y_A'' = \begin{bmatrix} \underline{y}_D \\ \underline{y}_r \end{bmatrix} = \begin{bmatrix} \underline{F}_D \\ 0 \end{bmatrix} x_D + C_A y_r + \begin{bmatrix} \underline{e}_D \\ \underline{e}_r \end{bmatrix} \text{ constraint: } (\|x_u - x_r\| + b_c)^2 = y_r^2. \quad (5.14)$$

and a local minimum of the third objective function:

$$\begin{aligned} &\arg_{x_D, y_r} \left\| \underline{y}_A'' - \begin{bmatrix} \underline{F}_D \\ 0 \end{bmatrix} x_D - C_A y_r \right\|_{W_A''}^2 \\ &\text{subject to: } (\|x_D - \breve{x}_r\|_L + s^T x_D)^2 - y_r^2 = 0 \end{aligned} \quad (5.15)$$

corresponds to the rigorous solution $(\hat{\underline{x}}_D'', \hat{\underline{y}}_r'')$ of the third model, which can be obtained by searching a stationary point of the following Lagrange function:

$$F_A''(x_D, y_r, \lambda) = \left\| \underline{y}_A'' - \begin{bmatrix} \underline{F}_D \\ 0 \end{bmatrix} x_D - C_A y_r \right\|_{W_A''}^2 + \lambda \left((\|x_D - \breve{x}_r\|_L + s^T x_D)^2 - y_r^2 \right) \quad (5.16)$$

Assuming a similar structure of W_A'' as in (5.7):

$$W_A'' = \begin{bmatrix} W_D & 0 \\ 0 & w'' \end{bmatrix}, \quad (5.17)$$

Equation (5.16) can be rewritten as:

$$F_A''(x_D, y_r, \lambda) = \|\underline{y}_D - \underline{F}_D x_D\|_{W_D}^2 + \|\underline{y}_r - y_r\|_{w''}^2 + \lambda \left((\|x_D - \breve{x}_r\|_L + s^T x_D)^2 - y_r^2 \right) \quad (5.18)$$

As mentioned in Chapter 4, the direct solution to the non-linear positioning problem is usually achieved by introducing certain simplifications. Here, based on the resemblances among the three objective functions (5.9), (5.13) and (5.18), it is reasonable to assume that $\hat{\underline{x}}_D = \hat{\underline{x}}_D' = \hat{\underline{x}}_D''$, $\hat{\underline{y}}_r = \hat{\underline{y}}_r' = \hat{\underline{y}}_r''$ and $\hat{\underline{\lambda}} = \hat{\underline{\lambda}}' = \hat{\underline{\lambda}}''$.

Taking first order derivatives of (5.13) and (5.18) w.r.t. y_r^2 and y_r, respectively, and setting the result to zero, yields:

$$-2w(\underline{y}_r^2 - \hat{\underline{y}}_r^2) - \lambda = 0 \quad (5.19)$$

$$-w''(\underline{y}_r - \hat{\underline{y}}_r) - \lambda \hat{\underline{y}}_r = 0 \quad (5.20)$$

which is a set of two equations with two unknowns. Solving (5.20) gives us three pairs of $\hat{\underline{\lambda}}$ and $\hat{\underline{y}}_r$,

$$\hat{\underline{\lambda}}^\alpha = 0, \ \hat{\underline{y}}_r^\alpha = \underline{y}_r, \quad (5.21)$$

$$\hat{\underline{\lambda}}^\beta = w'' - w\underline{y}_r^2 - w\underline{y}_r \sqrt{\underline{y}_r^2 + \frac{2w''}{w}}, \ \hat{\underline{y}}_r^\beta = \frac{-\underline{y}_r + \sqrt{\underline{y}_r^2 + \frac{2w''}{w}}}{2} \quad (5.22)$$

$$\hat{\underline{\lambda}}^\gamma = w'' - w\underline{y}_r^2 + w\underline{y}_r \sqrt{\underline{y}_r^2 + \frac{2w''}{w}}, \ \hat{\underline{y}}_r^\gamma = \frac{-\underline{y}_r - \sqrt{\underline{y}_r^2 + \frac{2w''}{w}}}{2} \quad (5.23)$$

Based on (5.10), the results (5.21)-(5.23) can be used to calculate the three candidates for x_D, and the final solution should be chosen as the one

$$\hat{\underline{x}}_D^{\text{MOFII}} = \arg \min_{\hat{\underline{x}}_D^{\alpha}, \hat{\underline{x}}_D^{\beta}, \hat{\underline{x}}_D^{\gamma}} \|\underline{y} - A(\hat{\underline{x}}_D^k)\|_W^2, \; k = \alpha, \beta, \gamma, \tag{5.24}$$

that minimizes the least-squares objective function for the original model (3.2).

5.1.1.2 ToA/RSS

In case of clock offset free systems, only two models are involved in the MOF II method. The first model can be written as:

$$\underline{y}_A = \begin{bmatrix} \underline{y}_D \\ \underline{y}_r^2 \end{bmatrix} = \begin{bmatrix} F_D \\ 0 \end{bmatrix} x_D + C_A y_r^2 + \begin{bmatrix} \underline{e}_D \\ 2d_r \underline{e}_r + \underline{e}_r^2 \end{bmatrix} \text{ constraint: } \|x_D - x_r\|^2 = y_r^2 \tag{5.25}$$

Note that here $x_D = x_u$ and $F_D = A_D$. The choices for $\underline{y}_D$, A_D, and $\underline{e}_D$ are the same as for (5.3).

Based on the model above, a Lagrange function can be constructed as:

$$F_A(x_D, y_r^2, \lambda) = \left\| \underline{y}_A - \begin{bmatrix} F_D \\ 0 \end{bmatrix} x_D - C_A y_r \right\|_{W_A}^2 + \lambda \left(\|x_D - x_r\|^2 - y_r^2 \right). \tag{5.26}$$

and the solution $(\hat{\underline{x}}_D, \hat{\underline{y}}_r)$ to (5.25) can be found by searching a stationary point of the above function.

Again, assuming a special structure of W_A as in (5.7), the function (5.26) can be rewritten as:

$$F_A(x_D, y_r^2, \lambda) = \|\underline{y}_D - F_D x_D\|_{W_D}^2 + \|\underline{y}_r^2 - y_r^2\|_w^2 + \lambda \left(\|x_D - x_r\|^2 - y_r^2 \right). \tag{5.27}$$

Taking first order derivatives of (5.27), w.r.t. x_D and y_r^2, one obtains:

$$\hat{\underline{x}}_D = (F_D^T W_D F_D + \lambda I)^{-1} (F_D^T W_D \underline{y}_D + \lambda x_r) \tag{5.28}$$

$$0 = -2w(\underline{y}_r^2 - \hat{\underline{y}}_r^2) - \lambda. \tag{5.29}$$

If λ can be calculated non-iteratively, a closed-form solution of $\hat{\underline{x}}_D$ can be obtained based on (5.28). To do so, one more equation, in addition to (5.29), is required to calculate y_r and λ.

A model similar as (5.25) can be obtained, if $\underline{y}_r^2$ and y_r^2 are used with their non-squared form:

$$\underline{y}'_A = \begin{bmatrix} \underline{y}_D \\ \underline{y}_r \end{bmatrix} = \begin{bmatrix} F_D \\ 0 \end{bmatrix} x_D + C_A y_r + \begin{bmatrix} \underline{e}_D \\ \underline{e}_r \end{bmatrix} \text{ constraint: } \|x_D - x_r\|^2 = y_r^2 \tag{5.30}$$

The above positioning problem (5.30), can be solved by finding a local minimum of:

$$\begin{aligned} &\left\| \underline{y}'_A - \begin{bmatrix} F_D \\ 0 \end{bmatrix} x_D - C_A y_r \right\|^2_{W'_A} \\ &\text{subject to: } \|x_u - x_r\|^2 - y_r^2 = 0. \end{aligned} \tag{5.31}$$

which is then obtained by finding a stationary point of the following Lagrange function:

$$F'_A(x_D, y_r, \lambda) = \left\| \underline{y}'_A - \begin{bmatrix} F_D \\ 0 \end{bmatrix} x_D - C_A y_r \right\|^2_{W'_A} + \lambda \left(\|x_u - x_r\|^2 - y_r^2 \right). \tag{5.32}$$

The above function can be rewritten as:

$$F'_A(x_D, y_r, \lambda) = \|\underline{y}_D - F_D x_D\|^2_{W_D} + \|\underline{y}_r - y_r\|^2_{w'} + \lambda \left(\|x_u - x_r\|^2 - y_r^2 \right), \tag{5.33}$$

by assuming a structure of W'_A as in (5.7).

Taking first order derivatives of (5.33) w.r.t. x_D and y_r, and setting them to zero, yields

$$\begin{aligned} \underline{\hat{x}}'_D &= (F_D^T W_D F_D + \lambda I)^{-1} (F_D^T W_D \underline{y}_D + \lambda x_r) && (5.34) \\ 0 &= -2w'(\underline{y}_r - \underline{\hat{y}}_r) - 2\lambda \underline{\hat{y}}_r. && (5.35) \end{aligned}$$

Based on the similarities of the two models (5.25) and (5.30), it is reasonable to assume that $\underline{\hat{y}}_r = \underline{\hat{y}}'_r$ and $\underline{\hat{\lambda}} = \underline{\hat{\lambda}}'$, and therefore $\underline{\hat{x}}_D = \underline{\hat{x}}'_D$.

Combining (5.29) and (5.35), one gets

$$\underline{\hat{\lambda}} \left[\frac{1}{2w} \left(\frac{w' - \underline{\hat{\lambda}}}{w'} \right)^2 - \underline{y}_r^2 \frac{1}{w'} \left(\frac{2w' - \underline{\hat{\lambda}}}{w'} \right) \right] = 0 \tag{5.36}$$

which directly gives us three candidates of λ,

$$\begin{aligned} \underline{\hat{\lambda}}^{\alpha} &= 0 && (5.37) \\ \underline{\hat{\lambda}}^{\beta} &= w' - w\underline{y}_r^2 - w\underline{y}_r \sqrt{\underline{y}_r^2 + \frac{2w'}{w}} && (5.38) \\ \underline{\hat{\lambda}}^{\gamma} &= w' - w\underline{y}_r^2 + w\underline{y}_r \sqrt{\underline{y}_r^2 + \frac{2w'}{w}} && (5.39) \end{aligned}$$

Based on either (5.28) or (5.34), the results (5.37)-(5.39) can be used to calculate the three candidates for x_D, and the final solution should be chosen as the one

$$\underline{\hat{x}}_D^{\text{MOFII}} = \arg \min_{\underline{\hat{x}}_D^{\alpha}, \underline{\hat{x}}_D^{\beta}, \underline{\hat{x}}_D^{\gamma}} \|\underline{y} - A(\underline{\hat{x}}_D^k)\|^2_W, \ k = \alpha, \beta, \gamma, \tag{5.40}$$

similar as what we did for TDoA cases.

5.1.2 Error Analysis

Recall from (5.10), the expression for $\hat{\underline{x}}_D$ is:

$$\hat{\underline{x}}_D(\underline{e}) = \left[\underline{F}_D^T W_D \underline{F}_D + \hat{\underline{\lambda}}(L - ss^T)\right]^{-1} \left[\underline{F}_D^T W_D \underline{y}_D + \hat{\underline{\lambda}}(L\breve{x}_r + s\hat{\underline{y}}_r)\right] \tag{5.41}$$

and the expressions for three possible values of $\hat{\underline{\lambda}}$ and $\hat{y}_r$ can be found in (5.21)-(5.23).

5.1.2.1 TDoA

The error analysis can be performed in a similar way as we did for MOF methods in Section 4.2.5.2. Recall from (4.85)-(4.86), the bias in the estimator, defined as: $\mu_{\hat{x}_D} = \mathsf{E}\{\hat{\underline{x}}_D(\underline{e}) - x_D(0)\}$, and the variance matrix, defined as: $Q_{\hat{x}_D\hat{x}_D} = \mathsf{E}\left\{(\hat{\underline{x}}_D - \mathsf{E}\{\hat{\underline{x}}_D\})(\hat{\underline{x}}_D - \mathsf{E}\{\hat{\underline{x}}_D\})^T\right\}$, can be approximated as:

$$\mu_{\hat{x}_D} = \frac{1}{2}\mathsf{E}\left\{\underline{e}^i \partial^2_{ij} x_D(0) \underline{e}^j\right\} = \frac{1}{2}\partial^2_{i^2} x_D(0) \mathsf{E}\{(\underline{e}^i)^2\} = \frac{1}{2}\partial^2_{i^2} x_D(0)\sigma_i^2, \tag{5.42}$$

$$Q_{\hat{x}_D\hat{x}_D} = \mathsf{E}\left\{\partial_i x_D(0) \underline{e}^i \underline{e}^j \partial_j x_D^T(0)\right\} = \partial_i x_D(0)\sigma_i^2 \partial_i x_D^T(0). \tag{5.43}$$

where the use is made of Einstein's summation convention, with $\partial_i(.)$ denoting the partial derivative of $(.)$ w.r.t. the i-th element of $\underline{e}$, denoted as $\underline{e}^i$. Moreover, the fact that $\mathsf{E}\{\underline{e}^i \underline{e}^j\} = 0,$ for $i \neq j$, has been used.

The rest of the analysis is to derive the expressions for $\partial_i x_D(0)$ and $\partial^2_{i^2} x_D(0)$, which is done in Appendix A.2.3.

5.1.2.2 ToA/RSS

The error analysis for ToA/RSS cases can be derived in a similar way as what has been done for TDoA cases. In fact, the results derived in (5.42)-(A.10) can be directly used by choosing $L = I$, $s = 0$ and using proper dimensions.

5.1.3 Discussion

5.1.3.1 Improvement

The MOF II estimator is better than a SOF estimator in the sense that the former corresponds to an equal or smaller value of the original least-squares objective function (3.50).

Proof. Plugging $\hat{\underline{\lambda}}^\alpha = 0$ in (5.10) and (5.28), one obtains

$$\hat{\underline{x}}_u^\alpha = (\underline{F}_D^T W_D \underline{F}_D)^{-1} \underline{F}_D^T W_D \underline{y}_D, \tag{5.44}$$

which is identical to the SOF estimator (4.26). Since the final MOF II estimator $\hat{\underline{x}}_D^{\text{MOFII}}$ in both (5.24) and (5.40) is chosen as the one that minimizes the original least-squares objective function among all available candidates, one obtains

$$\|\underline{y} - A(\hat{\underline{x}}_D^{\text{MOFII}})\|_W^2 \leq \|\underline{y} - A(\hat{\underline{x}}_D^\alpha)\|_W^2 = \|\underline{y} - A(\hat{\underline{x}}_D^{SOF})\|_W^2. \tag{5.45}$$

Therefore, $\hat{\underline{x}}_D^{\text{MOFII}}$ is better than $\hat{\underline{x}}_D^{SOF}$.

5.1.3.2 Assumption on Weight Matrix

Now we look back on the assumption we made for the weight matrix. Here we address only TDoA cases, and the discussion for ToA/RSS cases can be readily performed in a similar way.

Concerning the model (5.3), the weight matrix W_A can be chosen as an identity matrix when the statistics of $\underline{y}_A$ is unknown, and (5.7) is fulfilled. Otherwise, W_A is usually chosen as $W_A = Q_{\underline{y}_A}^{-1}$. Since $\underline{y}_r^2$ is included in each equation of $\underline{y}_{D,i} = \underline{y}_i^2 - \underline{y}_r^2 - x_i^T x_i + x_r^T x_r$, $\underline{y}_r^2$ and $\underline{y}_D$ are of course correlated:

$$\mathsf{C}\{\underline{y}_{D,i}, \underline{y}_r^2\} = \mathsf{C}\{\underline{y}_r^2\}, \tag{5.46}$$

where we assume that the measurements are independent, and therefore $\underline{y}_i^2$ and $\underline{y}_r^2$ are uncorrelated.

The correlation, however, can become very small with a specific choice of the reference to form $\underline{y}_D$, e.g. choosing the average $\frac{1}{m}\sum_{i=1}^{m}\underline{y}_i^2$ instead of $\underline{y}_r^2$, and the correlation becomes:

$$\mathsf{C}\{\underline{y}_{D,i}, \underline{y}_r^2\} = \frac{1}{m} D(\underline{y}_r^2), \tag{5.47}$$

which is negligible for a large m. No big changes are required to implement this choice of the reference, one only needs to alter the transform matrix T_A in (5.2) to:

$$T_A = \begin{bmatrix} 1-\frac{1}{m} & -\frac{1}{m} & \cdots & -\frac{1}{m} \\ -\frac{1}{m} & \ddots & -\frac{1}{m} & \vdots \\ -\frac{1}{m} & -\frac{1}{m} & 1-\frac{1}{m} & -\frac{1}{m} \\ 0 & \cdots & 0 & 1 \end{bmatrix}_{m\times m} . \tag{5.48}$$

and follow the described mechanics of the MOF II method to achieve a direct solution.

For the model (5.14), with $\frac{1}{m}\sum_{i=1}^{m}\underline{y}_i^2$ chosen as the reference when generating the squared-range differences, the correlation between $\underline{y}_D$ and $\underline{y}_r$ is also scaled by a factor of $1/m$, which can also be neglected with a large m.

5.1.4 Validation and Results

In this section, the MOF II method is validated using actual UWB range measurements. But instead of using timing based measurements, we use RSS measurements. The error statistics of an RSS range measurement vector can be found in [60]. The geometry of the positioning system is the same as described in Section 4.4.

The tested positioning algorithms include the SOF method, the MOF method in [43], and the MOF II method. The MOF method is modified, i.e. $\kappa = (x_u - x_s)^T(x_u - x_s)$, with x_s chosen as the geometrical center of all the base stations. Please note the difference as compared to Figure 4.13, where x_s was chosen as the position of the base station which

corresponds to the largest range measurement. All methods are tested in both un-weighted and weighted cases. In the un-weighted cases, all the methods are weighted with identity matrices. In the weighted case, the matrices are chosen as (5.7).

The RMSE curves of the estimates obtained using different methods, un-weighted and weighted, are plotted in Figure 5.2 and Figure 5.3, respectively. On the horizontal axis, the range error STD is given.

In both un-weighted and weighted cases, the RMSE of the proposed estimator is smallest, which means that better accuracy can be achieved with the proposed method. In the un-weighted case, the improvement is much less than in the weighted case. However, even when the statistics of the measurements are not known perfectly, better weight matrices can be used instead of the identity matrices, e.g. based on the fact that large distance measurements contains larger uncertainties [36], weight $\underline{y}_i$ with $1/\underline{y}_i^2$, $\underline{y}_i^2$ with $1/\underline{y}_i^4$. In this way, similar results can be obtained as in Figure 5.3 (not shown here).

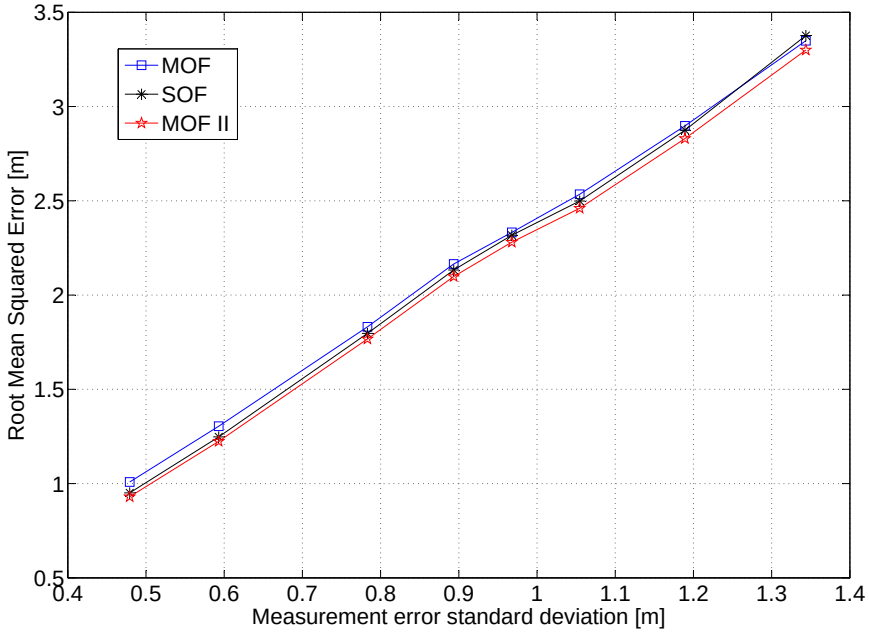

Figure 5.2: RMSE of the estimates for the un-weighted case, using different methods.

5.2 Iterative: A Transformed Least-squares Framework

In this section, a new framework is introduced for least-squares positioning. The multi-dimensional non-linear localization problem is first transformed to a lower dimension and then solved iteratively. Within the proposed Transformed Least-Squares (TLS) framework, we introduce a method in which the localization problem is transformed to one dimension (1-D). In this way, compared to the classical non-linear least-squares (NLS) type of methods, the amount of computations in each iteration is greatly reduced; a reduction of 67% for a 3-D positioning system is shown. Hence, the 1-D iterative (1DI) method is fairly light on the computational load. The way to choose the 1-D parameter is proposed,

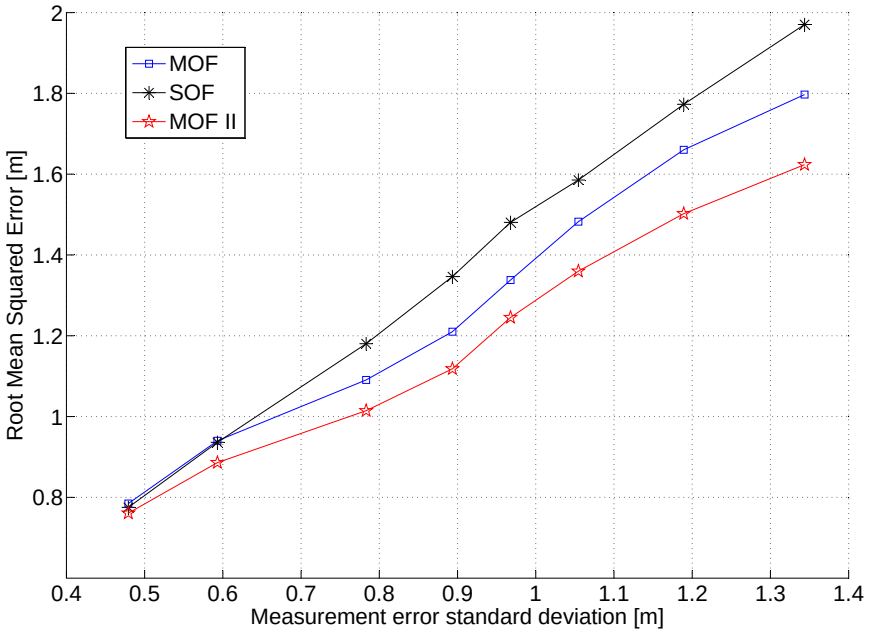

Figure 5.3: RMSE of the estimates for the weighted case, using different methods.

and theoretical expressions for the convergence rate and the RMSE of the 1DI estimator are derived. Validation is performed mainly based on actual UWB radio measurements, and some supplementary simulations are also included. Results show that, in terms of RMSE, the 1DI method performs better than the direct SOF methods, where the solution is obtained non-iteratively, and performs similarly as NLS methods, especially in TDoA cases.

5.2.1 Basic Idea

The idea consists of two steps. In Step 1, the key point is to reformulate or re-parametrize the positioning problem to reduce the number of unknown parameters. It is assumed that one can transform the unknown vector x (see Tab. 3.1) to another unknown vector z of smaller dimension, i.e. $x = L(z)$. We will come back to the point of how to obtain $x = L(z)$, shortly. A new model is then formed based on (3.2),

$$\underline{y} = A(x) + \underline{e} = A(L(z)) + \underline{e}. \tag{5.49}$$

In Step 2, the least-squares solution to (5.49) is obtained as

$$\hat{\underline{x}} = L(\hat{\underline{z}}) = L\left(\arg\min_{z} \left\|\underline{y} - A(L(z))\right\|^2_{Q_{yy}^{-1}}\right), \tag{5.50}$$

which can be done using the ID methods.

If the mapping $x = L(z)$ is error-free, the estimator (5.50) is better than (3.50), in the sense of accuracy and computational complexity, since the objective function in (5.50) contains fewer unknown parameters than (3.50) due to the *additional information* $x = L(z)$. Note that the computational complexity is reduced most when z is a scalar (1-D vector).

Now the unsolved issue is to find the mapping $L(z)$ in Step 1. In practice, an error-free mapping is usually not available and one has to estimate it using the data at hand. Hence, the mapping in Step 1 becomes $L(z, \underline{y})$. Since $\underline{y}$ contains measurement error $\underline{e}$ in (3.2), the mapping can also be written as $L(z, \underline{e})$, and (5.50) becomes

$$\underline{\hat{x}} = L(\underline{\hat{z}}, \underline{e}) = L\left(\arg\min_{z} \left\|\underline{y} - A(L(z, \underline{e}))\right\|^2_{Q_{yy}^{-1}}, \underline{e}\right). \tag{5.51}$$

The estimation of $L(z, \underline{e})$ depends on the formulated non-linear problem to solve. Fortunately, for positioning problems, $L(z, \underline{e})$ can be obtained e.g. based on direct methods. This point will be clearer with the aid of an example in the next section, where, as a realization of the TLS framework, the 1DI method is introduced. In the 1DI method, z is a scalar, $L(z, \underline{e})$ is estimated based on the LLS method, and the objective function in (5.51) is minimized using the Gauss-Newton method. We should note that the TLS framework provides a general way to solve the positioning problem, and the ways to obtain $L(z, \underline{e})$ and the final solution are not limited to the LLS method and the Gauss-Newton method, respectively.

5.2.2 An Example: 1DI Method

Actually there exists an unlimited number of choices for z in $L(z, \underline{e})$, and how to choose the parameter is discussed in Section 5.2.3. In the following derivations, it is assumed that for TDoA, the chosen parameter is $z = b_c$ while for ToA/RSS, the chosen parameter is one of the unknown coordinates, e.g. $z = x_k$, the k-th element of x.

5.2.2.1 TDoA

Recall from (3.26), the squared-range difference equations can be written as:

$$\underline{y}_i^2 - \underline{y}_r^2 = -2(x_i - x_r)^T x_u + 2(\underline{y}_i - \underline{y}_r)b_c + x_i^T x_i - x_r^T x_r + 2d_i\underline{e}_i - 2d_r\underline{e}_r + \underline{e}_i^2 - \underline{e}_r^2, \tag{5.52}$$

with $i = 1 \ldots m-1$, and the matrix form:

$$\underline{y}_D = A_D x_u + \underline{B}_D b_c + \underline{e}_D \tag{5.53}$$

where $\underline{y}_{D,i} = \underline{y}_i^2 - \underline{y}_r^2 - x_i^T x_i + x_r^T x_r$, $A_{D,i} = -2(x_i - x_r)^T$, $\underline{B}_{D,i} = 2(\underline{y}_i - \underline{y}_r)$, $C_{D,i} = 0$, $\underline{e}_{D,i} = 2d_i\underline{e}_i + \underline{e}_i^2 - 2d_r\underline{e}_r - \underline{e}_r^2$.

Based on (5.53), it follows

$$\underline{\hat{x}}_u(b_c) = \arg\min_{x_u} \|\underline{y}_D - A_D x_u - \underline{B}_D b_c\|^2_{W_D} = E\underline{y}_D - E\underline{B}_D b_c, \tag{5.54}$$

where $E = (A_D^T W_D A_D)^{-1} A_D^T W_D$. The key term $L(b_c, \underline{e})$ in Step 1 can then be written as

$$L(b_c, \underline{e}) = [\underline{\hat{x}}_u(b_c)^T, b_c]^T. \tag{5.55}$$

Further denote

$$P_i(b_c, \underline{e}) = A_i(L(b_c, \underline{e})) = \|E\underline{y}_D - E\underline{B}_D b_c - x_i\| + b_c, \;\; i = 1, ..., m, \tag{5.56}$$

then, based on (3.3), a new model can be given as

$$\underline{y} = P(b_c, \underline{e}) + \underline{e}_P, \tag{5.57}$$

where $\underline{e}_P = A(x) - P(b_c, \underline{e}) + \underline{e}$. The effect of $\underline{e}_P$ on the final estimator is addressed in Section 5.2.4.

In Step 2, the parameter b_c can be estimated based on (5.51) as

$$\hat{\underline{b}}_c = \arg\min_{b_c} \left\| \underline{y} - P(b_c, \underline{e}) \right\|_W^2, \tag{5.58}$$

where the weight matrix is taken as $W = Q_{yy}^{-1}$, and in this way, when the uncertainties in $P(b_c, \underline{e})$ are small, (5.58) is a good approximation to (3.50).

Equation (5.58) is a non-linear problem, which can be solved by means of ID algorithms, e.g. the Steepest Descent method, the Newton method, the Gauss-Newton method and the Trust Region method. Concerning estimation quality, these methods provide equivalent solutions when they successfully converge. In the 1DI method, the Gauss-Newton method is chosen to solve (5.58), since it is widely used in positioning systems, e.g. in GPS [1]. Furthermore, the Gauss-Newton method does not have the zigzag problem of the Steepest Descent method, and it is of lower complexity without requiring the evaluation of second order derivatives of $A(x)$, as compared to the Newton method.

First, (5.57) is linearized as

$$\underline{y} \approx P(b_c^0, \underline{e}) + \partial_{b_c} P(b_c^0, \underline{e})(b_c - b_c^0) + \underline{e}_P, \tag{5.59}$$

where b_c^0 is the initial guess for b_c, which can be obtained using a direct method.

Based on (5.59), the parameter b_c can be estimated iteratively and in the $(j+1)$-th iteration, the estimate is updated based on the approximation of (5.58)

$$\hat{\underline{b}}_c^{j+1} = \arg\min_{b_c} \left\| \underline{y} - P(b_c^j, \underline{e}) - \partial_{b_c} P(b_c^j, \underline{e})(b_c - b_c^j) \right\|_W^2. \tag{5.60}$$

The mechanics can be summarized as follows,

Step 1 Estimate $L(b_c, \underline{e})$,

1a. choose b_c as the parameter to refine,

1b. calculate E, $\underline{B}_D$ in (5.54),

1c. define the new model $\underline{y} = P(b_c, \underline{e}) + \underline{e}_P$,

Step 2 Calculate x based on the refinement of b_c,

2a. choose an initial guess b_c^0,

2b. calculate $\partial_{b_c} P(b_c^0, e)$ based on (5.56),

2c. calculate

$$\hat{b}_c = \frac{\partial_{b_c} P^T(b_c^0, e) W}{\partial_{b_c} P^T(b_c^0, e) W \partial_{b_c} P(b_c^0, e)} (y - P(b_c^0, e)) + b_c^0, \tag{5.61}$$

and terminate the iteration if

$$\|\hat{b}_c - b_c^0\|^2 < \epsilon, \tag{5.62}$$

otherwise, choose $b_c^0 = \hat{b}_c$, and go to Step 2b.

2d. finally, calculate $\hat{x}_u$ based on (5.55).

Note that the norm in (5.62) should be taken w.r.t. the inverse of the variance of $\hat{b}_c$ to make the it insensitive to scale changes [41]. However, in this paper, we keep the norm un-weighted for simplicity.

5.2.2.2 ToA/RSS

Here, the offset b_c does not exist, which makes it a bit simpler than TDoA. The derivations for TDoA can be easily adopted for ToA/RSS and only minor changes are required.

For ToA/RSS, it is assumed that the k-th element of x is chosen, the other elements of x can be estimated as functions of x_k as

$$[\underline{\hat{x}}_1(x_k), ..., \underline{\hat{x}}_{k-1}(x_k), \underline{\hat{x}}_{k+1}(x_k), ..., \underline{\hat{x}}_n(x_k)]^T = (B^T W_D B)^{-1} B^T W_D (\underline{y}_D - C x_k), \tag{5.63}$$

where B is an $(m-1) \times (n-1)$ matrix that contains all columns of A_D, except the k-th column C.

Further denote $D_{(n-1)\times(m-1)} = (B^T W_D B)^{-1} B^T W_D$, $F_{n\times 1} = [0, ..., 0, \underbrace{1}_{k-th}, 0..., 0]^T$, and $E_{n\times(m-1)} = [D_1^T, ..., D_{k-1}^T, 0, D_{k+1}^T, ..., D_{n-1}^T]^T$, with D_1 stands for the 1-st row of D. Note that E here is different from the one in (5.54). The key expressions can be given as,

$$\begin{aligned}
L(x_k, \underline{e}) &= \underline{\hat{x}}_u(x_k) = E\underline{y}_D + (F - EC)x_k, && (5.64)\\
P_i(x_k, \underline{e}) &= \|E\underline{y}_D + (F - EC)x_k - x_i\|, && (5.65)\\
\partial_{x_k} P_i(x_k, \underline{e}) &= \frac{(E\underline{y}_D + (F - EC)x_k - x_i)^T (F - EC)}{\|E\underline{y}_D + (F - EC)x_k - x_i\|}, && (5.66)\\
\hat{x}_k &= \frac{\partial_{x_k} P^T(x_k^0, e) W}{\partial_{x_k} P^T(x_k^0, e) W \partial_{x_k} P(x_k^0, e)} (y - P(x_k^0, e)) + x_k^0. && (5.67)
\end{aligned}$$

The mechanics of the 1DI method in the ToA/RSS case can be derived in a very similar way as for TDoA, and is therefore not repeated here.

5.2.3 Parameter Choice

In Step 1, z can be chosen as a linear combination of the elements of x or even a non-linear combination, e.g. $x^T L x$. The best parameter to iterate on should be chosen based on

the requirements of the application, e.g. minimum RMSE or maximum likelihood. However, finding the parameter is in general a non-linear problem and can be computationally involved. This can jeopardize the low complexity nature of the 1DI method, which is a favorable property compared with the NLS estimator. Simplifications are therefore desirable.

Based on the fact that the major reduction of RMSE of the 1DI estimator, compared to the LLS estimator, comes from the refinement of the parameter chosen in Step 1, we propose to choose a single element in x, which has worst precision.

5.2.3.1 TDoA

In this case, the design matrix $[A_D,\ \underline{B}_D]$ in (5.53) is usually badly-conditioned, due to the last column $\underline{B}_D = -2[\underline{y}_r - \underline{y}_1, ..., \underline{y}_r - \underline{y}_{(m-1)}]$, corresponding to the unknown b_c. If the parameter is chosen as $x_k, k = 1, ..., n$ and $\underline{B}_D$ remains in the new design matrix, it is still bad-conditioned, and the estimation (5.55) is of very bad accuracy. Thus, in TDoA cases, we propose to choose b_c as the parameter to be refined in Step 2. The resulting design matrix A_D is in a much better condition and is free of random variables.

5.2.3.2 ToA/RSS

The parameter of worst precision is not so obvious to choose, since the design matrix A_D is usually in a reasonable to good condition. From a geometric point of view, this is equivalent to choosing the coordinate on the longest axis of the (hyper) ellipsoid defined by $x^T Q_{\hat{x}\hat{x}}^{-1} x = c$, with $Q_{\hat{x}\hat{x}}$ the variance matrix of the SOF estimator and c a constant. Based on this fact, the parameter of worst precision can be picked by reparameterization, making use of EVD. The resulting parameter is a linear combination of the elements in x, and more details can be found in [63], where it has been shown that the mentioned way of parameter choice actually improves the performance of the 1DI method. However, EVD itself is usually a non-linear problem that does not have easy direct solutions. For a solution with low-computational load, here we narrow the choice of parameter in ToA/RSS cases to the elements in x, i.e. x_k.

The parameter x_k can be chosen such that the RMSE can be minimized, based on the expressions given in Section 5.2.4. However, to calculate the RMSE value for n different parameter choices can still be computationally heavy. Fortunately, in most of the indoor situations, one can simply choose the height (z-coordinate) as the parameter to refine. The reason is that most of the indoor rooms have smaller heights than lengths and widths, and the transmitters are usually installed in the corners of a room or even all on the ceiling above the user (to avoid blockage). This type of system geometry makes the confidence region in the local height direction much more elongated than in the horizontal plane. This can be easily verified with the theoretical error analysis in [90], for non-linear least-squares estimators. As a result, the estimated height is usually of worst precision. This point is verified in Section 5.2.7, by comparing the outcomes of two different ways of choosing x_k.

5.2.4 Error Analysis

In this section, the RMSE of the 1DI estimator is derived, and we assume that the elements in $\underline{e}$ are independent and $\mathsf{E}\{\underline{e}\} = 0$. As described in Section 3.3.4, the unknown parameter z in (5.49) can be related to y with a non-linear mapping $M(.)$, i.e. $z = M(y)$, and we define $\hat{\underline{z}} = M(\underline{y})$. Taylorizing $M(\underline{y})$ at $y = E\{\underline{y}\}$, with $\underline{e} = \underline{y} - y$, gives:

$$\hat{\underline{z}} = M(y) + \partial_y M(y)^T \underline{e} + \frac{1}{2}\underline{e}^T \partial^2_{yy^T} M(y)\underline{e} + \cdots \tag{5.68}$$

Define the bias in $\hat{\underline{z}}$ as $\underline{\mu}_{\hat{z}} = \hat{\underline{z}} - z$, which is given, together with the variance of $\hat{\underline{z}}$, as

$$\begin{aligned} \mu_{\hat{z}} &= \mathsf{E}\{\underline{\mu}_{\hat{z}}\} = \frac{1}{2} trace\{\partial^2_{yy^T} M(y) Q_{ee}\}, & (5.69)\\ Q_{\hat{z}\hat{z}} &= Q_{\mu_{\hat{z}}\mu_{\hat{z}}} = \mathsf{E}\{(\underline{\mu}_{\hat{z}} - \mu_{\hat{z}})(\underline{\mu}_{\hat{z}} - \mu_{\hat{z}})^T\} = \partial_y M(y)^T Q_{ee} \partial_y M(y), & (5.70) \end{aligned}$$

where the terms with order higher than 2 are ignored.

5.2.4.1 TDoA

In this case, $z = b_c$, and based on (5.55), one obtains $\hat{\underline{x}}_u(\hat{\underline{b}}_c) = L(\hat{\underline{b}}_c, \underline{e}) = G\underline{y}_D - G\underline{B}_D\hat{\underline{b}}_c + F\hat{\underline{b}}_c$, where $G_{(n+1)\times(m-1)} = [E^T, 0]^T$, $F_{(n+1)\times 1} = [0, ..., 0, 1]^T$. Denoting the error term $\underline{\mu}_{\hat{x}} = \hat{\underline{x}} - x$, the bias and variance of $\hat{\underline{x}}$ can be given based on (5.69) and (5.70) as,

$$\begin{aligned} \mu_{\hat{x}} &= \mathsf{E}\{\hat{\underline{x}} - x\} = G\mu_{e_D} - GK\partial_y M(y) + (F - GB_D)\tfrac{1}{2} trace\Big\{\partial^2_{yy^T} M(y) Q_{ee}\Big\}, & (5.71)\\ Q_{\hat{x}\hat{x}} &= \mathsf{E}\Big\{(\hat{\underline{x}} - x - \mu_{\hat{x}})(\hat{\underline{x}} - x - \mu_{\hat{x}})^T\Big\} \\ &= G\mathsf{E}\Big\{\underline{e}_D\underline{e}_D^T\Big\}G^T + J + J^T + (F - GB_D)\partial_y M(y)^T Q_{ee}\partial_y M(y)(F - GB_D)^T - \mu_{\hat{x}}\mu_{\hat{x}}^T, & (5.72) \end{aligned}$$

where $J = \mathsf{E}\Big\{G\underline{e}_D\underline{\mu}_{\hat{b}_c}^T(F - GB_D)^T\Big\}$. The details of the terms $\partial_y M(y)$, $\partial^2_{yy^T} M(y)$, K, and J are derived in Appendix A.3.2

Based on (5.71) and (5.72), the RMSE can be written as

$$\begin{aligned} \mathsf{RMSE}_{\hat{x}} &= trace\{Q_{\hat{x}\hat{x}}\} + \mu_{\hat{x}}^T\mu_{\hat{x}} & (5.73)\\ &= trace\{G\mathsf{E}\Big\{\underline{e}_D\underline{e}_D^T\Big\}G^T + J + J^T + (F - GB_D)\partial_y M(y)^T Q_{ee}\partial_y M(y)(F - GB_D)^T\}. \end{aligned}$$

The first term $G\mathsf{E}\left\{\underline{e}_D\underline{e}_D^T\right\}G^T$ comes from the error in the estimation of $L(z, \underline{e})$ in Step 1, $J + J^T$ represent the cross terms of the estimation errors in Step 1 and Step 2, and the last term comes from the estimation errors in Step 2. Equation (5.73) requires knowledge of the true value x, which is not available in practice. Therefore, the estimate $\hat{x}$ can be used instead. If one wants to use the theoretical RMSE expression for parameter choice in Step 1, before $\hat{x}$ is estimated, the initial guess obtained using a direct method can be used instead.

5.2.4.2 ToA/RSS

Here, one has $z = x_k$ and it can be shown in a similar way that

$$\begin{aligned} \mu_{\hat{x}} &= E\mu_{e_D} + (F - EC)\tfrac{1}{2} trace\Big\{\partial^2_{yy^T} M(y) Q_{ee}\Big\}, && (5.74) \\ Q_{\hat{x}\hat{x}} &= E\mathsf{E}\Big\{\underline{e}_D \underline{e}_D^T\Big\} E^T + J + J^T + (F - EC)\partial_y M(y)^T Q_{ee} \partial_y M(y)(F - EC)^T - \mu_{\hat{x}}\mu_{\hat{x}}^T, && (5.75) \\ RMSE_{\hat{x}} &= trace\{E\mathsf{E}\Big\{\underline{e}_D \underline{e}_D^T\Big\} E^T + J + J^T + (F - EC)\partial_y M(y)^T Q_{ee} \partial_y M(y)(F - EC)^T\}. && (5.76) \end{aligned}$$

The derivation of $\partial_y M(y)$, $\partial^2_{yy^T} M(y)$ and J in ToA/RSS cases can be done based on the same idea. The obtained results are very similar as for TDoA, and one can directly use the expression of $\partial_y M(y)$ (A.19), $\partial^2_{yy^T} M(y)$ (A.20) and J (A.21), and only needs to change $G \to E$, $B_D \to C$ and $K \to 0$. Again, we should note that the expressions for E and F in ToA/RSS cases are very similar but *not* identical to the ones in TDoA cases.

5.2.5 Convergence Rate

Recall from (5.61) and (5.67), the chosen parameter is updated in the $j+1$-th iteration as

$$z^{j+1} = \frac{\partial_z P^T(z^j, e) W}{\partial_z P^T(z^j, e) W \partial_z P(z^j, e)} (y - P(z^j, e)) + z^j. \quad (5.77)$$

Taylorizing (5.77) at the final solution $\hat{z}$, together with $\partial_{\hat{z}} P^T(\hat{z}, e) W (y - P(\hat{z}, e)) = 0$, gives

$$z^{j+1} - \hat{z} = \frac{\partial^2_{z^2} P^T(\hat{z}, e) W}{\partial_z P^T(\hat{z}, e) W \partial_z P(\hat{z}, e)} (y - P(\hat{z}, e))(z^j - \hat{z}) + o(\|z^j - \hat{z}\|). \quad (5.78)$$

With z^j close to $\hat{z}$, the term $o(\|z^j - \hat{z}\|)$ can be ignored, and the 1DI method has a linear rate of convergence. Note that (5.78) is a special case of the general expression in [41].

5.2.6 Flop Count

A flop is defined as a floating point operation, e.g. an add, a multiplication or a square root. The number of required flops to calculate the 1DI estimator (5.51) and the NLS estimator (3.50) obtained using the Gauss-Newton method, are compared in this section, for both TDoA and ToA/RSS, based on the number of flops required in each iteration.

5.2.6.1 TDoA

In Appendix A.3.3.1, we derived that in each iteration, the number of required flops for the 1DI method, reads

$$\text{flops}_{1DI} = 2m^2 + 3mn + 9m + 2n + 2. \quad (5.79)$$

If W is simply a diagonal matrix, the total number of flops becomes

$$\text{flops}'_{1DI} = 3mn + 11m + 2n + 2. \quad (5.80)$$

In each iteration, the calculations for the Gauss-Newton method need:

$$\text{flops}_{GN} = n^3/3 + 2m^2n + 2mn^2 + 2n^2 + 7mn + 4m + 6n + 4/3. \tag{5.81}$$

If W is simply a diagonal matrix, with unequal diagonal entries, the total number of flops becomes

$$\text{flops}'_{GN} = n^3/3 + 2mn^2 + 2n^2 + 9mn + 4m + 6n + 4/3. \tag{5.82}$$

5.2.6.2 ToA/RSS

For the ToA/RSS case, the derivations are given in Appendix A.3.3.2.

In each iteration, the number of required flops for the 1DI method, reads

$$\text{flops}_{1DI} = 2m^2 + 3mn + 7m + 2n + 2, \tag{5.83}$$

and if W is diagonal the total number of flops becomes

$$\text{flops}'_{1DI} = 3mn + 9m + 2n + 2. \tag{5.84}$$

The total number of flops for the Gauss-Newton method reads

$$\text{flops}_{GN} = n^3/3 + 2m^2n + 2mn^2 + n^2 + 6mn + m + 2n - 1. \tag{5.85}$$

If W is diagonal the total number of flops becomes

$$\text{flops}'_{GN} \;=\; n^3/3 + 2mn^2 + n^2 + 8mn + m + 2n - 1. \tag{5.86}$$

One may argue that if the number of flops, required to obtain the initial guess, is dominating the total computational load, then the saved computations by using the 1DI method may be negligible. Fortunately, the major computations involved in this initialization stage are no more than the major computations in one iteration of the Gauss-Newton method with full dimension. Suppose that the initial guess is obtained with the LLS method, the number of rows in A_L is $m-1$, and the dominant term in the flop count becomes $2(m-1)^2n$, smaller than the one for the Gauss-Newton method. Note that the estimation of $L(z, \underline{e})$ required by the 1DI method can also be done based on the same direct method, and no extra computations are involved.

5.2.7 Validation and Results

In this section, we validate the proposed 1DI method using timing based actual UWB range measurements. The details of the measurement campaign setup and the geometry of the positioning system for validation can be found in Chapter 2.

Supplementary (full) simulations are carried out with the geometry in Figure 5.4 (simulating both geometry and measurements), where 5 transmitters are installed in the corners of a

$6 \times 6 \times 3$ m^3 room, and the user positions are uniformly distributed in a space of $4 \times 4 \times 1$ m^3, in the room center.

The SOF method, the 1DI method, and the Gauss-Newton method (solving (3.50)) are tested in weighted cases. The weight matrices are chosen based on the statistics of the range error $\underline{e}$. Here, the range errors are modeled according to (2.5) and (2.6):

$$\begin{aligned} \mu_i(B, d_i) &= 0.0148 \exp(-B/0.48) d_i \text{ [m]}, && (5.87) \\ \sigma_i(B, d_i) &= \sqrt{[Q_{ee}(B,d)]_{ii}} = 0.016(0.64 \exp(-B/0.6) d_i^{1.5} + 1) \text{ [m]}, && (5.88) \\ Q_{ee}(B,d)]_{ij} &= 0, \ \ i \neq j. && (5.89) \end{aligned}$$

with d in meter and B in GHz. In practice, the true distances are not known. Equation (5.87) and (5.88) need to be evaluated based on the range measurements, or the estimated distances using the initial guess. In our validation, it is assumed that the term $\mu(B, d)$ has already been canceled, i.e. $\underline{e} \sim N_m(0, Q_{ee}(B, d))$.

For ToA, with 1DI, two different ways of choosing the parameter to iterate in Step 2 are tested, as suggested in Section 5.2.3. In Setup 1, the coordinate on the z-axis (the height) is chosen, while in Setup 2, the parameter is chosen out of the k elements of x as the one that minimizes (5.76), which is evaluated using the initial guess. For ToA, the initial guess is obtained using the SOF method. For TDoA, to avoid the numerical problem with the SOF method, we adapt it based on the idea of the MOF method in [43]:

1. Obtain the mapping (5.55).
2. Plugging $L(b_c, \underline{e})$ into the smallest (pseudo)range measurement in (3.2), ending up with a quadratic equation in b_c.
3. Solving the equation gives us two candidates for b_c and therefore two candidates for x. The initial guess for b_c is picked from: $\hat{x} = \arg\min_{x_1, x_2} \|\underline{y} - A(x)\|^2_{Q_{yy}^{-1}}$.

Furthermore, the threshold ϵ in (5.62) is chosen as 10^{-24}, and the algorithms will be intentionally terminated if the number of spent iterations hits 50. The same rules hold for the Gauss-Newton method.

Results are plotted in Figure 5.5 to Figure 5.9. On the horizontal axis, the range error STDs obtained under different signal bandwidths are given, and the STD corresponding to one specific bandwidth is empirically calculated using 400 measurements.

5.2.7.1 RMSE

The empirical RMSE curves of the tested methods obtained with real data are plotted in Figure 5.5. The Gauss-Newton estimator has the lowest RMSE. The performance of the 1DI estimator is better than the SOF estimator, and is very close to the Gauss-Newton estimator, especially in the TDoA case. Due to numerical problems, the RMSE of the SOF estimator in the TDoA case is of more than hundreds of meters, and is not shown in the figure.

5.2.7.2 Setup 1 vs Setup 2

For ToA/RSS cases, the 1DI method is tested with two setups with different parameter choices. In general, Setup 2 is always equal to, or better than Setup 1. This point is supported by the results in Figure 5.5. We should note that Setup 1 is worse than Setup 2 because the system geometry in Figure 5.5 is generated in a randomly uniform way, see Section 2.2.1.2. In a typical indoor room, e.g. Figure 5.4, the precision of the local height is usually the worst, and the corresponding coordinate can be directly chosen in Step 1. This point is verified by simulation, where Setup 1 and Setup 2 behave very similarly, see Figure 5.6.

5.2.7.3 Empirical vs Theoretical

In Figure 5.7, the empirical RMSE curves obtained using the real data are compared to the theoretical RMSE curves (5.73) and (5.76). The empirical and theoretical curves are close but still deviate. This may be caused by the imperfectness of the model (5.88), where the parameters are estimated using only 400 measurements. In Figure 5.8, we show the empirical and theoretical curves obtained with the simulation, and the results show that the empirical and theoretical curves match perfectly with each other for both TDoA and ToA/RSS cases.

5.2.7.4 Convergence Rate

In general, it is difficult to claim which one of the 1DI and the Gauss-Newton method converges faster, since the convergence rate is dependent on both geometry and the error $\underline{e}$, see (5.78) and [41]. In the *current test* with real data, the mean numbers of the required iterations of the two methods are shown in Figure 5.9, where the 1DI method requires fewer iterations to converge, especially in TDoA cases.

5.2.7.5 Flop Count

The calculations of the flop count for our test, with $m = 5$ and $n = 3$, can be done based on equations (5.80), (5.82), (5.84) and (5.86). Note that these equations are chosen since the weight matrix W is diagonal, see (5.89). The results show that the proposed 1DI method, as compared to the Gauss-Newton method, saves 63% and 59% on flops in TDoA and ToA/RSS cases, respectively.

5.3 Summary

In this chapter, we proposed two possible improvements on the existing methods.

Concerning the direct methods, more specifically, the SOF methods, a new non-iterative localization technique MOF II was proposed. It has been proved that the proposed method provides a better estimator then the SOF methods, and it avoids iteration by exploiting the similarity of multiple full positioning models. Validation was performed based on real

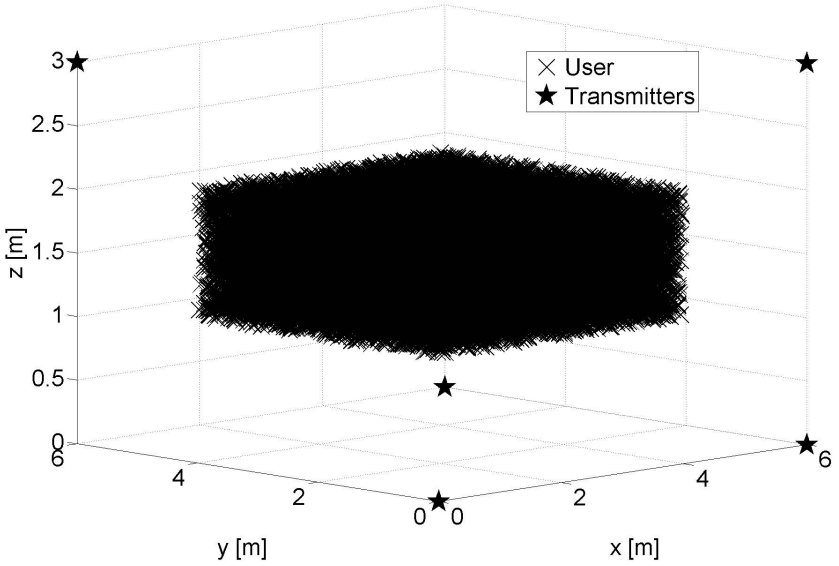

Figure 5.4: Geometry setup used in simulations, with 5 transmitters.

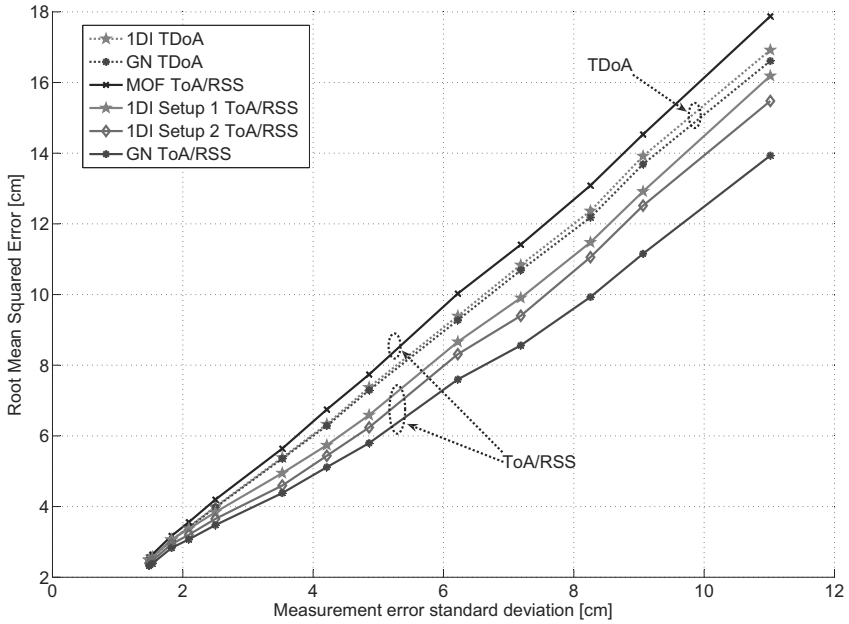

Figure 5.5: RMSE of the estimates obtained using the real UWB data, for different methods.

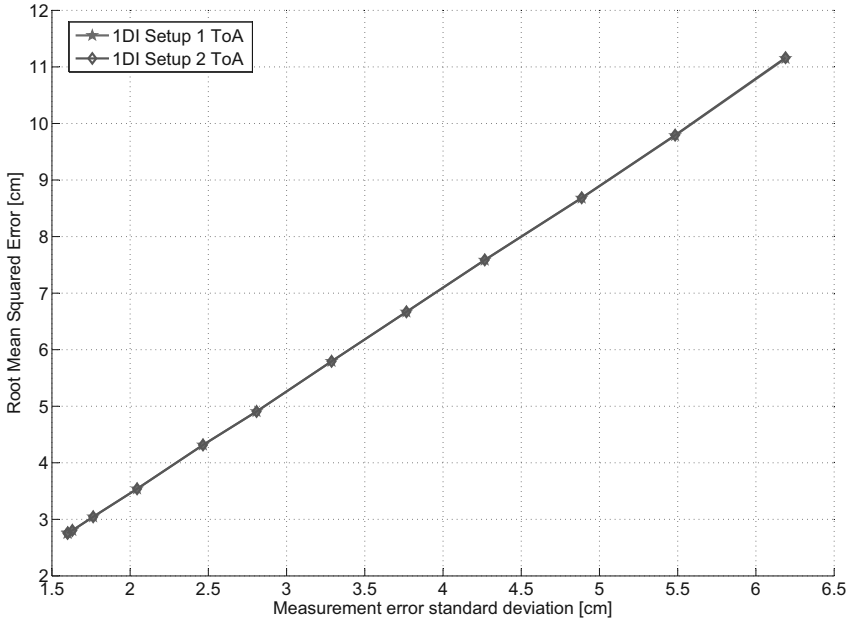

Figure 5.6: Simulated RMSE curves of the estimates obtained with two setups of the 1DI method.

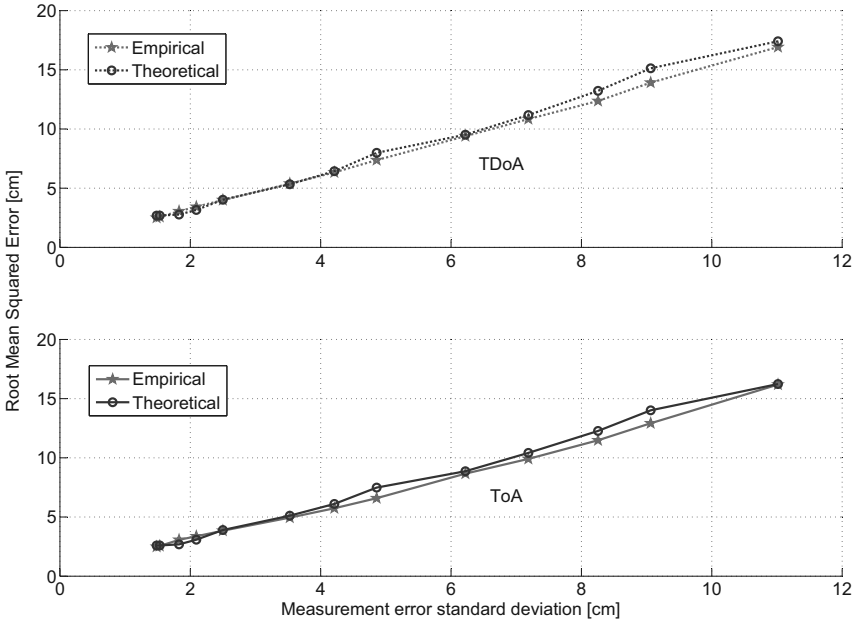

Figure 5.7: The theoretical RMSE curve v.s. the empirical curve obtained using real UWB data, for the 1DI method with Setup 1.

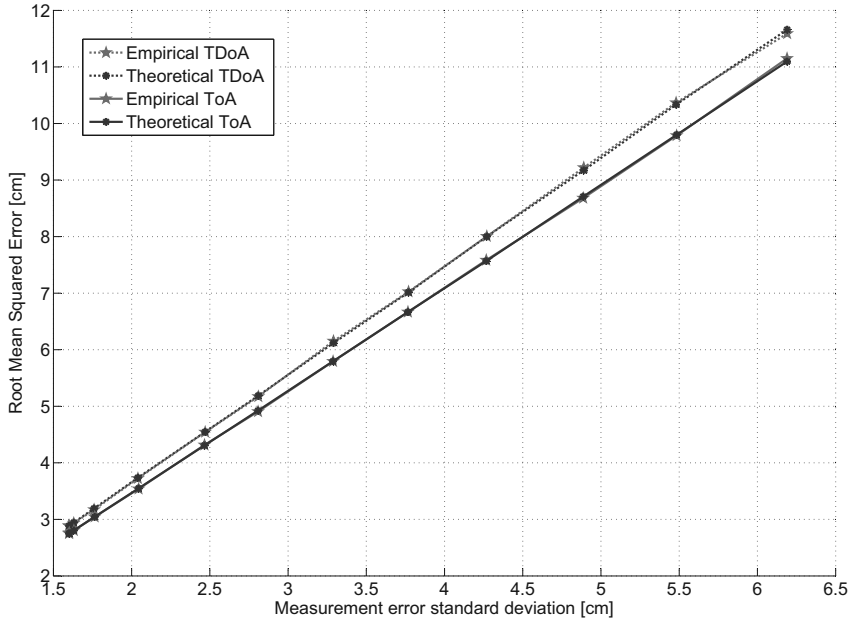

Figure 5.8: The theoretical RMSE curve v.s. the empirical curve obtained using simulation, for the 1DI method with Setup 1.

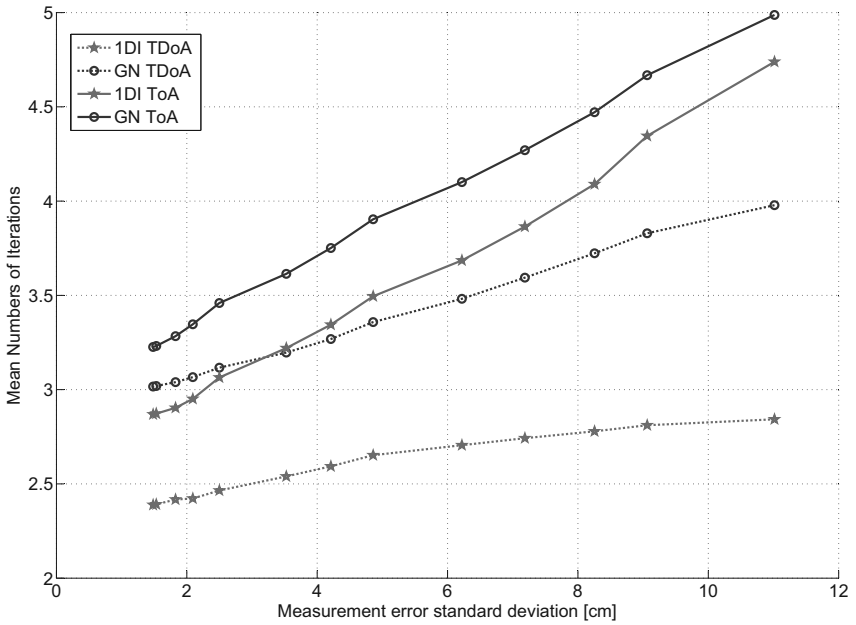

Figure 5.9: Mean of the required number of iterations by the 1DI and the Gauss-Newton methods, obtained using the real UWB data.

UWB measurements collected in typical office scenarios. Results show that the proposed method works better than any other tested non-iterative positioning methods, since it provides estimators with the lowest RMSE, both in un-weighted and weighted cases. In practice, when the position application aims for m- or dm-level accuracy, the estimates of the proposed method can be directly used; when cm-level or even higher level accuracy is required, the ranges needs to be estimated based on ToA/TDoA, and the proposed method can provide estimates as the initial guess to start ID methods such as the Gauss-Newton method.

Concerning the iterative methods, the TLS framework, consisting of 2 steps, is proposed to realize low-complexity positioning with high precision estimators, covering both TDoA and ToA/RSS cases. The low-complexity nature of the framework comes from Step 1, where the positioning problem is transformed from $(n+1)$- or n-dimension to a lower dimension. In Step 2, the high precision estimator is obtained by iteratively refining the unknown with all of the original range measurements. Within the TLS framework, the 1DI method is introduced, which applies the SOF method in Step 1 to reduce the number of unknowns to one and the Gauss-Newton method to estimate the single unknown in Step 2. The way of choosing the parameter in Step 1 of the 1DI method has been discussed. The theoretical performance of the 1DI method is also analyzed, including the expression for RMSE, the convergence rate and the required number of flops in each iteration. It has been shown that in each iteration, the 1DI method can save up to 67% (theoretically when m is large) of the computations, compared to the Gauss-Newton method. Validation of the work is done mainly based on real UWB radio measurements, together with some supplementary simulations. Results show that the 1DI method is better than the SOF method, and its performance is close to the Gauss-Newton method, especially in the TDoA case. The derived theoretical RMSE perfectly matches the empirical value in simulations, and small deviations exist with real data, which can be caused by the limited number of available data.

Although both methods were introduced and applied to the self positioning scenario (like satellite navigation), they also work for tracking systems in which the user sends signals to the base stations where signal processing and position estimation are accomplished.

6 NLoS Identification and Mitigation

In Chapters 4-5, we have discussed several positioning techniques which use range measurements to estimate the unknown position. The range measurements are usually obtained by exploiting timing information (ToA, TDoA) and RSS. It has been shown that cm-level positioning accuracy can be achieved with accurate range ToA/TDoA measurements, i.e. ranging using UWB signals with cm-level accuracy. However, in the validation work up till now, all the range measurements are obtained under LoS propagation, where part of the received signal travels from the transmitter to the receiver without interacting with any material but air. In practice, a lot of objects in indoor environments can block the direct ray between the transmitter and the receiver, which cause NLoS effects. The LoS scenario and several NLoS scenarios are depicted in Figure 6.1. In this chapter, the part of the received signal, which reaches the receiver without being reflected, will be referred as the *direct path*.

In NLoS situations, compared to LoS, estimated ToAs contain extra delays and estimated RSSs contain extra pathlosses. In some extreme cases, the received signal contains no direct path, e.g. when metallic objects are standing between the transmitter and the receiver. These NLoS effects introduce extra biases in range measurements, and propagate into position estimates as errors. In wireless positioning, the NLoS effect is one of the dominant factors that affect positioning accuracy.

This chapter focuses on the topic: NLoS identification and mitigation. First, a brief review of existing techniques is provided. Some of these ideas helped to develop our test schemes. We use both timing-based range and RSS measurements in the proposed schemes, which include:

1. A most general classification between multiple composite hypotheses based on Generalized Likelihood (GL) considering *all* measurements. Given m transmitters (links), the number of hypotheses is 2^m. Under each hypothesis, the corresponding GL is computed and the hypothesis corresponding to the largest one is selected.

2. Similar to the first one, except there are some of the unknown parameters are precalculated using a model, hence less unknown parameters in evaluating the GLs and computational load is reduced.

3. The values for the unknown distances required to evaluate the likelihoods of RSS measurements are approximated by the estimates obtained using ToA-ranges, no matter they are LoS or NLoS. Hence, there is no need to compute GLs for all possible hypotheses which are replaced by the approximated likelihoods. Testing is

done per link for m rounds, rather than per hypothesis for 2^m rounds. In this way, the computational load is greatly reduced.

4. ToA-ranges are directly used as the values for the unknown distances to evaluate the likelihoods of RSS measurements, and no estimation is performed for this approximation. Therefore, computational load is further reduced.

Validation is performed in the end, and results show that a high detection rate of 99% can be achieved, when radio UWB timing-based range measurements obtained under the full bandwidth of 7.5 GHz are used.

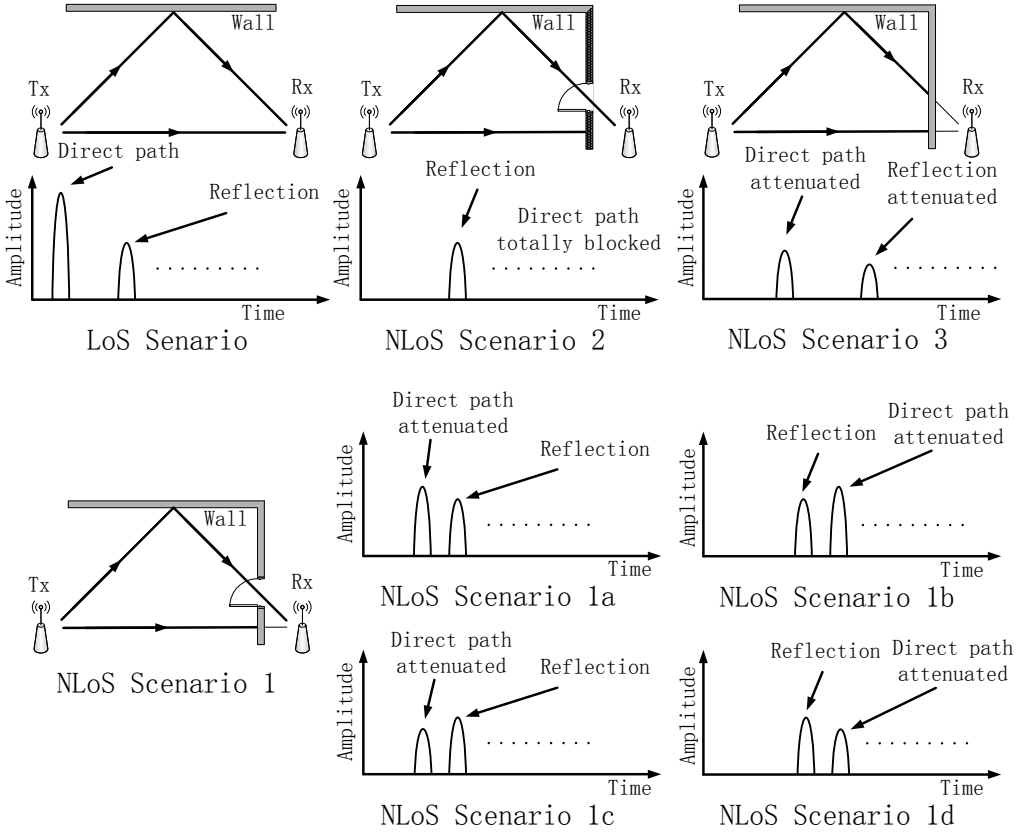

Figure 6.1: LoS and NLoS scenarios where only the direct path and the first reflection are considered, and later paths are represented as '· · · · · ·' in the figure. LoS scenario: direct path arrives earlier than the reflection and is stronger. In NLoS scenarios, the direct path is attenuated and delayed. NLoS scenario 1a: direct path arrives earlier than the reflection and is stronger; NLoS scenario 1b: direct path arrives later than the reflection but is stronger; NLoS scenario 1c: direct path arrives earlier than the reflection but is weaker; NLoS scenario 1d: direct path arrives later than the reflection and is weaker. Please note that NLoS scenario 3 can be looked on as an scaled version of the LoS scenario, but possibly with larger separation in time between the two paths. Although, here we assume that the receiver is able to correctly detect all attenuated paths, the case where an attenuated direct path is un-detected can be described in a similar way as the NLoS scenario 2.

6.1 Existing Techniques: Brief Review

The NLoS identification and mitigation techniques have been extensively studied in literature. These techniques solve the problem of identification by addressing characteristics of

the received signal that are different in LoS and NLoS conditions. Based on the targeted characteristic(s), these techniques can be divided into different groups, which are reviewed as follows.

6.1.1 Full Channel Impulse Response and Received Signal

6.1.1.1 Confidence Metric

In [8], to achieve better ToA estimation, a NLoS identification method called Confidence Metric (CM) was introduced. This method focuses on two features of the CIR, 1) Signal-to-Noise Ratio (SNR) of the first path and 2) the difference between the first path power and the maximum path power. The decision is made as follows:

$$T_{CM} = \log_{10}\left(\frac{\alpha_1^2}{\sigma_N^2}\right) + \log_{10}\left(\frac{\alpha_1^2}{\alpha_m^2}\frac{1}{A_0(\tau_1,\tau_m)}\right) \underset{\text{NLoS}}{\overset{\text{LoS}}{\gtrless}} \Theta_{CM}, \tag{6.1}$$

where σ_N^2 is the noise variance, α_1 and τ_1 are the amplitude and ToA of the first path, α_m and τ_m are of the path with maximum power, and $A_0(\tau_1,\tau_m)$ is a term used to compensate α_m^2 for the additional pathloss corresponding to the distance between the first and the strongest path. In [8], $A_0(\tau_1,\tau_m)$ is chosen as:

$$A_0(\tau_1,\tau_m) = \frac{\tau_1^n}{\tau_m^n}, \tag{6.2}$$

where n is the pathloss exponent, which can be obtained empirically.

Note that the first term $\log_{10}\left(\frac{\alpha_1^2}{\sigma_N^2}\right)$ in (6.1) accounts for SNR of the first path and the second term $\log_{10}\left(\frac{\alpha_1^2}{\alpha_m^2}\frac{1}{A_0(\tau_1,\tau_m)}\right)$ accounts for the power difference (in dB) between the first and the strongest path.

The reasoning and performance of the CM method can be described as follows:

- in a LoS scenario, generally, the SNR of the first path is large, and the first path power is larger than the power of any other detected path,
- in NLoS scenario 1a, the SNR of the first path is reduced due to the NLoS attenuation in the direct path, and the power difference is also reduced. Hence, it is possible to decide this scenario as NLoS using (6.1), with both terms in (6.1) becoming smaller. The more the direct path is attenuated, the larger the chance of a (correct) NLoS decision is.
- in NLoS scenario 1b, the first path corresponds to the reflection, so the SNR is reduced, and the power difference is greatly reduced, since the later arrived direct path is stronger. A decision of NLoS can therefore be made based on (6.1), where the power difference term plays a more important role than the SNR term.
- in NLoS scenario 1c, the SNR of the first path is greatly reduced due to the strong NLoS attenuation in the direct path, and the power difference is also greatly reduced since the later arrived reflection is stronger. Therefore, scenario 1c can be easily detected as NLoS.

- in NLoS scenario 1d, the first path corresponds to a reflection with reduced SNR, but it still has stronger power than the later arrived attenuated direct path. In this case, a decision of NLoS is possible, but dependent on the direct path power. The more the direct path is attenuated, the larger the power difference term is, and the smaller the chance of a (correct) NLoS decision is.
- in NLoS scenario 2, the situation is similar as a LoS case with a larger true distance, implying a difficult detection of NLoS. The decision is likely to be made as LoS.
- in NLoS scenario 3, the first path SNR is reduced due to NLoS attenuation, but the power difference is similar as in a LoS case. Thus, a decision of NLoS can therefore be made with the SNR term in (6.1) playing an important role in the detection.

To summarize, if both terms in (6.1) are large, resulting in a large confidence metric T_{CM}, then the condition is very likely to be LoS, otherwise, a possible NLoS condition is indicated.

One important issue of applying the CM method is to determine the threshold Θ_{CM}, which is unfortunately not covered in [93]. Later in [94], a distance dependent expression for Θ_{CM} is proposed:

$$\Theta_{CM} = \begin{cases} \Theta_{\max} + log_{10}\left(\frac{d_{\max}}{c\tau_1}\right), & \Theta_{CM} \geq \Theta_{\max} \\ \text{unchanged}, & \Theta_{CM} < \Theta_{\max} \end{cases} \tag{6.3}$$

where c is the speed of light, $d_{\max}$ represents the maximum reachable distance in LoS situation, and $\Theta_{\max}$ is the corresponding confidence metric. Note that both $d_{\max}$ and $\Theta_{\max}$ needs to be obtained experimentally.

6.1.1.2 Kurtosis, Mean Excess Delay, Root Mean Square Delay Spread, and Their Combination

In literature, the possibility of distinguishing LoS and NLoS propagation by exploiting CIR statistics has been widely explored. The most popular parameters include, the kurtosis, the Mean Excess Delay (MED) and the Root Mean Square Delay Spread (RMS-DS), which are used to capture both the amplitude and the delay statistics of LoS and NLoS scenarios.

The theoretical definition of parameter κ, the kurtosis of a CIR $h(t)$, can be formulated as follows [95]:

$$\kappa = \frac{\mathsf{E}\left\{(|h(t)| - \mu_{|h|})^4\right\}}{\mathsf{E}\left\{(|h(t)| - \mu_{|h|})^2\right\}^2}, \tag{6.4}$$

where $\mu_{|h|}$ is the mean of $|h(t)|$.

As the kurtosis of a CIR characterizes how peaky the CIR is [95], it can be used to distinguish LoS and NLoS channels [96]. Since a LoS CIR usually has a very strong and sharp LoS direct path, a CIR with a high kurtosis value is likely to represent a LoS channel. Thus, a Simple Likelihood Ratio Test (SLRT) based on kurtosis can be performed as:

$$\frac{f_{\underline{\kappa}}(\kappa|\mu_{\kappa,\mathsf{LoS}}, \sigma_{\kappa,\mathsf{LoS}})}{f_{\underline{\kappa}}(\kappa|\mu_{\kappa,\mathsf{NLoS}}, \sigma_{\kappa,\mathsf{NLoS}})} \underset{\mathsf{NLoS}}{\overset{\mathsf{LoS}}{\gtrless}} 1, \tag{6.5}$$

where $f_{\underline{\kappa}}(\kappa|\mu_{\kappa,\text{LoS}}, \sigma_{\kappa,\text{LoS}})$ and $f_{\underline{\kappa}}(\kappa|\mu_{\kappa,\text{NLoS}}, \sigma_{\kappa,\text{NLoS}})$ are the PDFs of the kurtosis in LoS and NLoS cases, with $\mu_{.}$ and $\sigma_{(.)}$ the mean and STD of $\underline{\kappa}$ under the corresponding case. The underscore indicating that the kurtosis is determined based on observed data in practice. In [95], the PDFs in both cases are determined with simulations as log-normal, and the details of their expression can be found therein. Furthermore, the threshold on the r.h.s. is taken as 1, with the assumption on equal a-priori probabilities, i.e. $P(\text{LoS}) = P(\text{NLoS})$.

The kurtosis characterizes the statistics of the path amplitudes in a CIR, but it does not provide any information on the delay statistics. In literature, the MED [95] and the RMS-DS [95, 97] have also been investigated, which characterize delay statistics of the channel, for LoS and NLoS identification.

Given a CIR, the MED and the RMS-DS can be calculated as given in [95, 82]:

$$\tau_{MED} = \frac{\int_{-\infty}^{+\infty} t|h(t)|^2 dt}{\int_{-\infty}^{+\infty} |h(t)|^2 dt}, \tag{6.6}$$

$$\tau_{RMS-DS} = \sqrt{\frac{\int_{-\infty}^{+\infty} (t-\tau_{MED})^2 |h(t)|^2 dt}{\int_{-\infty}^{+\infty} |h(t)|^2 dt}}. \tag{6.7}$$

It can be observed from the above two equations that the MED characterizes the weighted mean arrival time of the detected paths, while the RMS-DS is a weighted mean of the variability in the ToAs of the paths against the MED. These two parameters can be used for LoS and NLoS identification, since

1. the direct path under NLoS is delayed and attenuated/blocked, which results in a larger weighted mean of arrival time of the paths (MED) than under LoS,
2. suppose that the paths other than the direct one undergo less NLoS effects, e.g. in NLoS scenarios 1a-1d, then these paths are less separated from the larger MED under NLoS, resulting in a smaller RMS-DS value.

Hence, a large MED and a small RMS-DS of a CIR indicate that the signal is likely to have been received through a NLoS propagation. It was verified in [95] that both the MED and RMS-DS follow log-normal distributions. The corresponding SLRTs can be respectively performed as:

$$\frac{f_{\underline{\tau}_{MED}}(\tau_{MED}|\mu_{\tau_{MED},\text{LoS}}, \sigma_{\tau_{MED},\text{LoS}})}{f_{\underline{\tau}_{MED}}(\tau_{MED}|\mu_{\tau_{MED},\text{NLoS}}, \sigma_{\tau_{MED},\text{NLoS}})} \underset{\text{NLoS}}{\overset{\text{LoS}}{\gtrless}} 1, \tag{6.8}$$

$$\frac{f_{\underline{\tau}_{RMS-DS}}(\tau_{RMS-DS}|\mu_{\tau_{RMS-DS},\text{LoS}}, \sigma_{\tau_{RMS-DS},\text{LoS}})}{f_{\underline{\tau}_{RMS-DS}}(\tau_{RMS-DS}|\mu_{\tau_{RMS-DS},\text{NLoS}}, \sigma_{\tau_{RMS-DS},\text{NLoS}})} \underset{\text{NLoS}}{\overset{\text{LoS}}{\gtrless}} 1. \tag{6.9}$$

The notations in the above equations can be explained in a similar way as for (6.5). Now let us have a conceptual check on how the kurtosis, the MED and the RMS-DS behave in the LoS and the NLoS scenarios depicted in Figure 6.1:

- in the LoS scenario, with a strong and sharp LoS peak, the corresponding kurtosis is large, the MED is quite close to the arrival time of the LoS path, and the RMS-DS is larger than in the NLoS cases, which are described in the sequel.

- in NLoS scenario 1a, 1b, 1c, 1d and 2, the kurtosis is smaller than the LoS case with the direct path attenuated/blocked, the MED is larger due to the NLoS delay and attenuation, and the RMS-DS is reduced, since the paths are more closely distributed around the increased MED. Thus, the decision is very likely to be made as NLoS using any one of (6.5), (6.8) or (6.9).
- in NLoS scenario 3, the kurtosis is reduced and the MED is increased, just as in other NLoS cases. The RMS-DS, however, is increased, since those reflected paths experience longer NLoS delays and the corresponding arrival times are more separated. Hence, a decision of NLoS can be made based on either (6.5) or (6.8), but with (6.9), the scenario is likely to be detected as LoS.

In summary, NLoS CIRs are characterized by smaller kurtosis values and larger MEDs. The RMS-DS in most of the NLoS cases are smaller, but in a case like NLoS scenario 3, it can be larger than in a LoS case. These behaviors can be verified by the simulation results shown in [95], where NLoS identification based on the RMS-DS has slightly worse performance than using the kurtosis or the MED.

To obtain better identification results, the authors in [95] further proposed a test where the joint PDF of the kurtosis, the MED and the RMS-DS was used:

$$\frac{f_{joint}(\kappa, \tau_{MED}, \tau_{RMS-DS}|\text{LoS})}{f_{joint}(\kappa, \tau_{MED}, \tau_{RMS-DS}|\text{NLoS})} \underset{\text{NLoS}}{\overset{\text{LoS}}{\gtrless}} 1, \tag{6.10}$$

where, for simplicity, we have used LoSand NLoSto represent the mean and STD values of the kurtosis, the MED and the RMS-DS under the corresponding condition.

However, due to the difficulty in obtaining the expression of $f_{joint}(\kappa, \tau_{MED}, \tau_{RMS-DS})$, a simplified test was suggested in stead of (6.10):

$$\frac{f_{\underline{\kappa}}(\kappa|\text{LoS})}{f_{\underline{\kappa}}(\kappa|\text{NLoS})} \cdot \frac{f_{\underline{\tau}_{MED}}(\tau_{MED}|\text{LoS})}{f_{\underline{\tau}_{MED}}(\tau_{MED}|\text{NLoS})} \cdot \frac{f_{\underline{\tau}_{RMS-DS}}(\tau_{RMS-DS}|\text{LoS})}{f_{\underline{\tau}_{RMS-DS}}(\tau_{RMS-DS}|\text{NLoS})} \underset{\text{NLoS}}{\overset{\text{LoS}}{\gtrless}} 1, \tag{6.11}$$

where the kurtosis, the MED and the RMS-DS were assumed independent.

6.1.1.3 Received Signal Envelope

Other than the CIR, LoS/NLoS identification can also be achieved by exploiting the full received signal. In [98], the authors use the so called Rician K factor to distinguish LoS and NLoS propagations. Given $R(t)$, the envelope of the received signal, the Rician K factor can be estimated in a closed form as [99, 98]:

$$\hat{K} = \frac{-2\hat{\mu}_2^2 + \hat{\mu}_4 - \hat{\mu}_2\sqrt{2\hat{\mu}_2^2 - \hat{\mu}_4}}{\hat{\mu}_2^2 - \hat{\mu}_4}, \tag{6.12}$$

where $\hat{\mu}_j$ is the estimated j-th moment of the Rician distribution, calculated as:

$$\hat{\mu}_j = \frac{1}{N}\sum_{i=0}^{N-1} R^j(iT_s), \tag{6.13}$$

with N the number of available samples, and T_s the sampling interval.

The Rician K factor is defined as the ratio of power in the direct path to power in the scattered paths. With $K > 1$, power of the direct path is greater than the power of the multipath components, indicating a LoS condition; otherwise, with $K < 1$, power of the direct path is considered not strong enough as compared to power of the multipath components, indicating a NLoS condition. Together with (6.12), the corresponding test can be formulated as:

$$\hat{K} \underset{\text{NLoS}}{\overset{\text{LoS}}{\gtrless}} 1. \tag{6.14}$$

According to the definition of the Rician K-factor, it serves in a similar way as the term $\log_{10}\left(\frac{\alpha_1^2}{\alpha_m^2}\frac{1}{A_0(\tau_1,\tau_m)}\right)$ in (6.1), and therefore, similar behaviors are expected in the scenarios in Figure 6.1.

6.1.2 ToA, RSS Measurements and Their Combinations

6.1.2.1 Running Variance of ToA Range Measurements

As described in Chapter 2, a range measurement obtained from ToA is usually modeled as:

- the true distance plus a Normally distributed noise term with zero mean and LoS variance, under the LoS condition,
- the true distance plus a NLoS bias and a Normal noise term with zero mean and NLoS variance, under the NLoS condition.

In practice, given a single ToA range measurement with unknown true distance and unknown possible NLoS bias, it is not possible to perform a SLRT or a Generalized Likelihood Ratio Test (GLRT). Thus, instead of using hypothesis model testing based on a single ToA range measurement, the authors of [100] proposed to identify NLoS based on time series of range measurements.

The idea is that the variance of NLoS range measurements is larger than the variance of LoS measurements. The corresponding test can be interpreted as the following steps:

1. estimate the variance of the range measurements obtained from a transmitter, using the time series of N measurements,
2. compare the estimated variance to a threshold Θ_{RV}, which is related to σ^2_{LoS}, the variance of ToA range measurements in LoS situations.

and the expression of the test is given as:

$$\frac{\sum_{j=1}^{N}(\underline{y}_j - \mu_y)^2}{N-1} \underset{LoS}{\overset{\text{NLoS}}{\gtrless}} \Theta_{RV}, \tag{6.15}$$

where $\underline{y}_j$ represents the j-th history range measurement obtained from one transmitter, with $j = 1, ..., N$, and μ_y is the empirical mean, with $\mu_y = \frac{1}{N}\sum_{j=1}^{N} \underline{y}_j$.

The main problems of this test include:

- Distinguishing two variables based on the difference of their variances is usually difficult and implies poor results. Identifying two variables with different mean values would be much easier.

- The receiver should be static or be moving very slowly. Even in LoS cases, fast moving unit will increase the variance of the N range measurements, which can lead to a wrong decision. For fast moving units, a new threshold is introduced in [94], which takes into consideration the empirical maximum speed of a moving target:

$$\Theta_{RV} = \sigma^2_{\text{LoS}} + \frac{N(N+1)}{12}(v_{max}t_U), \tag{6.16}$$

  where t_U is the update interval. The new threshold, however, fixes wrong decisions with a moving unit to a very limited extent, since the threshold only works properly when a moderate pair of update interval and maximum velocity can be assured [94]. A too large value of $v_{max}t_U$ will disable the detection of NLoS measurements.

- The quality of the variance estimator depends on N, the size of history data. However, a large N inevitably introduces latency in detection. This point can be explained with an example in Figure 6.2. Starting at the 11-th sample, the scheme detects the current situation as NLoS, since the empirical variance is larger than the threshold. When the target moves from NLoS into LoS at the 21-th sample, with 9 NLoS samples and a new LoS sample, the empirical variance is increased, due to the sudden reduction in the new LoS sample. Hence, latency is expected for the running variance scheme to make a correct decision of LoS with a transition from NLoS to LoS. Until the 30-th sample, the updated variance drop below the threshold, and a decision of LoS is made. The latency in this case equals 9 samples. On the contrary, a NLoS sample can be instantly detected with a transition from LoS to NLoS at sample 41, where the new NLoS sample instantly increases the variance to a value above threshold, and the correct decision of NLoS is made. The latency problem of the running variance scheme has been reported in [94].

6.1.2.2 Change of RSS

To avoid the main issues in applying the running variance for LoS/NLoS identification, the authors of [94] proposed a simple test based on RSS, achieving very promising results with both simulations and actual measurements. The test works based on the fact that a sudden decrease of RSS usually indicates a transition from LoS into NLoS condition, and vice versa [94].

Denoting α_j the amplitude of the received signal at update time t_j, the test can be

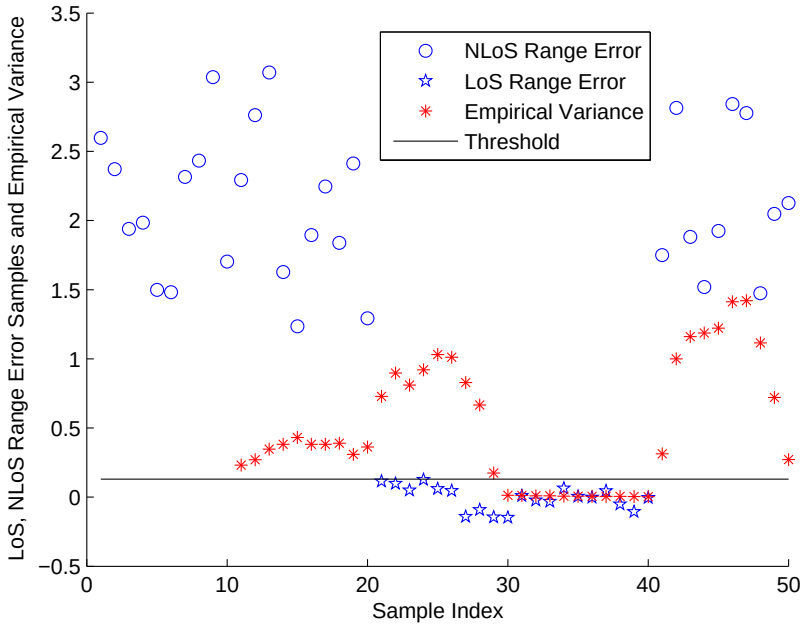

Figure 6.2: A simulated series of LoS and NLoS range error samples together with the empirical variance obtained with (6.15). In total, 50 samples are simulated, with 20 NLoS followed by 20 LoS, and then at last 10 NLoS. The value for N is chosen as 10. The LoS range error is distributed as $\mathcal{N}(0 \text{ m}, 0.01 \text{ m}^2)$ and the NLoS error follows $\mathcal{N}(3 \text{ m}, 0.25 \text{ m}^2)$. The decision threshold is chosen as the mean of LoS and NLoS variances, i.e. $(0.01 + 0.25)/2 = 0.13$.

formulated as:

$$\frac{\alpha_j^2 - \alpha_{j-1}^2}{\max\{\alpha_j^2, \alpha_{j-1}^2\}} \begin{cases} > +\Theta_{CR}, & \text{NLoS} \Rightarrow \text{LoS} \\ < -\Theta_{CR}, & \text{LoS} \Rightarrow \text{NLoS} \\ \text{otherwise}, & \text{unchanged} \end{cases} \tag{6.17}$$

where the threshold Θ_{CR} is expressed as:

$$\Theta_{CR} = 1 - 10^{\frac{-PL}{10}}. \tag{6.18}$$

The term PL in dB accounts for a typical reduction in RSS, when the unit suddenly changes from a LoS to a NLoS condition. The value of PL can be obtained empirically. For example, RSS values obtained under 7.5 GHz bandwidth using the radio UWB system described in Chapter 2 are shown again in Figure 6.3. Based on the separation between LoS and NLoS RSS measurements, the PL value for this data set can be chosen around 7 dB.

Identification based on the change of RSS is simple and easy to implement. Moreover, it has overcome all the listed problems of the running variance scheme described in Section 6.1.2.1:

- Instead of targeting the measurement variance, identification based on the change of RSS in (6.17) is done by comparing the normalized RSS values (normalized to $\max\{\alpha_j^2, \alpha_{j-1}^2\}$ in (6.17)) of the currently detected first path and the first path detected in the previous update interval.

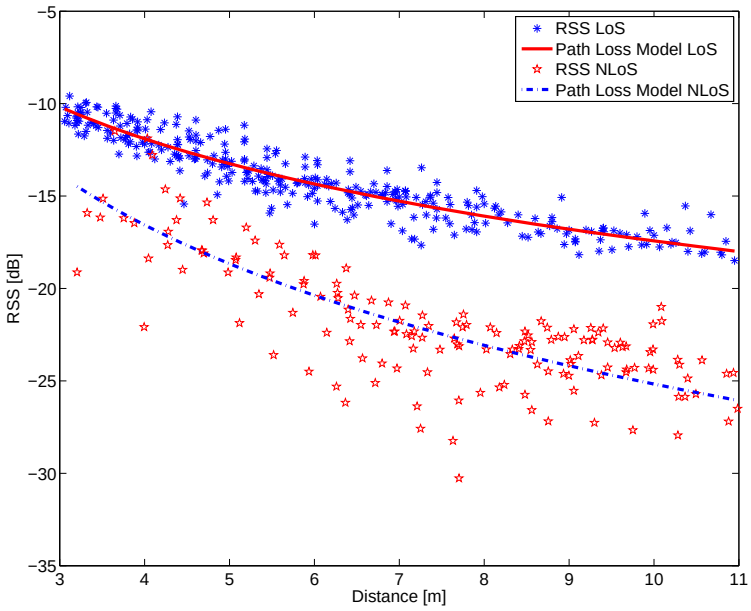

Figure 6.3: Actual UWB radio RSS measurements and pass loss models obtained under LoS and NLoS conditions, respectively. The NLoS measurements are collected in scenario 3, where the walls are made of concrete or brick. The signal bandwidth is 7.5 GHz.

- Identification based on the change of RSS is more robust with a moving target. It can be observed from Figure 6.3 that, with true distance around 4 m, the reduction of RSS due to the transition from LoS to NLoS is about 5 dB. However, a 5 dB pathloss corresponds to a movement from 4 m to about 10 m under LoS. If the sample period is small and the target is not moving too fast, the drop in RSS due to movement under LoS will be negligible compared to the RSS loss due to a transition from LoS to NLoS. Similarly, the increment in RSS due to a transition from NLoS to LoS is much larger than an increment due to a moderate movement toward the transmitter. Hence, the change of RSS due to movement in the same scenario will not be false alarmed as a transition.

- No latency is introduced, since the detection totally based on the difference between the current RSS and the previous RSS.

Identification based on the change of RSS gives quite promising results in detecting the transitions of a unit [94]. However, it has an uncovered issue about how to determine the initial condition of the unit. If no transition is detected, then the scheme fails to give any information on whether a unit is under LoS or NLoS condition. Fortunately, this problem can be easily solved by combining the ToA and RSS measurements, which is reviewed next.

6.1.2.3 ToA and RSS Combined

In [101], the Walfisch-Ikegami pathloss model is adopted for an outdoor celluar system, where the pathloss L is expressed as:

$$\underline{L} = \begin{cases} A_{\mathsf{LoS}} + 26\log_{10} d + \underline{S}_{\mathsf{LoS}}, & \mathsf{LoS} \\ A_{\mathsf{NLoS}} + 38\log_{10} d + \underline{S}_{\mathsf{NLoS}}, & \mathsf{NLoS} \end{cases} \tag{6.19}$$

where A_{LoS} and A_{NLoS} are parameters that should be obtained experimentally in practice. They are dependent on the signal carrier frequency, the transmitter and receiver antenna heights, the structure of buildings and roads, and the street orientation relative to the direct path. The symbol d is the true distance, and $\underline{S}_{\mathsf{LoS}}$ and $\underline{S}_{\mathsf{NLoS}}$ are the pathloss model errors. Note that for indoor systems, log-normal pathloss models are often used instead, as discussed in Chapter 3.

In practice, the true distance is unknown, which makes it impossible to apply hypothesis model testing directly based on (6.19). Based on the fact that a ToA range measurement contains the information of the corresponding true distance, the authors of [101] came up with the idea to combine the ToA and RSS pass loss models. Given the models for ToA range measurements:

$$\underline{y} = \begin{cases} d + \underline{e}_{\mathsf{LoS}}, & \mathsf{LoS} \\ d + \underline{\nabla} + \underline{e}_{\mathsf{NLoS}}, & \mathsf{NLoS} \end{cases} \tag{6.20}$$

where $\underline{e}_{\mathsf{LoS}}$ and $\underline{e}_{\mathsf{NLoS}}$ are the model errors, and $\underline{\nabla}$ is a random variable that accounts for the NLoS bias.

Plugging (6.20) into (6.19), the combined models are expressed as:

$$\underline{L} = \begin{cases} A_{\mathsf{LoS}} + 26\log_{10}(\underline{y} - \underline{e}_{\mathsf{LoS}}) + \underline{S}_{\mathsf{LoS}}, & \mathsf{LoS} \\ A_{\mathsf{NLoS}} + 38\log_{10}(\underline{y} - \underline{\nabla} - \underline{e}_{\mathsf{NLoS}}) + \underline{S}_{\mathsf{NLoS}}. & \mathsf{NLoS} \end{cases} \tag{6.21}$$

Furthermore, assuming $\underline{y} >> \underline{e}_{\mathsf{LoS}}$ and $\underline{y} >> \underline{\nabla} + \underline{e}_{\mathsf{NLoS}}$, the following approximations can be applied [101]:

$$\log_{10}(\underline{y} - \underline{e}_{\mathsf{LoS}}) \approx \log_{10}\underline{y} - \frac{\underline{e}_{\mathsf{LoS}}}{2.3\underline{y}} \tag{6.22}$$

$$\log_{10}(\underline{y} - \underline{\nabla} - \underline{e}_{\mathsf{NLoS}}) \approx \log_{10}\underline{y} - \frac{\underline{\nabla} + \underline{e}_{\mathsf{NLoS}}}{2.3\underline{y}} \tag{6.23}$$

and the model (6.21) can be approximated as:

$$\underline{L} \approx A'_{\mathsf{LoS}} + \underline{S}'_{\mathsf{LoS}} \text{ under LoS} \tag{6.24}$$

$$A'_{\mathsf{LoS}} = A_{\mathsf{LoS}} + 26\log_{10}\underline{y} \text{ and } \underline{S}'_{\mathsf{LoS}} = \underline{S}_{\mathsf{LoS}} - \frac{26\underline{e}_{\mathsf{LoS}}}{2.3\underline{y}}$$

$$\underline{L} \approx A'_{\mathsf{NLoS}} + \underline{S}'_{\mathsf{NLoS}} \text{ under NLoS} \tag{6.25}$$

$$A'_{\mathsf{NLoS}} = A_{\mathsf{NLoS}} - 38\log_{10}\underline{y} \text{ and } \underline{S}'_{\mathsf{NLoS}} = \underline{S}_{\mathsf{NLoS}} - \frac{38(\underline{\nabla} + \underline{e}_{\mathsf{NLoS}})}{2.3\underline{y}}$$

In [101], $\underline{S}_{\mathsf{LoS}}$, $\underline{S}_{\mathsf{NLoS}}$, $\underline{e}_{\mathsf{LoS}}$ and $\underline{e}_{\mathsf{NLoS}}$ are modeled as Gaussian random variables, and $\underline{\nabla}$ is an Exponential random variable, with mutual independence, known means and variances. Then, given a measurement y, the PDF of $\underline{L}$ under LoS and NLoS conditions can be derived as:

$$f_{\underline{L}}(L|y,\mathsf{LoS}) = f_{\underline{S}'_{\mathsf{LoS}}}(L - A'_{\mathsf{LoS}}|y,\mathsf{LoS}) \quad (6.26)$$

$$f_{\underline{L}}(L|y,\mathsf{NLoS}) = f_{\underline{S}'_{\mathsf{NLoS}}}(L - A'_{\mathsf{NLoS}}|y,\mathsf{NLoS}) \quad (6.27)$$

and the detailed expressions can be found in [101].

The final decision is then made as:

$$\frac{f_{\underline{L}}(L|y,\mathsf{LoS})}{f_{\underline{L}}(L|y,\mathsf{NLoS})} \underset{\mathsf{NLoS}}{\overset{\mathsf{LoS}}{\gtrless}} \Theta_{TR}, \quad (6.28)$$

where Θ_{TR} is a threshold, depending on the preassigned probability of false alarm.

Please note that (6.26) and (6.27) are obtained with the simplification that the random variable $\underline{y}$ in (6.21) is replaced by one of its realization y (one measurement). A general way to deal with LoS/NLoS identification using combined ToA/RSS measurements would be:

$$\frac{f_{\underline{L},\underline{y}}(L,y|\mathsf{LoS})}{f_{\underline{L},\underline{y},\underline{\nabla}}(L,y,\nabla|\mathsf{NLoS})} \underset{\mathsf{NLoS}}{\overset{\mathsf{LoS}}{\gtrless}} \Theta_{TR}, \quad (6.29)$$

where $f_{\underline{L},\underline{y}}(L,y|\mathsf{LoS})$ is the joint PDF of $\underline{L}$ and $\underline{y}$ under LoS and $f_{\underline{L},\underline{y},\underline{\nabla}}(L,y,\nabla|\mathsf{NLoS})$ is the joint PDF of $\underline{L}$, $\underline{y}$ and $\underline{\nabla}$ under NLoS. Since both L and y are functions of d, where d is unknown, (6.29) can be further extended as:

$$\frac{\max\limits_{d} f_{\underline{L},\underline{y}}(L,y|\mathsf{LoS})}{\max\limits_{d} f_{\underline{L},\underline{y},\underline{\nabla}}(L,y,\nabla|\mathsf{NLoS})} \underset{\mathsf{NLoS}}{\overset{\mathsf{LoS}}{\gtrless}} \Theta_{TR}. \quad (6.30)$$

This point will be discussed in detail in Section 6.2.

Equation (6.28) provides a simple solution for LoS/NLoS identification. Later in Section 6.2, we will show that this scheme can be further simplified, and the corresponding results are still very promising.

6.1.3 Nonparametric Methods

Suppose that, within an area, the PDF of the range measurements $\underline{y}$, the PDF of the user positions $\underline{x}$ and their joint PDF $f_{\underline{y},\underline{x}}(y,x)$ are known, estimators such as a ML estimator:

$$\hat{\underline{x}} = \arg\max_{x} f_{\underline{y}}(\underline{y}|x), \quad (6.31)$$

a Maximum A Posterior (MAP) estimator:

$$\hat{\underline{x}} = \arg\max_{x} f_{\underline{x}|\underline{y}}(x|\underline{y}), \quad (6.32)$$

or the estimator adopted in [24]:

$$\begin{aligned}\hat{\underline{x}} &= \mathsf{E}\{\underline{x}|\underline{y}\} = \int_{-\infty}^{+\infty} x f_{\underline{x}|\underline{y}}(x|\underline{y})dx \\ &= \frac{\int_{-\infty}^{+\infty} x f_{\underline{y},\underline{x}}(\underline{y},x)dx}{\int_{-\infty}^{+\infty} f_{\underline{y},\underline{x}}(\underline{y},x)dx} \\ &= \frac{\int_{-\infty}^{+\infty} x f_{\underline{y}}(\underline{y}|x) f_{\underline{x}}(x)dx}{\int_{-\infty}^{+\infty} f_{\underline{y}}(\underline{y}|x) f_{\underline{x}}(x)dx},\end{aligned} \tag{6.33}$$

can be obtained. These three solutions solve the problems of LoS/NLoS identification, mitigation and position estimation in one go, since the statistics of LoS and NLoS measurements and the positions within the area are all characterized by the PDFs.

The PDFs of $\underline{y}$ and x usually can be obtained or approximated in practice. The difficulty lies in determining the joint PDF. This joint PDF can be estimated in parametric or nonparametric ways. Parametric estimation is done by first choosing a PDF model for $f_{\underline{y},\underline{x}}(y,x)$, and then estimate the required parameters that define the model, based on a set of survey measurements. However, obtaining a proper parametric PDF model for $f_{\underline{y},\underline{x}}(y,x)$ is generally a nontrivial task, especially for NLoS cases.

In [24], the nonparametric way is chosen. Instead of choosing a predefined PDF model, the nonparametric estimation is done by approximating the PDF that created the given data as a sum of kernel functions. Each kernel function is centered at a single survey measurement, and its value decreases monotonically as the distance from the survey point increases. The expression for $f_{\underline{y},\underline{x}}(y,x)$ is given in [24] as:

$$\hat{f}_{\underline{y},\underline{x}}(y,x) = \frac{1}{N}\sum_{j=1}^{N} {h_y}^{-m_j} h_x^{-n} K_y(\frac{y-y_j}{h_y}) K_x(\frac{x-x_j}{h_x}), \tag{6.34}$$

where N is the number of survey points, $\underline{y}_j$ and x_j denotes the measurement vector and the position vector at the survey point j, m_j and n are the lengths of $\underline{y}_j$ and x_j, respectively. The smoothing constants h_y and h_x determine the width of the kernel functions K_y and K_x. The choice for these parameters has been proposed in [24]. With a good selection of kernel functions, this technique can create good approximations of the PDF.

Once the joint PDF is estimated, given a measurement vector y, the user position can be estimated using ML, MAP estimation or based on (6.33). For example, given a measurement set y and a PDF $f_{\underline{x}}(x)$ (uniform in [24]) of x over the region of interest, the ML estimator is obtained as:

$$\hat{\underline{x}} = \arg\max_x \frac{\hat{f}_{\underline{y},\underline{x}}(y,x)}{f_{\underline{x}}(x)}. \tag{6.35}$$

Given the fact that a kernel function $K_y(\frac{y-y_j}{h_y})$ is large at $y = y_j$ and is monotonically decreasing as y gets farther away from y_j, then for a y close to y_j, the estimate $\hat{x}$ in (6.35) will be close to x_j, since $\hat{f}_{\underline{y},\underline{x}}(y,x)$ can be simplified as:

$$\hat{f}_{\underline{y},\underline{x}}(y,x) = \frac{1}{N} {h_y}^{-m_j} h_x^{-n} K_y(\frac{y-y_j}{h_y}) K_x(\frac{x-x_j}{h_x}), \tag{6.36}$$

which is large at x close to x_j.

Moreover, given a uniform $f_{\underline{x}}(x)$, the solution in (6.33) is equivalent to a weighted average of the survey point positions with the weights being determined by the measurement data [24]:

$$\hat{x} = \sum_{j=1}^{N} x_j w_j(y), \tag{6.37}$$

where

$$w_j(y) = \frac{K_y(\frac{y-y_j}{h_y})}{\sum_{k=1}^{N} K_y(\frac{y-y_k}{h_y})}. \tag{6.38}$$

The performance of the nonparametric location estimators is dependent on

- the selection of the smoothing constants and the kernel functions,
- the number of survey points,
- the distribution of the survey point locations.

Details about how to correctly choose these parameters such that the performance and the survey complexity is well balanced, can be found in [24].

6.1.4 Least-Squares Residuals

In [102], a LoS/NLoS identification and mitigation scheme based on least-squares residual has been proposed. Variations have been reported later in [103, 104].

Recall from (3.50), under LoS condition, the unknown vector x is estimated from the measurements in $\underline{y}$ by

$$\hat{x} = \arg\min_{x} \|y - A(x)\|_W^2. \tag{6.39}$$

The key idea is that, with a set of LoS measurements, the least-squares residuals should be smaller than in the case where NLoS measurements are present. This is due to the fact that

1. LoS measurements have better quality than NLoS measurements, i.e. smaller biases and variance, see Section 6.1.2.1,
2. The mapping $A(x)$ in (6.39) does not take into account the biases in NLoS measurements. Hence, the un-modeled NLoS biases will propagate into the final estimate as errors, and lead to larger residuals.

The full procedure consists of three steps in [102],

1. choose different subsets of the m measurements, with the number of measurements in each set larger than the number of unknown parameters, n, and the total number of subsets would be

$$N = \sum_{k=n}^{m} \begin{pmatrix} m \\ k \end{pmatrix} \tag{6.40}$$

2. with the k-th subset $\underline{y}_k$, calculate an intermediate solution $\hat{x}_k$, and a corresponding weight $w_k = \frac{\|y_k - A(\hat{x}_k)\|^2_{W_k}}{\text{dimension of } y_k}$, where W_k is the weight matrix corresponds to subset k,

3. calculate the final estimate as the weighted linear combination of the intermediate estimates from Step 2:

$$\hat{x} = \frac{\sum_{k=1}^{N} \hat{x}_k w_k^{-1}}{\sum_{k=1}^{N} w_k^{-1}}. \tag{6.41}$$

Please note that (6.41) is a weighted mean of the intermediate estimates obtained using N measurement subsets, and the idea can be regarded as dating back to Mayer's method [105], which had been one of the early possible ways to handle redundant measurements before least-squares was actually introduced. The difference is that in [105] the measurement subsets are not overlapping, while in [102], some subsets are overlapping with each other.

Equation (6.41) provides a simple solution of NLoS identification/mitigation when the PDFs of LoS and NLoS measurements are unknown, however, it brings little improvement if the majority of the measurements are NLoS. The cases where the PDFs of the LoS and NLoS measurements are available will be addressed in the next section.

At the end of this chapter, the pros and cons of the reviewed methods will be summarized and compared with the tests we developed, which are introduced next.

6.2 ToA and RSS Combined

In this section, we introduce a scheme that can complete LoS/NLoS identification and mitigation. The scheme works with ToA range and RSS measurements, and the motivation is the following:

1. As described in Chapter 2, UWB signals offer great time resolution and the potential to have (at least) some frequency components penetrate through or around obstacles. Ranging based on timing information is considered one of the best ways to exploit the fine time resolution provided by UWB signals [2, 21]. Additionally, RSS measurements can be obtained readily in a timing based system.

2. The pass loss models of RSS measurements under LoS and NLoS are well separated, see Figure 6.3, which indicates a good possibility of LoS/NLoS identification.

The scheme, working with both ToA range and RSS measurements, is introduced next, starting with the models and assumptions.

6.2.1 ToA/RSS Models and Assumptions

Considering m links, the models for the ToA range measurements, $\underline{y} \in \mathbb{R}^m$, and the RSS measurements, $\underline{P} \in \mathbb{R}^m$, are given as the following:

$$\underline{y}_{\mathsf{LoS},i} = A_i(x_u) + \underline{e}_{\mathsf{LoS},i} \tag{6.42}$$

$$\underline{y}_{\mathsf{NLoS},i} = A_i(x_u) + \mu_i + \underline{e}_{\mathsf{NLoS},i} \tag{6.43}$$

$$\underline{\hat{P}}_{\mathsf{LoS},i} = P_0 + 10n_{\mathsf{LoS}} \log_{10} \frac{A_i(x_u)}{d_0} + \underline{s}_{\mathsf{LoS},i} \tag{6.44}$$

$$\underline{\hat{P}}_{\mathsf{NLoS},i} = P_0 + 10n_{\mathsf{NLoS}} \log_{10} \frac{A_i(x_u)}{d_0} + \underline{s}_{\mathsf{NLoS},i} \tag{6.45}$$

where $A_i(x_u) = d_i = \sqrt{(x_u - x_i)^T(x_u - x_i)}$ is the true distance, with x_u the unknown user position of interest, x_i the known position of the i-th transmitter, $\underline{e}_{(.)}$ is the error in ToA range measurements, μ_i is the unknown NLoS bias, P_0 is the reference power collected at the reference distance d_0, $n_{(.)}$ and $\underline{s}_{(.)}$ are the pathloss exponent and the shadowing effect, respectively. The pathloss exponents need to be obtained empirically. Note that with RSS in the NLoS case, no extra range bias is considered as compared to ToA cases. The differences compared to the LoS case include a different pathloss exponent and different shadowing effect.

The general assumptions are:

1. Gaussian distribution for the measurement error: $\underline{e}_{\mathsf{LoS},i} \sim \mathcal{N}(0, \sigma^2_{\mathsf{ToA,LoS}})$, $\underline{e}_{\mathsf{NLoS},i} \sim N(0, \sigma^2_{\mathsf{ToA,NLoS}})$, $\underline{s}_{\mathsf{LoS},i} \sim N(0, \sigma^2_{\mathsf{RSS,LoS}})$, and $\underline{s}_{\mathsf{NLoS},i} \sim \mathcal{N}(0, \sigma^2_{\mathsf{RSS,NLoS}})$. All the variances are empirically obtained from a measurement campaign.

2. the ToA and RSS measurements for different links are independent.

3. the probability of condition for a certain link is set to 1/2 for LoS and to 1/2 for NLoS, i.e. P(LoS)=P(NLoS). This is quite realistic for indoor environments. Please note that the later introduced NLoS identification/mitigation schemes are not limited by this assumption and can be easily adapted if more accurate probabilities can be obtained, e.g. via measurement campaign.

4. the ToA and RSS measurements for the same link are correlated, but the correlation is empirically determined from survey data. Denote $\sigma_{\mathsf{TR,LoS}}$ and $\sigma_{\mathsf{TR,NLoS}}$ as the covariances of a ToA and a RSS measurements in LoS and NLoS situations, respectively.

5. with the previous assumptions, the PDF $f_{\underline{y}_{\mathsf{all}}}$ of the vector $\underline{y}_{\mathsf{all}} = [\underline{y}^T \ \underline{P}^T]^T_{2m\times 1}$, containing m ToA measurements and m RSS measurements, is also Normal, with known mean and variance matrix.

The following notation is used in the remainder of this chapter:

$\underline{S}_{m\times 1}$	the random state vector, $P(\underline{S} = S_k) = \frac{1}{2^m}$, with $k = 0, 1, ..., 2^m - 1$
$[S_k]_{m\times 1}$	possible realizations of the state vector, with 0 entries representing LoS links and 1 entries representing NLoS ones
C_k	the state matrix of size $2m \times q_k$, with q_k being the number of NLoS links
∇_k	the NLoS bias vector

Please note that $P(\underline{S} = S_k) = \frac{1}{2^m}$ is based on 1) $P(\mathsf{LoS}) = P(\mathsf{NLoS}) = 1/2$ is assumed for each link, and 2) different links are independent.

Given a sufficient number of ToA and RSS measurements, with the models and the assumptions listed above, it is possible to perform a classification. The details are described in the next section, and several possible simplified versions are elaborated on in the later sections.

6.2.2 Classification Type I

With m transmitters and 1 receiver, there are m links with m ToA range measurements and m RSS measurements. Each link can be under a LoS or a NLoS condition, hence, there are 2^m different states or possibilities. The problem of deciding between p states or possibilities is generally referred to as *classification* [89].

The following hypotheses are made:

$$
\begin{aligned}
H_0 \ : \ & \text{all LoS, no } \nabla \text{ exists,} \\
& S_0 = [0, \cdots, 0]^T, q_0 = 0, C_0 = 0, \nabla_0 = 0, \\
& f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_0) = (\det(2\pi Q_0))^{-1/2} \exp\{-\frac{1}{2}\|y_{\mathsf{all}} - F_0(x_u)\|^2_{Q_0^{-1}}\}, \\
& F_0(x_u) = \begin{bmatrix} A_1(x_u) \\ \vdots \\ A_m(x_u) \\ P_0 + 10 n_{\mathsf{LoS}} \log_{10} \frac{A_1(x_u)}{d_0} \\ \vdots \\ P_0 + 10 n_{\mathsf{LoS}} \log_{10} \frac{A_m(x_u)}{d_0} \end{bmatrix}_{2m\times 1}
\end{aligned}
\tag{6.46}
$$

$$H_1 \;:\; m-1 \text{ LoS, } 1 \text{ NLoS, NLoS bias } \nabla_1 = [\mu_m] \text{ in } \underline{y}_m \tag{6.47}$$

$$S_1 = [0, \cdots, 0, 1]^T, q_1 = 1, C_1 = [0, \cdots, 0, 1, 0, ..., 0]^T_{2m\times 1}, \nabla_1 = [\mu_m],$$

$$f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_1) = (\det(2\pi Q_1))^{-1/2} \exp\{-\frac{1}{2}\|y_{\mathsf{all}} - F_1(x_u, C_1\nabla_1)\|^2_{Q_1^{-1}}\},$$

$$F_1(x_u, \nabla_1) = \begin{bmatrix} A_1(x_u) \\ \vdots \\ A_{m-1}(x_u) \\ A_m(x_u) + \mu_m \\ P_0 + 10n_{\mathsf{LoS}} \log_{10} \frac{A_1(x_u)}{d_0} \\ \vdots \\ P_0 + 10n_{\mathsf{NLoS}} \log_{10} \frac{A_m(x_u)}{d_0} \end{bmatrix}_{2m\times 1}$$

$$\vdots$$

$$H_{2^m-1} \;:\; \text{all NLoS,} \tag{6.48}$$

$$S_{2^m-1} = [1, \cdots, 1]^T, q_{2^m-1} = m, C_{2^m-1} = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}_{2m\times m},$$

$$\nabla_{2^m-1} = [\mu_1, ..., \mu_m]^T,$$

$$f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_{2^m-1})$$

$$= (\det(2\pi Q_{2^m-1}))^{-1/2} \exp\{-\frac{1}{2}\|y_{\mathsf{all}} - F_{2^m-1}(x_u, C_{2^m-1}\nabla_{2^m-1})\|^2_{Q^{-1}_{2^m-1}}\},$$

$$F_{2^m-1}(x_u, C_{2^m-1}\nabla_{2^m-1}) = \begin{bmatrix} A_1(x_u) + \mu_1 \\ \vdots \\ A_m(x_u) + \mu_{2^m-1} \\ P_0 + 10n_{\mathsf{NLoS}} \log_{10} \frac{A_1(x_u)}{d_0} \\ \vdots \\ P_0 + 10n_{\mathsf{NLoS}} \log_{10} \frac{A_m(x_u)}{d_0} \end{bmatrix}_{2m\times 1}$$

The expression of the $2m \times 2m$ variance matrix Q_k for state S_k is calculated as

$$Q_k = \begin{bmatrix} \sigma^2_{\mathsf{ToA,LoS}} T_{\mathsf{LoS}} + \sigma_{\mathsf{ToA,NLoS}} T_{\mathsf{NLoS}} & \sigma_{\mathsf{TR,LoS}} T_{\mathsf{LoS}} + \sigma_{\mathsf{TR,NLoS}} T_{\mathsf{NLoS}} \\ \sigma_{\mathsf{TR,LoS}} T_{\mathsf{LoS}} + \sigma_{\mathsf{TR,NLoS}} T_{\mathsf{NLoS}} & \sigma^2_{\mathsf{RSS,LoS}} T_{\mathsf{LoS}} + \sigma_{\mathsf{RSS,NLoS}} T_{\mathsf{NLoS}} \end{bmatrix} \tag{6.49}$$

where T_{LoS} and T_{NLoS} are diagonal matrices with diagonal entries equal to $|S_k - 1|$ and S_k, respectively. The operator $|.|$ here denotes an absolute value.

Given the multiple hypotheses, the final decision is made based on multiple GLRT [89]:

$$\begin{aligned} \text{accept } H_k \text{ if } & \max_{x_{u,j},\nabla_j} \left\{ \max_{[x_{u,j},\nabla_j]\in\Phi_j} f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_j) \right\} \\ = & \max_{[x_{u,k},\nabla_k]\in\Phi_k} f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_k) \end{aligned} \tag{6.50}$$

where

$$\mathsf{GL}_j = \max_{[x_{u,j},\nabla_j]\in\Phi_j} f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_j) \tag{6.51}$$

is the GL. The GL_j implies maximization of the likelihood within the parameter space Φ_j, put forward by the j-th hypothesis H_j. Under hypothesis H_0, we have $\Phi_0 = \{x_{u,j} \in \mathbb{R}^n, \nabla = 0\}$.

The steps of Classification Type I can be summarized in a similar way as in [79] as follows:

1. For each of the 2^m states, calculate the position estimate $\hat{x}_u$ and $\hat{\nabla}$ that maximizes the corresponding likelihood function. Since the PDF of ToA and RSS measurements are all (assumed) Gaussian, the estimates can be obtained via least-squares, see Chapter 3,

2. based on the estimates obtained in the previous step to evaluate the corresponding GL using (6.51),

3. choose the final estimates as the ones that correspond to the maximum GL of all the states, see (6.50). The LoS/NLoS identification results can be found in the chosen state vector S_k, with 0 representing LoS and 1 representing NLoS.

One issue of using (6.50), where the largest GL is picked, needs to be pointed out. In an extreme case, when the number of unknown parameters in H_{2^m-1} is equal to the number of measurements, the corresponding least squares residual is equal to 0. Hence the absolute minimum value for $\|y_{\mathsf{all}} - F_{2^m-1}(x_u, C_{2^m-1}\nabla_{2^m-1})\|^2_{Q^{-1}_{2^m-1}}$, which is 0. In this case, (6.50) will always select the H_{2^m-1}, since it chooses the most likely one, no matter how likely it is. Although we haven not considered the effect of the Q-matrix (different per hypothesis), there is a tendency that the hypotheses with large number of NLoS biases are preferred, due to the above reasoning, even when the above extreme case is not present, and the models are very much relaxed with the number of measurements larger than the number of unknown parameters.

However, although this tendency will reduce the rate of correct decisions on the state, the effect on final positioning accuracy should be subtle. This is because, treating LoS measurements as NLoS only introduces more parameters to estimate, which should be close to zero in the final outcomes. The effect of the above tendency can then be described as:

1. the estimator is still unbiased,

2. the variance of the estimator will be (a bit) larger with more unknown parameters.

Table 6.1: Decision matrix of classification among 2^m hypothesis [89].

	reality				
decision	H_0 true	H_1 true	H_2 true	$\cdots$	H_{2^m-1} true
accept H_0	OK	$\beta_{1,0}$	$\beta_{2,0}$	$\cdots$	$\beta_{2^m-1,0}$
accept H_1	$\beta_{0,1}$	OK	$\beta_{2,1}$	$\cdots$	$\beta_{2^m-1,0}$
accept H_2	$\beta_{0,2}$	$\beta_{1,2}$	OK	$\cdots$	$\beta_{2^m-1,2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
accept H_{2^m-1}	$\beta_{0,2^m-1}$	$\beta_{1,2^m-1}$	$\beta_{2,2^m-1}$	$\cdots$	OK

6.2.2.1 Probability of Incorrect and Correct Decisions

The decision matrix of classification among 2^m hypothesis is shown in Table 6.1. The incorrect decisions are denoted by their probabilities $\beta_{k,j}$, with the subscript (k, j) standing for H_k true but decision made as H_j. The sum of β's per column in Table 6.1:

$$\beta_k = \sum_{j=0\ j\neq k}^{2^m-1} \beta_{k,j}, \tag{6.52}$$

gives the probability of incorrect decision when H_k is true [89]. The power of H_k, the probability of correct decision when H_k is true, simply follows as $\gamma_k = 1 - \beta_k$.

With multiple hypotheses, the calculation of power γ_k is usually easier than β_k, since the latter needs to calculate $\beta_{k,j}$ for all $j \neq k$. This can be shown based on an example in Figure 6.4 with $m = 2$ and $m^2 = 4$ hypotheses. Without loss of generality, we calculate

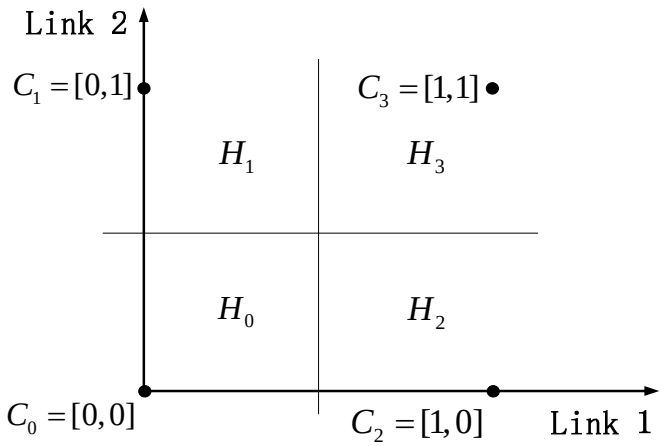

Figure 6.4: Decision regions for four hypotheses with two links (two transmitters and one receiver) in $\mathbb{R}^2$.

β_k and γ_k for $k = 3$, and the expressions for both terms with $k = 0, 1, 2$ can be given in a similar way.

Based on (6.52), β_3 can be given as:

$$\begin{aligned}\beta_3 &= \beta_{3,0}+\beta_{3,1}+\beta_{3,2}\\ &= P\left(\text{Link 1}=0,\text{Link 2}=0|\text{Link 1}=1,\text{Link 2}=1\right)\\ &+ P\left(\text{Link 1}=0,\text{Link 2}=1|\text{Link 1}=1,\text{Link 2}=1\right)\\ &+ P\left(\text{Link 1}=1,\text{Link 2}=0|\text{Link 1}=1,\text{Link 2}=1\right)\end{aligned} \tag{6.53}$$

Since the ToA range and RSS measurements of different links are *independent*, we have

$$\begin{aligned}\beta_3 &= P\left(\text{Link 1}=0|\text{Link 1}=1,\text{Link 2}=1\right)P\left(\text{Link 2}=0|\text{Link 1}=1,\text{Link 2}=1\right)\\ &+ P\left(\text{Link 1}=0|\text{Link 1}=1,\text{Link 2}=1\right)P\left(\text{Link 2}=1|\text{Link 1}=1,\text{Link 2}=1\right)\\ &+ P\left(\text{Link 1}=1|\text{Link 1}=1,\text{Link 2}=1\right)P\left(\text{Link 2}=0|\text{Link 1}=1,\text{Link 2}=1\right)\end{aligned} \tag{6.54}$$

First, we assume that x_u, μ_1 and μ_2 are all known, (6.54) can be written in detail as:

$$\begin{aligned}&\beta_3\\ =\ & P\left(\text{chose } H_0 \text{ over } H_1|\text{given } H_3 \text{ true}\right)P\left(\text{ chose } H_0 \text{ over } H_2|\text{given } H_3 \text{ true}\right)\\ +\ & P\left(\text{chose } H_1 \text{ over } H_0|\text{given } H_3 \text{ true}\right)P\left(\text{ chose } H_1 \text{ over } H_3|\text{given } H_3 \text{ true}\right)\\ +\ & P\left(\text{chose } H_2 \text{ over } H_0|\text{given } H_3 \text{ true}\right)P\left(\text{ chose } H_2 \text{ over } H_3|\text{given } H_3 \text{ true}\right),\\ =\ & P\left(f_{\underline{y}_{\text{all}}|\underline{S}}(y_{\text{all}}|S_0)>f_{\underline{y}_{\text{all}}|\underline{S}}(y_{\text{all}}|S_1)\right)P\left(f_{\underline{y}_{\text{all}}|\underline{S}}(y_{\text{all}}|S_0)>f_{\underline{y}_{\text{all}}|\underline{S}}(y_{\text{all}}|S_2)\right)\\ +\ & P\left(f_{\underline{y}_{\text{all}}|\underline{S}}(y_{\text{all}}|S_1)>f_{\underline{y}_{\text{all}}|\underline{S}}(y_{\text{all}}|S_0)\right)P\left(f_{\underline{y}_{\text{all}}|\underline{S}}(y_{\text{all}}|S_1)>f_{\underline{y}_{\text{all}}|\underline{S}}(y_{\text{all}}|S_3)\right)\\ +\ & P\left(f_{\underline{y}_{\text{all}}|\underline{S}}(y_{\text{all}}|S_2)>f_{\underline{y}_{\text{all}}|\underline{S}}(y_{\text{all}}|S_0)\right)P\left(f_{\underline{y}_{\text{all}}|\underline{S}}(y_{\text{all}}|S_2)>f_{\underline{y}_{\text{all}}|\underline{S}}(y_{\text{all}}|S_3)\right).\end{aligned} \tag{6.55}$$

where the expression for $f_{\underline{y}_{\text{all}}|\underline{S}}(y_{\text{all}}|S_k)$ can be found in (6.46)-(6.48).

The calculation for the probability correct decision γ_3, on the other hand, is much simpler, and can be done as:

$$\gamma_3 = P\left(f_{\underline{y}_{\text{all}}|\underline{S}}(y_{\text{all}}|S_3)>f_{\underline{y}_{\text{all}}|\underline{S}}(y_{\text{all}}|S_1)\right)P\left(f_{\underline{y}_{\text{all}}|\underline{S}}(y_{\text{all}}|S_3)>f_{\underline{y}_{\text{all}}|\underline{S}}(y_{\text{all}}|S_2)\right). \tag{6.56}$$

The important conclusion, drawn based on (6.56) with $m=2$, is that γ_k is equal to the product of *two* probabilities that the likelihood of H_k is larger than the likelihoods of its two 'neighbor' hypotheses, where a 'neighbor' hypothesis of H_k, say H_j, is defined as the hypothesis whose state vector S_j has the unit euclidean distance to S_k,

$$\|S_j - S_k\| = 1, \text{ for } j=1,...,2^m-1, j\neq k. \tag{6.57}$$

This conclusion can be easily extended to a case with any m, that:

$$\gamma_k = \prod_{j\in\{j\,|\,\|S_j-S_k\|=1\}} P\left(f_{\underline{y}_{\text{all}}|\underline{S}}(y_{\text{all}}|S_k)>f_{\underline{y}_{\text{all}}|\underline{S}}(y_{\text{all}}|S_j)\right) \tag{6.58}$$

In our case, since x_u and ∇_k's are unknown, we will replace them with the corresponding estimates, i.e. replace the likelihoods in (6.58) with the corresponding GL's:

$$\gamma_k = \prod_{j\in\{j\,|\,\|S_j-S_k\|=1\}} P\left(\max_{[x_{u,k},\nabla_k]\in\Phi_k} f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_k) > \max_{[x_{u,j},\nabla_j]\in\Phi_j} f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_j)\right) \quad (6.59)$$

To this end, it is quite clear that the power γ_k of H_k is much easier to calculate than β_k, the probability incorrect decision when H_k is true. In the rest of this section, we will first calculate γ_k based on (6.59) for Classification Type I, and then β_k can be readily calculated as $\beta_k = 1-\gamma_k$.

Derivation for γ_k The expression for γ_k, the probability of correct decision when H_k is true, is performed as follows. Concerning a hypothesis H_j, with $j = 0, ..., 2^m-1$, we have:

$$\begin{aligned} & \max_{[x_{u,j},\nabla_j]\in\Phi_j} f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_j) \quad (6.60)\\ = & \max_{[x_{u,j},\nabla_j]\in\Phi_j} (\det(2\pi Q_j))^{-1/2}\exp\{-\frac{1}{2}\|y_{\mathsf{all}}-F_j(x_u,C_j\nabla_j)\|^2_{Q_j^{-1}}\}\\ = & \max_{[x_{u,j},\nabla_j]\in\Phi_j}\left\{-m\ln(2\pi)-\frac{1}{2}\ln(\det(Q_j))-\frac{1}{2}\|y_{\mathsf{all}}-F_j(x_u,C_j\nabla_j)\|^2_{Q_j^{-1}}\right\}\end{aligned}$$

The maximizer $\hat{x}_{u,j}, \hat{\nabla}_j$ of the likelihood function can be found via least-squares estimation. Denoting the corresponding least-squares solution of the measurements as $\hat{y}_{\mathsf{all},j} = F_j(\hat{x}_{u,j}, C_j\hat{\nabla}_j)$, the residual then can be expressed as:

$$\hat{e}_j = y_{\mathsf{all}} - \hat{y}_{\mathsf{all},j} = y_{\mathsf{all}} - F_j(\hat{x}_{u,j}, C_j\hat{\nabla}_j). \quad (6.61)$$

A correct decision between H_j and H_k, with H_k true and $k\neq j$, is then equivalent to

$$\begin{aligned} \text{accept } H_k \text{ if} \quad & f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_k) > f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_j)\\ \Leftrightarrow \quad & T_{k,j} = \hat{e}_j^T Q_j^{-1}\hat{e}_j - \hat{e}_k^T Q_k^{-1}\hat{e}_k > k_{k,j}, \end{aligned} \quad (6.62)$$

where $k_{k,j} = \ln(\frac{\det(Q_j)}{\det(Q_k)}$.

The statistics of $T_{k,j}$ are quite difficult to evaluate, given the fact that the function $F(.)$ in (6.61) is non-linear, and it is more convenient to approximate the non-linear part with its linearized version. For any $j = 0, ..., 2^m-1$, the least-squares estimate vector corresponding H_j, $\hat{x}_{u,j}, \hat{\nabla}_j$, and the least-squares solution of the measurements $\hat{y}_{\mathsf{all},j}$ can be approximated according to [89]. Assuming that the second- and higher- order terms in $\hat{x}_{u,j}, \hat{\nabla}_j$ and $\hat{y}_{\mathsf{all},j}$ are negligible, their approximations, can be given as:

$$[(\hat{x}_{u,j}-x_u)^T, \hat{\nabla}_j]^T \approx (A_j^T Q_j^{-1} A_j)^{-1} A_j^T Q_j^{-1}(y_{\mathsf{all}} - F_j(x_u)) \quad (6.63)$$

$$\hat{y}_{\mathsf{all},j} - F_j(x_u) \approx A_j(A_j^T Q_j^{-1} A_j)^{-1} A_j^T Q_j^{-1}(y_{\mathsf{all}} - F_j(x_u)) \quad (6.64)$$

where the notation $A_j = [\partial_{x_u^T} F_j(x_u), C_j]$ is used, x_u is the true value of the unknown position, and $F_j(x_u) = F_j(x_u, C_0\nabla_0)$ with $C_0 = 0$ and $\nabla_0 = 0$.

Moreover, (6.61) can be approximated as:

$$\hat{e}_j = y_{\text{all}} - \hat{y}_{\text{all},j} \approx P^{\perp}_{A_j,Q_j}(y_{\text{all}} - F_j(x_u)), \tag{6.65}$$

with $P_{A_j,Q_j} = A_j(A_j^T Q_j^{-1} A_j)^{-1} A_j^T Q_j^{-1}$ and $P^{\perp}_{A_j,Q_j} = I - A_j(A_j^T Q_j^{-1} A_j)^{-1} A_j^T Q_j^{-1}$.

Given the general assumption that $\underline{y}_{\text{all}} \sim \mathcal{N}(F(x_u) + C\nabla, Q_{\text{all}})$, with $F(.)$, C, ∇ and Q_{all} the correct function, state matrix, NLoS bias vector and measurement variance matrix, the distribution of $\underline{\hat{e}}_j$ can be derived based on (6.65) as:

$$\underline{\hat{e}}_j \sim \mathcal{N}(P^{\perp}_{A_j,Q_j}(F(x_u) - F_j(x_u) + C\nabla), P^{\perp}_{A_j,Q_j} Q^{-1}_{\text{all}} (P^{\perp}_{A_j,Q_j})^T). \tag{6.66}$$

Based on (6.66), the distribution of $\underline{T}_j$ in (6.62) can be given using the theorem in [106]:

Theorem 6.1 *Let the $b \times 1$ vector $\underline{t}$ be Normally distributed with mean $E\{\underline{t}\} = t$ and positive definite covariance matrix Q_t. Let B be a symmetric non-negative definite matrix of order b. Then there exists a positive-definite diagonal matrix $\Lambda_r = \text{diag}(\lambda_1, ..., \lambda_r)$ and a vector $u \in \mathbb{R}^r$, such that*

$$\underline{t}^T B \underline{t} = (\underline{z} + u)^T \Lambda_r (\underline{z} + u) = \sum_{i=1}^{r} \lambda_i (\underline{z}_i + u_i)^2 \tag{6.67}$$

where $\underline{z}$ has the standard Normal distribution, i.e. $\underline{z} \sim \mathcal{N}(0, I_r)$. The number r is the rank of BQ_t or Q_tB. The diagonal elements of Λ_r are the r positive eigenvalues of BQ_t or Q_tB. And if $U_r \Lambda_r U_r^T$ is the singular value decomposition of $Q_t^{1/2} B Q_t^{1/2}$, i.e., $Q_t^{1/2} B Q_t^{1/2} = U_r \Lambda_r U_r^T$, with $Q_t^{1/2}$ a square-root of Q_t, i.e., $Q_t = Q_t^{1/2} Q_t^{1/2}$, then the $r \times 1$ vector u can be computed as

$$u = U_r^T Q_t^{1/2} t. \tag{6.68}$$

The above theorem says that $\underline{t}^T A \underline{t}$ is equal to a linear combination of r independent non-central Chi-square distribution with 1 degree of freedom and non-centrality parameters u_i^2, $i = 1, ..., r$, i.e.,

$$\underline{t}^T A \underline{t} = \sum_{i=1}^{r} \lambda_i \underline{\chi}_i, \tag{6.69}$$

where $\underline{\chi}_i \sim \chi^2(1, u_i^2)$.

The distribution of $\hat{e}_j^T Q_j^{-1} \hat{e}_j$ can be given based on the above theorem:

$$\hat{e}_j^T Q_j^{-1} \hat{e}_j = \sum_{i=1}^{r_j} \lambda_{j,i} \underline{\chi}_{j,i}, \tag{6.70}$$

$$\underline{\chi}_{j,i} \sim \chi^2(1, u_{j,i}^2) \tag{6.71}$$

where r_j denotes the rank of $Q_j^{-1} P^{\perp}_{A_j,Q_j} Q^{-1}_{\text{all}} (P^{\perp}_{A_j,Q_j})^T$, and $\lambda_{j,i}$ is the i-th diagonal entry of Λ_j, which contains the r_j eigenvalues of $Q_j^{-1} P^{\perp}_{A_j,Q_j} Q^{-1}_{\text{all}} (P^{\perp}_{A_j,Q_j})^T$. Furthermore,

$u_{j,i} = U_j^T \left[P^{\perp}_{A_j,Q_j} Q_{\text{all}}^{-1} (P^{\perp}_{A_j,Q_j})^T\right]^{1/2} P^{\perp}_{A_j,Q_j} C\nabla$, where U_j can be obtained via singular value decomposition:

$$U_j \Lambda_{r_j} U_j^T = \left[P^{\perp}_{A_j,Q_j} Q_{\text{all}}^{-1} (P^{\perp}_{A_j,Q_j})^T\right]^{1/2} Q_j^{-1} \left[P^{\perp}_{A_j,Q_j} Q_{\text{all}}^{-1} (P^{\perp}_{A_j,Q_j})^T\right]^{1/2}.$$

And similarly for $\hat{e}_k^T Q_k^{-1} \hat{e}_k$:

$$\hat{e}_k^T Q_k^{-1} \hat{e}_k = \sum_{i=1}^{r_k} \lambda_{k,i} \underline{\chi}_{k,i}, \tag{6.72}$$

$$\underline{\chi}_{k,i} \sim \chi^2(1, u_{k,i}^2) \tag{6.73}$$

where r_k denotes the rank of $Q_k^{-1} P^{\perp}_{A_k,Q_k} Q_{\text{all}}^{-1} (P^{\perp}_{A_k,Q_k})^T$, and $\lambda_{k,i}$ is the i-th diagonal entry of Λ_k, which contains the r_k eigenvalues of $Q_k^{-1} P^{\perp}_{A_k,Q_k} Q_{\text{all}}^{-1} (P^{\perp}_{A_k,Q_k})^T$. Furthermore, $u_{k,i} = U_k^T \left[P^{\perp}_{A_k,Q_k} Q_{\text{all}}^{-1} (P^{\perp}_{A_k,Q_k})^T\right]^{1/2} P^{\perp}_{A_k,Q_k} C\nabla$, where U_k can be obtained via singular value decomposition:

$$U_k \Lambda_{r_k} U_k^T = \left[P^{\perp}_{A_k,Q_k} Q_{\text{all}}^{-1} (P^{\perp}_{A_k,Q_k})^T\right]^{1/2} Q_k^{-1} \left[P^{\perp}_{A_k,Q_k} Q_{\text{all}}^{-1} (P^{\perp}_{A_k,Q_k})^T\right]^{1/2}.$$

Combining (6.62), (6.70) and (6.72) we have:

$$\underline{T}_{k,j} = \sum_{i=1}^{r_j} \lambda_{j,i} \underline{\chi}_{j,i} - \sum_{i=1}^{r_k} \lambda_{k,i} \underline{\chi}_{k,i}. \tag{6.74}$$

Finally, the full distribution of $\underline{T}_{k,j}$ can be derived based on the joint PDF of the $r_j + r_k$ Chi-square distributed random variables based on (6.74), see [89]. If H_j is true, then the relevant parameters in (6.74) should be calculated by choosing $x_u = x_{u,j}$, $\nabla = \nabla_j$ and $Q_{\text{all}} = Q_j$, and if H_k is true, the calculation should be done by choosing $x_u = x_{u,k}$, $\nabla = \nabla_k$ and $Q_{\text{all}} = Q_k$. In practice, since the knowledge of the true values are not available, the corresponding estimates obtained under correct hypothesis will be used instead.

Without loss of generality, we assume that H_k is the correct state, then γ_k can be calculated based on (6.59) and (6.62) as:

$$\gamma_k = \prod_{j \in \{j \,|\, \|S_j - S_k\| = 1\}} P\left(T_{k,j} > k_{k,j}\right) \tag{6.75}$$

To this end, the probability of correct decision when H_k true, can be verified with the following steps:

1. calculate the mean values and variances of $\underline{\hat{e}}_j$ and $\underline{\hat{e}}_k$ based on (6.66), where the true position vector should be chosen as $x_u = x_{u,k}$, the state matrix C as C_k, the NLoS bias vector ∇ as ∇_k and the variance matrix Q_{all} as Q_k. In practice, the true values should be replaced with the corresponding estimates obtained under H_k.

2. calculate the relevant parameters, such as $u_{j,i}$, $\lambda_{j,i}$, $u_{k,i}$, $\lambda_{k,i}$, according to Theorem 6.1,

3. calculate the probability of correct decision based on (6.75).

Although the distribution of $T_{k,j}$ can be derived based on (6.74), it is still a difficult task which generally requires the aid of numerical tools. Actually, by examining alternative expression for $T_{k,j}$, an upper- and a lower- bounds for γ_k are derived in Appendix A.4.1, which can be easily evaluated with a standard table for Chi-square distribution after approximation.

The upper bound for the probability of correct decision when H_k is true is:

$$\begin{aligned}\gamma_k^{\mathsf{UB}} = \prod_{j\in\{j\,|\,\|S_j-S_k\|=1\}} & \left[a\,P^{\mathsf{UB}}\left(T_{k,j}^{\mathsf{UB}} > 0 | \underline{T}_{k,j}^{\mathsf{UB}} \sim \chi^2(2,\lambda^{\mathsf{UB}})\right)\right. \\ + & \left.(1-a)\,P^{\mathsf{UB}}\left(T_{k,j}^{\mathsf{UB}} < 0 | \underline{T}_{k,j}^{\mathsf{UB}} \sim \chi^2(2,0)\right)\right]\end{aligned} \tag{6.76}$$

and the lower bound:

$$\begin{aligned}\gamma_k^{\mathsf{LB}} = \prod_{j\in\{j\,|\,\|S_j-S_k\|=1\}} & \left[a\,P^{\mathsf{LB}}\left(T_{k,j}^{\mathsf{LB}} > 0 | \underline{T}_{k,j}^{\mathsf{LB}} \sim \chi^2(2,\lambda^{\mathsf{LB}})\right)\right. \\ + & \left.(1-a)\,P^{\mathsf{LB}}\left(T_{k,j}^{\mathsf{LB}} < 0 | \underline{T}_{k,j}^{\mathsf{LB}} \sim \chi^2(2,0)\right)\right]\end{aligned} \tag{6.77}$$

where the details of λ^{UB}, λ^{LB} and the controlling factor a can be found in Appendix A.4.1.

Classification Type I provides optimal LoS/NLoS identification/mitigation solution under the assumptions made at the beginning of this section. However, this method has an obvious drawback that the estimation of unknown parameters needs to be done for all 2^m hypotheses, which implies very demanding computational load with a large m. In the rest of this chapter, we will introduce three simplified versions of Classification Type I, aiming at not only reducing the computational load, but also keeping as much as possible the quality of the solution.

6.2.3 Classification Type II

The first simplified version of Classification Type I is performed by calculating ToA NLoS biases based on the model in [36]:

$$\mu_i = (0.027\exp(-B/0.47) + 0.013)d_i + 0.019, \tag{6.78}$$

where B in [GHz] is the signal bandwidth in use, and d_i in meter is the true distance. The so obtained bias parameters: μ, are assumed to be exact in the sequel. In practice, since d_i is not known, it is approximated with the ToA range measurement y_i.

The new hypotheses becomes:

$$
\begin{aligned}
H_0 \quad : \quad & \text{all LoS, no } \nabla \text{ exists,} \qquad (6.79)\\
& S_0 = [0, \cdots, 0]^T, q_0 = 0, C_0 = 0, \nabla_0 = 0,\\
& f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_0) = (\det(2\pi Q_0))^{-1/2} \exp\{-\frac{1}{2}\|y_{\mathsf{all}} - F_0(x_u)\|^2_{Q_0^{-1}}\},\\
& F_0(x_u) = \begin{bmatrix} A_1(x_u) \\ \vdots \\ A_m(x_u) \\ P_0 + 10 n_{\mathsf{LoS}} \log_{10} \frac{A_1(x_u)}{d_0} \\ \vdots \\ P_0 + 10 n_{\mathsf{LoS}} \log_{10} \frac{A_m(x_u)}{d_0} \end{bmatrix}_{2m\times 1}
\end{aligned}
$$

$$
\begin{aligned}
H_1 \quad : \quad & m-1 \text{ LoS, 1 NLoS, NLoS bias } \nabla_1 = [\mu_m] \text{ in } \underline{y}_m \qquad (6.80)\\
& S_1 = [0, \cdots, 0, 1]^T, q_1 = 1,\\
& C_1 = [0, \cdots, 0, 1, 0, \cdots, 0]^T_{2m\times 1}, \nabla_1 = [\mu_m],\\
& f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_1)\\
= \quad & (\det(2\pi Q_1))^{-1/2} \exp\{-\frac{1}{2}\|y_{\mathsf{all}} - C_1\nabla_1 - F_1(x_u)\|^2_{Q_1^{-1}}\},\\
& F_1(x_u) = \begin{bmatrix} A_1(x_u) \\ \vdots \\ A_m(x_u) \\ P_0 + 10 n_{\mathsf{LoS}} \log_{10} \frac{A_1(x_u)}{d_0} \\ \vdots \\ P_0 + 10 n_{\mathsf{NLoS}} \log_{10} \frac{A_m(x_u)}{d_0} \end{bmatrix}_{2m\times 1}\\
& \vdots\\
H_{2^m-1} \quad : \quad & \text{all NLoS,} \qquad (6.81)\\
& S_{2^m-1} = [1, \cdots, 1]^T, q_{2^m-1} = m, C_{2^m-1} = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}_{2m\times m},\\
& \nabla_{2^m-1} = [\mu_1, ..., \mu_m]^T,\\
& f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_{2^m-1})\\
= \quad & (\det(2\pi Q_{2^m-1}))^{-1/2} \exp\{-\frac{1}{2}\|y_{\mathsf{all}} - C_{2^m-1}\nabla_{2^m-1} - F_{2^m-1}(x_u)\|^2_{Q^{-1}_{2^m-1}}\},
\end{aligned}
$$

$$F_{2^m-1}(x_u) = \begin{bmatrix} A_1(x_u) \\ \vdots \\ A_m(x_u) \\ P_0 + 10n_{\mathsf{NLoS}} \log_{10} \frac{A_1(x_u)}{d_0} \\ \vdots \\ P_0 + 10n_{\mathsf{NLoS}} \log_{10} \frac{A_m(x_u)}{d_0} \end{bmatrix}_{2m\times 1}$$

The corresponding decision is made using MAP based on GL's, similar as in (6.50):

$$\text{accept } H_k \text{ if } \max_{x_{u,j}} \left\{ \max_{x_{u,j}} f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_j) \right\} = \max_{x_{u,k}} f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_k). \tag{6.82}$$

The steps of Classification Type II can be summarized as follows:

1. For each of the 2^m states, calculate the position estimate $\hat{x}_u$ that maximize the corresponding likelihood function. Again, least-squares theory can be applied for estimation,
2. based on the estimates obtained in the previous step to evaluate the corresponding GL $\max_{x_{u,j}} f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_j)$,
3. choose the final estimate for x_u as the one that corresponds to the maximum GL of all the states. The LoS/NLoS identification results can be found in the chosen state vector S_k, with 0 representing LoS and 1 representing NLoS.

6.2.3.1 Probability of Incorrect and Correct Decisions

The decision matrix of classification among 2^m hypothesis is the same as for Classification Type I, as shown in Table 6.1. The probability of correct decision when H_k is true can be given as:

$$\gamma_k = \prod_{j\in\{j \,|\, \|S_j - S_k\|=1\}} P\left(f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_k) > f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_j) \right) \tag{6.83}$$

where the NLoS biases are pre-calculated based on (6.78).

Derivation for γ_k In a similar way as we did for Classification Type I, the decision between H_j and H_k is then equivalent to

$$\begin{aligned} \text{accept } H_k \text{ if } \quad & f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_k) > f_{\underline{y}_{\mathsf{all}}|\underline{S}}(y_{\mathsf{all}}|S_j) \\ \Leftrightarrow \quad & T_{k,j} = \hat{e}_j^T Q_j^{-1} \hat{e}_j - \hat{e}_k^T Q_k^{-1} \hat{e}_k > k_{k,j}, \end{aligned} \tag{6.84}$$

where $k_{k,j} = \ln(\frac{\det(Q_j)}{\det(Q_k)}$.

Denoting the corresponding least-squares solution of the measurements as $\hat{y}_{\mathsf{all},j} = F_j(\hat{x}_{u,j}) + C_j\nabla_j$, the residual then can be expressed as:

$$\hat{e}_j = y_{\mathsf{all}} - \hat{y}_{\mathsf{all},j} = y_{\mathsf{all}} - C_j\nabla_j - F_j(\hat{x}_{u,j}). \tag{6.85}$$

Assuming that the second- and higher- order terms in $\hat{x}_{u,j}$ and $\hat{y}_{\text{all},j}$ are negligible, their approximations, can be given as:

$$(\hat{x}_{u,j} - x_u) \approx (A_j^T Q_j^{-1} A_j)^{-1} A_j^T Q_j^{-1} [y_{\text{all}} - C_j \nabla_j - F_j(x_u)] \quad (6.86)$$

$$\hat{y}_{\text{all},j} - F_j(x_u) - C_j \nabla_j \approx A_j (A_j^T Q_j^{-1} A_j)^{-1} A_j^T Q_j^{-1} [y_{\text{all}} - C_j \nabla_j - F_j(x_u)], \quad (6.87)$$

where the notation $A_j = \partial_{x_u^T} F_j(x_u)$ is used and x_u is the true value of the unknown position.

Moreover, (6.85) can be approximated as:

$$\hat{e}_j = y_{\text{all}} - \hat{y}_{\text{all},j} \approx P_{A,j}^{\perp} [y_{\text{all}} - C_j \nabla_j - F_j(x_u)], \quad (6.88)$$

with $P_{A_j,Q_j} = A_j (A_j^T Q_j^{-1} A_j)^{-1} A_j^T Q_j^{-1}$ and $P_{A_j,Q_j}^{\perp} = I - A_j (A_j^T Q_j^{-1} A_j)^{-1} A_j^T Q_j^{-1}$.

Given the general assumption that $\underline{y}_{\text{all}} \sim \mathcal{N}(F(x_u) + C\nabla, Q_{\text{all}})$, with $F(.)$, C, ∇ and Q_{all} the correct function, state matrix, NLoS bias vector and measurement variance matrix, the distribution of $\underline{\hat{e}}_j$ can be derived based on (6.88) as:

$$\underline{\hat{e}}_j \sim \mathcal{N}(P_{A,j}^{\perp} (F(x_u) - F_j(x_u) + C\nabla - C_j \nabla_j), P_{A,j}^{\perp} Q_{\text{all}}^{-1} (P_{A,j}^{\perp})^T). \quad (6.89)$$

Then, based on Theorem 6.1, it can be easily shown that:

$$\underline{T}_{k,j} \sim \sum_{i=1}^{r_j} \lambda_{j,i} \chi^2(1, u_{j,i}^2) - \sum_{i=1}^{r_k} \lambda_{k,i} \chi^2(1, u_{k,i}^2), \quad (6.90)$$

where the parameters r_j, $\lambda_{j,i}$, $u_{j,i}$, r_k, $\lambda_{k,i}$ and $u_{k,i}$ can be computed in a similar way as in (6.70)-(6.72).

Finally, the full distribution of $\underline{T}_{k,j}$ can be derived based on the joint PDF of the $r_j + r_k$ Chi-square distributed random variables based on (6.90). If H_j is true, then the relevant parameters in (6.90) should be calculated by choosing $x_u = x_{u,j}$, $\nabla = \nabla_j$ and $Q_{\text{all}} = Q_j$, and if H_k is true, the calculation should be done by choosing $x_u = x_{u,k}$, $\nabla = \nabla_k$ and $Q_{\text{all}} = Q_k$. In practice, since the knowledge of the true values are not available, the corresponding estimates obtained under correct hypothesis will be used instead.

Without loss of generality, we assume that H_k is the correct state, then γ_k can be calculated based on (6.59) and (6.84) as:

$$\gamma_k = \prod_{j \in \{j \,|\, \|S_j - S_k\| = 1\}} P\left(T_{k,j} > k_{k,j}\right) \quad (6.91)$$

To this end, for Classification Type II, the probability of correct decision γ_k when H_k true, can be verified with the following steps:

1. calculate the mean values and variances of $\underline{\hat{e}}_j$ and $\underline{\hat{e}}_k$ based on (6.89), where the true state matrix C should be chosen as C_k,

2. calculate the relevant parameters in (6.90), such as $u_{j,i}$, $\lambda_{j,i}$, $u_{k,i}$, $\lambda_{k,i}$, according to Theorem 6.1,

3. calculate the probability of correct decision based on (6.91).

Similar as for Classification Type I, here we also give the expressions of an upper- and a lower bounds for $\underline{T}_{k,j}$. The upper bound on the probability of correct decision when H_k is true is given as:

$$\begin{aligned}\gamma_k^{\mathsf{UB}} = \prod_{j\in\{j\,|\,\|S_j-S_k\|=1\}} & \left[a\,P^{\mathsf{UB}}\left(T_{k,j}^{\mathsf{UB}} > k_{k,j}^{\mathsf{UB}} | \underline{T}_{k,j}^{\mathsf{UB}} \sim \chi^2(1,\lambda^{\mathsf{UB}})\right)\right. \\ + & \left.(1-a)\,P^{\mathsf{UB}}\left(T_{k,j}^{\mathsf{UB}} < k_{k,j}^{\mathsf{UB}} | \underline{T}_{k,j}^{\mathsf{UB}} \sim \chi^2(1,0)\right)\right]\end{aligned} \tag{6.92}$$

and the lower bound:

$$\begin{aligned}\gamma_k^{\mathsf{LB}} = \prod_{j\in\{j\,|\,\|S_j-S_k\|=1\}} & \left[a\,P^{\mathsf{LB}}\left(T_{k,j}^{\mathsf{LB}} > k_{k,j^{\mathsf{LB}}} | \underline{T}_{k,j}^{\mathsf{LB}} \sim \chi^2(1,\lambda^{\mathsf{LB}})\right)\right. \\ + & \left.(1-a)\,P^{\mathsf{LB}}\left(T_{k,j}^{\mathsf{LB}} < k_{k,j}^{\mathsf{LB}} | \underline{T}_{k,j}^{\mathsf{LB}} \sim \chi^2(1,0)\right)\right]\end{aligned} \tag{6.93}$$

The details about $k_{k,j}^{\mathsf{UB}}$, $k_{k,j}^{\mathsf{LB}}$, λ^{UB}, λ^{LB} and a can be found in Appendix A.4.2, where the full derivations are given.

The current test is still quite complicated, since one needs to perform estimation for all 2^m states. Based on the fact that the ToA range measurements are much more accurate than the RSS ones, we would like to explore the possibility in the next section, where position estimates are calculated based on ToA measurements and NLoS identification is done with RSS measurements.

6.2.4 Simple Likelihood Ratio Test Type I

The computations involved in Classification Type II are still quite cumbersome, since non-linear least-squares estimation needs to be performed for each hypothesis. To reduce the computational load, we introduce here a further simplified test that requires only two rounds of least-squares estimation, one for identification and one for calculating the final estimate. The testing scheme is named as SLRT Type I, which contains the following steps:

1. Estimate $\hat{x}_u$ using all the measurements, assuming they are all collected under LoS:

$$\begin{aligned}\hat{x}_u &= \arg\max_{x_u}(\mathsf{det}(2\pi Q_0))^{-1/2}\exp\{-\frac{1}{2}\|y_{\mathsf{all}} - F(x_u, C_0\nabla_0)\|^2_{Q_0^{-1}}\} \\ &= \arg\max_{x_u}\|y_{\mathsf{all}} - F(x_u)\|^2_{Q_0^{-1}}\end{aligned} \tag{6.94}$$

   where the expression of $F(x_u)$ can be found with H_0 in Section 6.2.3.

2. Calculate the least-squares solution for the measurement vector $\hat{y} = A(\hat{x}_u)$.

3. Establish the following hypotheses on (only) the RSS measurements for the i-th link, $i = 1, ..., m$:

$$H_0 \ : \ \underline{P}_i \sim N(m_{L,i}, \sigma^2_{L,i}) \tag{6.95}$$
$$H_1 \ : \ \underline{P}_i \sim N(m_{N,i}, \sigma^2_{N,i}) \tag{6.96}$$

   where $m_{L,i} = P_0 + 10n_{\mathsf{LoS}} \log_{10} \frac{\hat{y}_i}{d_0}$, $m_{N,i} = P_0 + 10n_{\mathsf{NLoS}} \log_{10} \frac{\hat{y}_i}{d_0}$, $\sigma^2_{L,i} = \sigma^2_{RSSLOS,i}$, and $\sigma^2_{N,i} = \sigma^2_{RSSNLOS,i}$.

4. Perform SLRT with the two distributions in (6.95) and (6.96) for each i, and construct the state vector S_i:

$$\underline{T}_i = (\underline{P} - \frac{\sigma^2_{N,i} m_{L,i} - \sigma^2_{L,i} m_{N,i}}{\sigma^2_{N,i} - \sigma^2_{L,i}})^2 \underset{\mathsf{NLoS}, S_i=1}{\overset{\mathsf{LoS}, S_i=0}{\lessgtr}} \frac{\sigma^2_{N,i}\sigma^2_{L,i}(m_{L,i} - m_{N,i})^2}{(\sigma^2_{N,i} - \sigma^2_{L,i})^2} - \frac{2\sigma^2_{N,i}\sigma^2_{L,i}}{\sigma^2_{N,i} - \sigma^2_{L,i}} \ln \frac{\sigma_{L,i}}{\sigma_{N,i}}. \tag{6.97}$$

   Please note that $\underline{P}$ is in dB and is of negative values, see Figure 6.3. Moreover, we have assumed equal probabilities of LoS and NLoS conditions, i.e. the threshold of likelihood ratio is chosen as 1.

5. Calculate the corresponding NLoS biases based on the model (6.78) used for Classification Type II, subtracting these biases from the corresponding identified NLoS ToA range measurements.

6. Obtain the final least-squares position estimate based on the RSS and the corrected ToA range measurements (with calculated NLoS biases (6.78) subtracted), with variance matrix corresponding to the state vector $S = [S_1, ..., S_m]^T$.

Please note that the models in (6.95) and (6.96) are the approximations for the exact models, where a least-squares estimate for $\hat{y}_i$ is used to approximate its unknown true value d_i.

The motivation of the SLRT Type I can be described as follows:

1. With the presence of m ToA range measurements, the least-squares estimate for measurement $\hat{y}_i$ should be a good approximation to d_i, since all $2m$ measurements are used in calculating x_u.

2. RSS models are relatively in-sensitive to a small change in d_i, i.e., the LoS and NLoS power delay profiles are still well separated after we replace d_i by $\hat{y}_i$.

6.2.4.1 Level of Significance and Probability of Missed Detection

Given the distributions of $\underline{P}_i$ under H_0 and H_1 in (6.95) and (6.96), respectively, the test statistics can be verified for both hypotheses as follows:

$$H_0 : \sigma^{-2}_{L,i}\underline{T}_i \sim \chi^2(1, T^2_{L,i}/\sigma^2_{L,i}) \text{ and } H_1 : \sigma^{-2}_{N,i}\underline{T}_i \sim \chi^2(1, T^2_{N,i}/\sigma^2_{N,i}) \tag{6.98}$$

where $\underline{T}_i = (\underline{P} - \frac{\sigma^2_{N,i} m_{L,i} - \sigma^2_{L,i} m_{N,i}}{\sigma^2_{N,i} - \sigma^2_{L,i}})^2$, $T_{L,i} = m_{L,i} - \frac{\sigma^2_{N,i} m_{L,i} - \sigma^2_{L,i} m_{N,i}}{\sigma^2_{N,i} - \sigma^2_{L,i}}$ and $T_{N,i} = m_{N,i} - \frac{\sigma^2_{N,i} m_{L,i} - \sigma^2_{L,i} m_{N,i}}{\sigma^2_{N,i} - \sigma^2_{L,i}}$.

The level of significance α_i and the probability of missed detection β_i for the i-th link are then calculated as:

$$\alpha_i = P(\sigma^{-2}_{L,i}\underline{T}_i < k_{L,i}), \text{ with distribution in (6.98) under } H_0 \quad (6.99)$$

$$\beta_i = P(\sigma^{-2}_{N,i}\underline{T}_i > k_{N,i}), \text{ with distribution in (6.98) under } H_1 \quad (6.100)$$

where

$$k_{L,i} = \sigma^{-2}_{L,i}\left(\frac{\sigma^2_{N,i}\sigma^2_{L,i}(m_{L,i} - m_{N,i})^2}{(\sigma^2_{N,i} - \sigma^2_{L,i})^2} - \frac{2\sigma^2_{N,i}\sigma^2_{L,i}}{\sigma^2_{N,i} - \sigma^2_{L,i}} \ln \frac{\sigma_{L,i}}{\sigma_{N,i}}\right) \quad (6.101)$$

$$k_{N,i} = \sigma^{-2}_{N,i}\left(\frac{\sigma^2_{N,i}\sigma^2_{L,i}(m_{L,i} - m_{N,i})^2}{(\sigma^2_{N,i} - \sigma^2_{L,i})^2} - \frac{2\sigma^2_{N,i}\sigma^2_{L,i}}{\sigma^2_{N,i} - \sigma^2_{L,i}} \ln \frac{\sigma_{L,i}}{\sigma_{N,i}}\right) \quad (6.102)$$

To this end, it is clear that the SLRT Type I greatly reduces the computational load compared to the two previous described tests by reducing the number of estimation rounds from 2^m to 2. Actually, this scheme can be even further simplified. The reasoning and the procedure of the simplification is introduced next.

6.2.5 Simple Likelihood Ratio Test Type II

The SLRT Test can be further simplified, if we skip the step where $\hat{y}_{\mathsf{all},i}$ is estimated, and use the measurement $y_{\mathsf{all},i}$ to replace its true value in the models. The steps can be described as follows in a similar way as for SLRT Type I:

1. Establish the following hypotheses for the i-th link, $i = 1, ..., m$:

$$H_0 : \underline{P}_i \sim N(m_{L,i}, \sigma^2_{L,i}) \quad (6.103)$$

$$H_1 : \underline{P}_i \sim N(m_{N,i}, \sigma^2_{N,i}) \quad (6.104)$$

   where $m_{L,i} = P_0 + 10 n_{\mathsf{LoS}} \log_{10} \frac{d_i}{d_0}$, $m_{N,i} = P_0 + 10 n_{\mathsf{NLoS}} \log_{10} \frac{d_i}{d_0}$, $\sigma^2_{L,i} = \sigma^2_{RSSLOS,i}$, and $\sigma^2_{N,i} = \sigma^2_{RSSNLOS,i}$.

2. Perform SLRT for each i, and construct the state vector S_i:

$$\underline{T}_i = (\underline{P} - \frac{\sigma^2_{N,i} m_{L,i} - \sigma^2_{L,i} m_{N,i}}{\sigma^2_{N,i} - \sigma^2_{L,i}})^2 \underset{\mathsf{NLoS}, S_i=1}{\overset{\mathsf{LoS}, S_i=0}{\lessgtr}} \frac{\sigma^2_{N,i}\sigma^2_{L,i}(m_{L,i} - m_{N,i})^2}{(\sigma^2_{N,i} - \sigma^2_{L,i})^2} - \frac{2\sigma^2_{N,i}\sigma^2_{L,i}}{\sigma^2_{N,i} - \sigma^2_{L,i}} \ln \frac{\sigma_{L,i}}{\sigma_{N,i}} \quad (6.105)$$

3. Calculate the corresponding NLoS biases based on the model used for Classification Type II, subtracting these biases from the corresponding identified NLoS ToA range measurements.

4. Obtain the final least-squares position estimate based on the RSS and the corrected ToA range measurements (with calculated NLoS biases (6.78) subtracted), with variance matrix corresponding to the state vector $S = [S_1, ..., S_m]^T$.

Note that SLRT Type II is first proposed by [107], though the test is described in this section as a simplified version of the most general test Classification Type I.

6.2.5.1 Level of Significance and Probability of Missed Detection

The level of significance and the probability of missed detection in SLRT Type II can be derived in a similar way as we did for SLRT Type I. The expressions in (6.98)-(6.100) can be directly used for the distributions of $\underline{T}_i$, α and β, while we only need to change $\hat{y}_i$ to the corresponding measurement y_i in $m_{L,i}$ and $m_{N,i}$.

6.3 TDoA and RSS Combined

Up till now, we have assumed that all the m transmitters are synchronized, and there are no clock offset in the ToA measurements. Actually, it is rather straightforward to extend some of the tests from ToA/RSS measurements to TDoA/RSS measurements, by changing $A(x_u)$ in (6.42) and (6.43) to $A(x)$, with $A_i(x) = d_i + b_c = \sqrt{(x_u - x_i)^T(x_u - x_i)} + b_c$, where b_c is a clock offset, common in all pseudorange measurements. Please note that $A(x_u)$ in (6.44) and (6.45) is left unchanged, i.e. the RSS models are not changed.

Classification Type I and SLRT Type I can be adjusted to work for TDoA/RSS measurements by extending the x_u to $x = [x_u^T, b_c]^T$, and the derived equations can be easily modified accordingly.

Classification Type II and SLRT Type II can not be performed for TDoA/RSS measurements, due to the fact that, for Classification Type II, the calculation of (6.78) requires the knowledge of $A_i(x_u) = d_i$, which can not be approximated with the pseudorange measurements, and for SLRT Type II, the RSS models can not be approximated using the pseudorange measurements, since the pseudorange measurements contain a clock offset, which makes them not good approximations of the corresponding true distances.

6.4 Computational Load Comparison

In this section we will give an analysis on the computational loads of the four proposed schemes.

6.4.1 Classification Type I vs Type II

Assume that the Gauss-Newton method is used for estimation, the flop count is performed in Appendix (A.89), which shows that in each iteration, the number of required flops for

the Classification Type I is:

$$\begin{aligned} \text{flops}_{CTI} &= q^3/3 + (n+2)q^2 + (n^2 + 4mn + 3n + 4m)q \\ &+ n^3/3 + (4m+1)n^2 + 13mn + 4n + 1, \end{aligned} \tag{6.106}$$

where q is the number of NLoS biases. Please note that here we do not consider a clock offset, but cases with a clock offset can be easily analyzed in a similar way.

For Classification Type II, the biases are pre-calculated, i.e. the set of unknown parameters is always the same, but the vector of (corrected) measurements is different for each hypothesis. Hence, the number of flops in one iteration can be easily calculated by setting $q = 0$ in (6.106).

Thus, the computational difference between Classification Type I and II in one iteration, considering all 2^m states can be given as:

$$\text{flops}_{CTI-CTII} = \sum_{k=1}^{2^m-1} \left\{ q_k^3/3 + (n+2)q_k^2 + (n^2 + 4mn + 3n + 4m)q_k \right\}. \tag{6.107}$$

When m is large, the above equation implies a very big number, since there are one occurrence of $q_k = m$, m occurrences of $q_k = m-1$, $(m-1)m/2$ occurrences of $q_k = m-2$, and so on. This means that we can expect large computational savings, switching from Classification Type I to II.

6.4.2 Classification Type II vs SLRT Type I vs SLRT Type II

The computational difference between Classification Type II and SLRT Type I is easier to quantize. Both schemes involve parameter estimation with $\underline{S} = S_0$, i.e. NLoS free; then the Classification Type II will perform parameter estimations for further $2^m - 1$ states, each time with a different corrected set of measurements, while SLRT Type I perform (only once) the calculations for the picked hypothesis. The detailed steps of these two schemes can be found in Section 6.2.3 and Section 6.2.4, respectively. Thus, ignoring the common part, the computational load of SLRT Type I can be quantized as $\frac{1}{2^m-1}$ of that required by the Classification Type II. Please note that this ratio is only a rough number, since an accurate value depends on the correct state.

The comparison between SLRT Type I and SLRT Type II is even straightforward, where the difference is the calculations involved in the estimation for $k = 0$. The number of flops in each iteration can be calculated using (6.106) with $q = 0$, which gives

$$n^3/3 + (4m+1)n^2 + 13mn + 4n + 1.$$

6.5 Validation Results

In this section, we validate the proposed tests using actual UWB combined ToA/RSS and combined TDoA/RSS measurements measurements. First path RSS measurements are used, since the first path in LoS cases is directly related to the LoS path [71]. The

details of the measurement campaign setup can be found in Chapter 2. The geometry of the positioning system for validation is generated in a same way as in Section 4.4. The estimation method used in the validation is the Levenberg Marquardt method.

For both ToA/RSS and TDoA/RSS combinations are tested under two different setups: 1) 4 LoS and 1 NLoS 2) all 5 NLoS. The obtained results are shown in Table 6.2-6.5. For each setup, positioning is performed for $N = 5000$ rounds, each with a set of 5 random selected links. The quantities $P(\text{LoS}|\text{LoS})$ and $P(\text{LoS}|\text{LoS})$ in these tables denote the correct probability of detecting a LoS link as LoS and of detecting a NLoS link as NLoS, respectively. The term RMSE is empirically obtained as $\text{RMSE} = \sqrt{\frac{1}{N-1}\sum_{l=1}^{N}\|\hat{x}_u^l - x_u\|^2}$, with x_u the true position, $\hat{x}_u^l$ the estimates obtained in the l-th round.

As a general remark, the performance of the four discussed tests can be ranked as: Classification Type I better than Classification Type II better than SLRT Type I better than SLRT Type II. This is intuitively reasonable given the fact that the computational loads of the tests are ranked in a reversed way. Moreover, with 7.5 GHz of full UWB bandwidth, the differences among the performances are subtle.

Both ToA/RSS and TDoA/RSS, the results obtained when $S = [0, 0, 0, 0, 1]^T$ (Setup 1) are quite easy to interpret. But with $S = [1, 1, 1, 1, 1]^T$ (Setup 2), two phenomenons require further discussion:

1. classification Type I has the *largest* RMSE with 7.5 GHz bandwidth,
2. classification Type I has the smallest RMSE with 0.5 GHz bandwidth.

The possible reasons for the first phenomenon are:

1. given 5 (pseudo)range measurements, 5 RSS measurements, and $S = [1, 1, 1, 1, 1]^T$, there are 8 unknowns including 3 position coordinates and 5 NLoS biases in the ToA/RSS case and 9 unknown parameters in the TDoA/RSS cases with one additional parameter b_c. Therefore, redundancy is much reduced compared to the case with only 1 NLoS, and the estimation of the 5 NLoS biases rely heavily on the RSS measurements, which are of relatively bad quality. Thus, the final position RMSE is expected to be large.
2. for the other 3 tests, the NLoS biases are calculated via (6.78), and then treated as *known* parameters in a later stage when x_u (and an extra b_c with TDoA) is estimated. Thus, assuming an accurate NLoS bias model in (6.78), better positioning results can be achieved compared to the case where the NLoS biases have to be estimated together with x.

The reason for the second phenomenon is that the NLoS bias model in (6.78) has better quality under 7.5 GHz than 0.5 GHz, see [36]. Thus, with less accurate NLoS biases, calculated using (6.78), the quality of the final position estimates is of course affected, which explains the comparison results between Classification Type I and the other three tests under two different bandwidths.

Table 6.2: Results for proposed tests using ToA/RSS measurements. The correct state vector is $S = [0, 0, 0, 0, 1]^T$.

B=0.5 GHz	P(LoS|LoS)	P(NLoS|NLoS)	RMSE [m]
Classification Type I	93.4%	85.6%	0.086
Classification Type II	93.8%	83.2%	0.093
SLRT Type I	94.6%	82.8%	0.104
SLRT Type II	95.1%	81.8%	0.114
B=7.5 GHz	P(LoS|LoS)	P(NLoS|NLoS)	RMSE [m]
Classification Type I	99.0%	98.9%	0.016
Classification Type II	98.2%	98.0%	0.021
SLRT Type I	97.2%	98.1%	0.023
SLRT Type II	96.9%	98.1%	0.023

Table 6.3: Results for proposed tests using ToA/RSS measurements. The correct state vector is $S = [1, 1, 1, 1, 1]^T$.

B=0.5 GHz	P(NLoS|NLoS)	RMSE [m]
Classification Type I	89.7%	0.170
Classification Type II	85.4%	0.196
SLRT Type I	83.2%	0.204
SLRT Type II	81.5%	0.214
B=7.5 GHz	P(NLoS|NLoS)	RMSE [m]
Classification Type I	98.4%	0.102
Classification Type II	98.0%	0.070
SLRT Type I	98.0%	0.074
SLRT Type II	98.1%	0.076

Table 6.4: Results for proposed tests using TDoA/RSS measurements. The correct state vector is $S = [0, 0, 0, 0, 1]^T$.

B=0.5 GHz	P(LoS|LoS)	P(NLoS|NLoS)	RMSE [m]
Classification Type I	93.1%	85.2%	0.089
SLRT Type I	94.2%	82.6%	0.100
B=7.5 GHz	P(LoS|LoS)	P(NLoS|NLoS)	RMSE [m]
Classification Type I	96.8%	97.6%	0.019
SLRT Type I	98.1%	96.2%	0.021

Table 6.5: Results for proposed tests using TDoA/RSS measurements. The correct state vector is $S = [1, 1, 1, 1, 1]^T$.

B=0.5 GHz	P(NLoS|NLoS)	RMSE [m]
Classification Type I	88.1%	0.178
SLRT Type I	83.5%	0.179
B=7.5 GHz	P(NLoS|NLoS)	RMSE [m]
Classification Type I	98.5%	0.096
SLRT Type I	98.2%	0.064

6.6 Summary

In this chapter, we reviewed several existing NLoS identification/mitigation techniques, and also proposed four tests using combined ToA/RSS or TDoA/RSS. The features of all these techniques are summarized in Table 6.6.

The four proposed LoS/NLoS identification/mitigation schemes all work based on the models described in (6.42)-(6.45), where the combination of timing and RSS measurements are used because 1) the timing measurements are usually very accurate compared to RSS measurements, 2) the distribution of RSS measurements under LoS and NLoS conditions are well separated. The first proposed scheme, Classification Type I, is the most general test that utilizes all available information in the models (6.42)-(6.45), where the NLoS ToA biases are treated as unknowns in addition to the unknown position x_u (and b_c in TDoA case). In Classification Type II, simplification is introduced by pre-calculating the NLoS ToA biases using (6.78). Later, by further assuming that the estimate $\hat{y}_{\text{all}}$ obtained under H_0 in (6.46) is a good approximation of its true value in the RSS models (6.44)-(6.45), we reduce classification (multiple hypotheses testing) to SLRT Type I, where the NLoS identification for each link is done individually. Finally, SLRT Type I is further simplified to Type II, where the ToA range measurement is directly used as the approximation to its true value in (6.44)-(6.45). The computational load of the schemes decrease as simplified from Classification Type I to Classification Type II, to SLRT Type I and finally to the most simple SLRT Type II. The performance, however, is also degraded in general, except the case where most of the links are NLoS. Please note that, currently, Classification Type II and SLRT Type II only work for ToA/RSS cases.

Validation results show that, under full UWB signal bandwidth of 7.5 GHz, up to 99% correct decision rate can be achieved with Classification Type I. Meanwhile, the performance differences among these methods are subtle. Thus, with 7.5 GHz UWB radio signal, we suggest to use SLRT Type II for ToA/RSS systems and SLRT Type I for TDoA/RSS systems.

Compared to the reviewed methods in Section 6.1, the four schemes proposed in Section 6.2 have the following advantages:

1. In the first three schemes, the fact that the measurements from different links are related via the unknown position is exploited, rather than performing identification

based on information of only one link.

2. The achievable correct detection rate is very high, even using the last scheme with lowest computational load.

3. All the schemes are single epoch methods that do not require information of history data.

4. The effort required to evaluate the statistics of timing and RSS measurements, e.g. measurement variances, pathloss exponents etc., are moderate, and are anyway also required by some of the reviewed methods such as running variance.

Moreover, when history data is available, it is possible to achieve better results or reduce computational load by combining one of the four proposed tests with the reviewed Change of RSS scheme, and the latter has been validated with actual data in [94] showing promising results, see Table 6.6.

The discussion on the topic of LoS/NLoS identification/mitigation is close to an end now, but there are still several issues worth further investigation, e.g. the quality of the models (6.42)-(6.45), other ideas to reduce the 2^m rounds of estimation in Classification Type I and II, and possibilities to extend Classification Type II and SLRT Type II for TDoA/RSS cases etc.. Some discussions on these issues will be provided in the next chapter.

Table 6.6: Summary of reviewed and proposed techniques. The results for reviewed techniques are found in [94]. For the detection results with CT I, II and ST I, II, 7.5 GHz bandwidth was used with 4 LoS and 1 NLoS links.

	Major		Actual indoor results e.g. in office	
	advantages	disadvantages	$P(\text{NLoS}\|\text{NLoS})$	$P(\text{LoS}\|\text{LoS})$
Techniques using full CIR	all information can be utilized	obtaining full CIR is expensive	98.2% (obtained with Confidence Metric)	95.1%
Running Variance	simple method	receiver should be almost static; large latency	79.7%	94.3%
Change of RSS	simple method good results	previous RSS value needed	93.8%	97%
Nonparam-etric Methods	approximates actual PDF	heavily depends on the kernel functions and the number of survey points	N.A.	N.A.
Least-squares residual	does not need PDF	identification is hard with majority of NLoS; heavy computations	N.A.	N.A.
Classification Type I (CT I)	optimal results	Very heavy computations	98.9%	99.0%
Classification Type II (CT II)	good results; a bit worse than CT I	heavy but less than CT I	98.0%	98.2%
SLRT Type I (ST I)	a bit worse than CT II	much less than CT II	98.1%	97.6%
SLRT Type II (ST II)	a bit worse than ST I	slightly less than ST I	98.1%	96.9%

7 Conclusions and Recommendations

With the great success of the GPS, position-based applications have attracted more and more attention, not only in outdoor open areas, but also in indoor environments. However, due to dense multipaths effects and strong signal attenuation (even total blockage), providing positioning service in indoor environments using the GPS or any other satellite navigations system is currently not feasible. To find an alternative solution, the research project: "HERE: indoor positioning based on UWB radio signals", has been launched. The project is supported by the Dutch Technology Foundation STW (project no. 0.7343), and investigates the feasibility of providing indoor positioning services using a system which:

- performs ranging using UWB radio signals and
- applies optimal position estimation.

The HERE team is composed of two Ph.D. candidates and two supervisors. As one of the Ph.D. candidates, the author of this thesis focused on the second item in the list above, with the following main research subjects covered:

1. Investigation on the feasibility of applying least-squares theory for positioning, which is a non-linear estimation problem, in typical *indoor* environments.
2. Study and analysis of existing positioning algorithms, and development of improvements, aiming at either reducing computational load or increasing positioning accuracy.
3. Development of hypothesis testing for NLoS detection and mitigation, together with theoretical performance analysis.

In previous chapters, the details related to the above subjects have been presented. The objective of this chapter is to give a summary of the conclusions drawn from the previous chapters, and in the mean time give recommendations for future research.

As a general remark, self positioning scenario (like satellite navigation) is by default assumed in this thesis, but the content can also equally be applied to tracking systems in which the user sends signals to the base stations where signal processing and position estimation are accomplished.

7.1 Conclusions

7.1.1 Non-linear Least-Squares for Indoor Positioning

The reasons we chose least-squares based algorithms as the focus of this thesis are threefold.

First, least-squares estimation does not require knowledge of the measurement PDF. Concerning indoor environments, accurate PDFs for timing-based and RSS-based measurements are only available for LoS cases, assuming very good multipath resolvability; this modeling task is difficult for NLoS situations, since the material of the objects that cause NLoS effects can differ from place to place, e.g. the behavior of radio signals blocked by concrete and metallic objects are expected to be significantly different. Hence, it is practically impossible to find an universal PDF of measurements that suits all indoor environments.

Second, when the PDF of the measurements is known and elliptically contoured, the weighted least-squares solution with a weight matrix equal to the inverse of the measurement variance matrix,

- is equivalent to the result obtained by means of ML estimation, and
- closely approximates the CRLB, which represents the lowest achievable variance of estimators.

This statement was verified in Section 3.3.3. Please note that the requirement on a elliptically contoured PDF seems usually fulfilled in LoS cases using UWB signals.

Third, applying least-squares based algorithms for positioning is the default already for long time and has proved to be a success with the GPS.

7.1.1.1 Local/Global Minima

The range and Cartesian position coordinates are inherently related in a non-linear way, meaning that the least-squares objective function can have multiple minima. Multiple minima with similar values usually occur when the geometry of the system is configured close to coplanar. As shown in Section 3.3.2, this type of geometries introduces difficulties in finding the correct solution (regarding the actual position), since the global minimum does not always correspond to the correct solution. In Figure 3.3(b), such a case where the global minimum corresponds to a wrong solution has been shown.

> *In a system configured with a (close to) coplanar geometry, the correct solution, regarding the actual position, can not be guaranteed by picking the one that corresponds to the global minimum. Additional information is required for this purpose. Hence, (close to) coplanar type of geometries should be avoided at the system design phase.*

7.1.1.2 Bias due to Non-linearity

The non-linear least-squares estimators are inherently biased (3.65), even when the used measurements are unbiased, due to the fact that unbiasedness does not carry over through a non-linear relation. In a satellite navigation system, the bias may be negligible. This can be roughly seen by considering a circle with a satellite as the center and the distance between the satellite and a point on earth as the radius. Then the small segment of this circle containing the point on earth can be well approximated by a straight line segment (hence a linear relation), due to the extremely large radius, e.g. not less than about 20200 km regarding a GPS satellite. For indoor applications, this radius is much and much smaller, and therefore, a bias check is proposed.

> *We proposed to test significance of the bias in an estimator by comparing it to the STD of the estimator. If the bias is larger, it is considered significant.*

Validation results based on both simulation and actual data show that for fully synchronized systems, this bias is not significant while for systems with a clock offset, the bias can be a performance degrader when all the transmitters are on one side of the receiver.

7.1.2 Existing Range-based Least-Squares Algorithms

7.1.2.1 Iterative Descent Algorithms

The Iterative Descent (ID) algorithms iteratively calculate non-linear least-squares estimates by means of linear approximations, i.e. they start with an initial guess of the unknown parameter vector, which is then updated, from iteration to iteration, based on the rule defined by the specific method. The reviewed ID methods in this thesis include the Steepest Descent method, the Newton method, the Trust Region method, the Gauss-Newton method, and the Levenberg Marquardt method. A comparison of these ID methods regarding computational load and convergence rate can be found in Table 4.1.

> *A least-squares solution is determined by the objective function rather than the specific method used to search the solution. Despite the different linear approximations used by these ID methods, the final solutions obtained by all ID methods are equivalent when they are used to minimize the same objective function, and they all successfully converge.*

Initial Guess All the iterative algorithms require an initial guess of the unknown parameter to start. The initial guess should be chosen as close to the actual position as possible, since:

- iterative algorithms will converge to the solution that locates in the same 'valley' as the initial guess, and a non-linear problem may have multiple 'valleys';
- the algorithms will converge in less steps/iterations (computational advantage).

For on-earth satellite navigation, choosing a initial guess in the same 'valley' as the correct solution is not difficult, e.g. even the center of the earth works. However, for indoor positioning systems, it is not possible to find such an initial guess that works for all scenarios.

We proposed to obtain the initial guess by means of a direct method as a universal solution. In particular cases with healthy geometries where the target is surrounded in the three-dimensional space by the transmitters, a more simpler way is to use the geometrical center of the transmitters as the initial guess.

7.1.3 Direct Methods

In this thesis, we have cast all the direct methods in a universal structure (4.21), and the common idea of the direct methods is that they all apply a squaring operation to get rid of the square root in the range-coordinates equation. By further introducing simplifications, the position can be estimated in non-iterative ways. The reviewed direct methods can be divided into two groups based on the introduced simplifications, or more precisely the objective function to minimize.

7.1.3.1 Single Objective Function Methods

The SOF methods refer to the ones that use only one objective function. In this thesis, many SOF methods are reviewed, which are scattered across the fields of radar, aerospace engineering, oceanic engineering, (acoustic) signal processing and wireless communications.

The SOF methods are equivalent in the sense that they can be looked on as different realizations of one universal method by applying different weight matrices. Moreover, SOF methods are not suitable to solve positioning problem for systems with a clock offset (TDoA), due to rank defect problems in a typical geometry where transmitters surround the target.

7.1.3.2 Multiple Objective Function Methods

The MOF methods refer to the ones that use a *combined* objective function, i.e. the final solution is obtained by combining the results from minimizing two or more different objective functions.

The basic ideas of different MOF methods are equivalent, but the outcomes of MOF methods are not equivalent in general. Moreover, the MOF solutions are variant under a change of the choice for the quadratic constraint (Section 4.2.2).

In the best case, the MOF solution is equivalent to the solution obtained using an ID method, but in this case, knowledge of the actual position is required, see Section 4.2.2, which is not possible in practice. Finding a good choice for the quadratic constraint is similar as finding a good initial guess for the ID methods.

7.1.3.3 Constrained Iterative Methods

The constrained iterative methods achieve better estimation results compared to the direct methods by rigorously exploiting the quadratic constraint. In turn, the solution has to be obtained iteratively. The reviewed constrained iterative methods apply EVD to achieve 1-dimensional calculations in each iteration, implying lower computational load. We should note that the EVD itself is a non-linear problem which in general needs to be solved iteratively.

> *The EVD can be skipped in general and one can still solve the problem with only one unknown parameter in each iteration. This is helpful to interpret the computational load of the constrained iterative methods, which is still quite high, since one needs to inverse a* $(n+2)\times(n+2)$ *matrix, and perform matrix multiplications afterwards. Here* n *is the dimension of unknown coordinates.*

A full summary of all reviewed existing methods can be found in Table 4.4, concerning computational load, positioning accuracy and requirements on redundancy. We should note that when comparing the positioning accuracy of different methods, we are actually comparing the objective functions to minimize, since the quality of an estimator is solely determined by the objective function rather than the applied method to search the solution. For example, the minimizer of the objective function of the constrained iterative methods can also be searched by the ID methods, and same results will be obtained.

7.2 Improvements

In this thesis two possible ways to improve the existing methods are given:

1. improving the estimation accuracy of the SOF methods on one hand and keeping the method non-iterative on the other hand, results in the MOF Type II method;
2. reducing the computational load of the traditional ID methods while keeping the estimation accuracy as high as possible, results in a transformed least-squares framework.

7.2.1 MOF Type II Method

The MOF II method achieves better estimation accuracy than the SOF methods by exploiting multiple, but similar, fully constrained models. The solution is better, in the sense that it corresponds to an equal or smaller value of the original least-squares objective function (3.50).

> *The MOF II method is non-iterative and with guaranteed better estimation accuracy than the SOF methods. In our validation results, it also performs better than the MOF method. However, the MOF II method requires the evaluation of three*

candidates which implies heavier computational load than the SOF methods (one candidate) and the MOF methods (two candidates). Moreover, the MOF II method also suffers from the rank defect problem as the SOF methods.

7.2.2 Transformed Least-Squares Framework

The TLS framework, consisting of two steps, is proposed to realize low-complexity positioning with highly accurate estimators. The low-complexity nature of the framework comes from Step 1, where the positioning problem is transformed from $(n+1)$- or n-dimension to a lower dimension. In Step 2, the high accuracy estimator is obtained by iteratively refining the estimates from Step 1 for the unknowns with all of the original range measurements. Within the TLS framework, the one dimensional iterative (1DI) method is introduced, which applies a SOF method in Step 1 to reduce the number of unknowns to one and the Gauss-Newton method to estimate the single unknown in Step 2. Please note that in Step 2, other ID methods can also be used.

The 1DI method can save up to 67% of the computations, compared to the Gauss-Newton method. In fact, the Newton method can be used in Step 2 to achieve a quadratic rate of convergence, and the required computational load is still low with only one parameter to calculate in each iteration. Validaton work using actual UWB radio measurements shows that its performance is close to the traditional Gauss-Newton method.

7.3 Comparing All Methods

According to the achievable positioning accuracy and the required computational load, the methods considered can be ranked as follows:

ID methods $\approx$ *Constrained iterative methods* $>$ *1DI method* $>$ *MOF II method* $>$ *MOF methods* $>$ *SOF methods.*

Please note that the above conclusion holds as a general trend. In some special cases, the relations can be different. For example, when a very good choice of the quadratic constraint is chosen for the MOF method, it can perform better than the MOF II method, and get very close results as the iterative methods; when the measurements are of very good quality, the 1DI method may take only one or two iterations to converge, and therefore, less computations are involved than the MOF II method. Moreover, the above conclusion holds only for fully synchronized and non-timing based systems. In case of a system with a clock offset, the MOF II and SOF methods perform worse than the MOF method.

Moreover, the ID methods represent general ways to solve non-linear estimation problems, i.e. they can be applied to timing-based range, RSS-range (2.35) and RSS measurements (2.30). While the direct methods, the constrained iterative methods and the 1DI method, where a squaring operation is required, can only be applied to timing-based range and RSS-range measurements, not to RSS measurements.

7.4 NLoS Identification and Mitigation

Four NLoS identification and mitigation schemes have been proposed in Chapter 6, with different computational loads and final positioning accuracies through different rates of correct decision and measurement modeling. All the schemes use both timing-based range and RSS measurements.

7.4.1 Classification Type I

In this scheme, the timing NLoS biases in are treated as unknown parameters. The steps of Classification Type I can be summarized as follows:

1. Under each of the 2^m states, estimate the position and the timing NLoS biases in that maximize the corresponding likelihood function. Here m is the number of transmitter-receiver links.
2. Calculate the GL using the estimates.
3. Choose the final estimates as the ones that correspond to the maximum of the GL over all states.

Classification Type I represents the most general scheme where all the available information is exploited. Highest positioning accuracy is expected among all four schemes. However, the computational load of this scheme is also the highest, since both position coordinates and NLoS biases need to be estimated for all 2^m states.

7.4.2 Classification Type II

The difference here is that, in order to reduce the number of computations required by Classification Type I, the NLoS biases are pre-calculated by a model, and then treated as constants (in both the decision/detection, and the final parameter estimation).

Classification Type II requires less computations than Type I, since no NLoS biases need to be calculated for each of the states. On the other hand, degradation in final positioning accuracy may happen. Moreover, this scheme currently does not work for systems with a clock offset due to 1) the lack of a proper model to calculate NLoS biases in pseudoranges and 2) pseudoranges are not good enough approximations for the actual distances.

7.4.3 Simple Likelihood Ratio Test Type I

The computations involved in Classification Type II are still quite cumbersome, since non-linear least-squares estimation needs to be performed for each of the states. To further reduce the computational load, the SLRT Type I has been introduced, where the NLoS

identification stage is coarried out individually for each link using only the statistics of the RSS measurements, but the actual values required to evaluate the likelihoods of RSS measurements are approximated by the estimates obtained using timing-based ranges. Once the final state is chosen, the position coordinates and NLoS biases are calculated/estimated using both timing-based and RSS measurements.

> *Compared to Classification Type II, SLRT Type I requires only $\frac{1}{2^m-1}$ computations. Degradation in the performance may happen, but the scheme still works very good.*

Please note that the above ratio is only a rough number, due to the fact that an accurate value depends on the correct state, through its actual number of unknown parameters.

7.4.4 Simple Likelihood Ratio Test Type II

The difference between SLRT Type I and Type II is that the latter one directly use the timing-based range measurements, rather than the least-squares estimate, to approximate the actual distance values required to evaluate RSS measurement likelihoods.

> *Compared to SLRT Type I, SLRT Type II saves an extra round of position estimation. Again, degradation on performance may happen. Moreover, this scheme currently does not work with systems with a clock offset due to the fact that pseudoranges are not good enough approximations for the actual distances.*

7.5 Recommendations

Extension to Cooperative Positioning

In this thesis, it has been assumed by default that the information available at the user side includes only the measurements on signals received from multiple transmitters with known positions, and the corresponding measurement statistics. In practice, nodes in systems like wireless sensor networks are capable of pairwise ranging and exchanging information between each other, or at least between close neighboring nodes. This enables the possibility of cooperative positioning, where more information (better measurements availability) can be used to achieve better results with higher positioning accuracy.

It would be interesting to extend this thesis work to fit in a cooperative positioning context. A very simple example of such an extension would be to estimate the unknown positions of users one after another, and treat users with estimated positions as new reference nodes. In this way, positions of later users can be estimated with more measurements and better geometries. For this straightforward example, the methods mentioned in this thesis can all be directly used, whereas in a more complicated case, where positions of all nodes are unknown and measurements are only available at neighboring nodes, additional research work is required.

Bias Check

Validation works showed that the bias check we proposed in Equation (3.69):

$$\|\mu_{\hat{x}}\|_{Q_{\hat{x}\hat{x}}^{-1}} > 1, \tag{7.1}$$

works good with the tested geometries and measurements. However, it is obvious that the significance test depends on the choice of the threshold, currently chosen as '1'. Moreover, if the dimension of the vector $\mu_{\hat{x}}$ increases, the tendency is that the criteria will be violated sooner. To combat these two issues, one can choose the dimension of the vector $\mu_{\hat{x}}$ as the threshold,

$$\begin{aligned} & \|\mu_{\hat{x}}\|_{Q_{\hat{x}\hat{x}}^{-1}} > n \\ \Rightarrow \quad & \frac{1}{n}\|\mu_{\hat{x}}\|_{Q_{\hat{x}\hat{x}}^{-1}} > 1. \end{aligned} \tag{7.2}$$

In this way, the tendency of exceeding the threshold due to an increase in the dimension of $\mu_{\hat{x}}$ is reduced. The validation of (7.2) and finding a better criteria is left for future research.

Connecting Chapter 4-5 to Chapter 6

At current stage, the four schemes proposed in Chapter 6 only work with the ID methods. The reasons are twofold:

1. In case of a state with more than one NLoS bias, there will be more than one quadratic constraints after taking squaring operations on ToA- or pseudo-range measurements. Applying direct methods, constrained iterative methods and the 1DI method in this case requires further adaption.

2. Apart from the timing-based range measurements, the RSS measurements, which involve the distances in a non-linear logarithmic way, are used rather than RSS-ranges. Applying a squaring operation on RSS measurements does not give us a structure as in (4.21), and therefore, the direct methods, the constrained iterative methods and the 1DI method can not be applied.

An intuitively easy solution for Classification Type II and SLRT Type I-II, would be to use RSS-range measurements instead, and the above mentioned methods can all be applied, when the NLoS biases in timing-based ranges are pre-calculated at the final stage. However, one should keep in mind that the obtained least-squares solution is not equivalent to a ML estimate since the distribution of RSS-ranges is log-normal and not elliptically contoured.

Measurement Models

In Chapter 6, we have assumed simple Gaussian distributions for timing-based range and RSS measurements, under both LoS and NLoS. In reality, the NLoS measurements are

reported to follow more complicated distributions, e.g. a Nakagami distribution for NLoS RSS measurements. Performing NLoS identification and mitigation with more complicated models can be one of the future research topics.

Meanwhile, the Classification Type II and the SLRT Type II NLoS identification currently do not work with pseudoranges. As a future research subject, one may think of generating a RSS-difference model, hopefully also Gaussian, as a function of the corresponding actual range-difference. A link-difference can have four states: LoS-LoS, LoS-NLoS, NLoS-LoS and NLoS-NLoS. The situation can be simplified by assuming that there is always at least one LoS link, which corresponds to the largest RSS value, usually also to the smallest pseudorange. Taking this LoS link as the reference, the possible states are reduced to LoS-LoS and NLoS-LoS.

Then Classification Type II can work with pseudorange-difference and RSS-difference measurements, and the pre-calculation of NLoS bias-differences can be done using (2.20), by approximating $d_i - d_r$ with the corresponding pseudorange-difference measurement.

The SLRT Type II can also work by approximating the range-differences in the likelihoods of RSS-differences with the pseudorange-difference measurements.

Computational Load of Classification Type I and II

Currently, the two classification schemes calculate GLs for all possible states. This, however, is not necessary, and some of the states can be skipped based on already evaluated GLs. An easy way can be described as follows:

1. Choosing a state S_k and calculate the corresponding GL_k.
2. Calculate the GLs of S_j's that satisfy $\|S_j - S_k\| = 1,$ for $j = 1, ..., 2^m - 1, j \neq k$.
3. If GL_k is the largest, S_k is the correct state and the corresponding result is the final solution. Otherwise, choose the state that corresponds to the largest GL as the new S_k, then go to Step 2.

The reasoning behind the above steps is that to decide whether a state is correct or not, one only needs to compare it with its *neighboring* states, see Section 6.2.2.1.

When choosing the starting state vector S_i one should consider two important factors:

1. The starting state should be chosen as the one that corresponds to the largest probability of being correct. Anyway, given the assumption that all links are independent, the choice is either S_0 or S_{2^m-1}.
2. The choice should also be made, considering minimizing required computations.

Hence, with current assumption of $P(\text{LoS}) = P(\text{NLoS}) = 1/2$, one should choose the starting state as S_0, since for states with less NLoS biases, the computational load is lower.

Appendix

A.1 Double Differencing Model: Measurement and Error Statistics

Based on (3.42), the expressions $\partial_\alpha y_D(0)$ and $\partial^2_{\alpha^2} y_D(0)$ can be derived as, for $i = 1, ..., m-2$:

$$\begin{aligned}
\partial_\alpha y_{D,i}(0) &= 1 + \frac{x_i^T x_i - x_r^T x_r}{(y_i - y_r)^2}, \ \alpha = 1, ..., m-2 \\
\partial_\alpha y_{D,i}(0) &= -1 - \frac{x_{rr}^T x_{rr} - x_r^T x_r}{(y_{rr} - y_r)^2}, \ \alpha = m-1 \\
\partial_\alpha y_{D,i}(0) &= -\frac{x_i^T x_i - x_r^T x_r}{(y_i - y_r)^2} + \frac{x_{rr}^T x_{rr} - x_r^T x_r}{(y_{rr} - y_r)^2}, \ \alpha = m \\
\partial^2_{\alpha^2} y_{D,i}(0) &= -2\frac{x_i^T x_i - x_r^T x_r}{(y_i - y_r)^3}, \ \alpha = 1, ..., m-2 \\
\partial^2_{\alpha^2} y_{D,i}(0) &= 2\frac{x_{rr}^T x_{rr} - x_r^T x_r}{(y_{rr} - y_r)^3}, \ \alpha = m-1 \\
\partial^2_{\alpha^2} y_{D,i}(0) &= -2\frac{x_i^T x_i - x_r^T x_r}{(y_i - y_r)^3} + \frac{x_{rr}^T x_{rr} - 2x_r^T x_r}{(y_{rr} - y_r)^3}, \ \alpha = m
\end{aligned}$$

And similarly,

$$\begin{aligned}
\partial_\alpha e_{D,i}(0) &= \frac{2d_i}{y_i - y_r}, \ \alpha = 1, ..., m-2 \\
\partial_\alpha e_{D,i}(0) &= -\frac{2d_{rr}}{y_{rr} - y_r}, \ \alpha = m-1 \\
\partial_\alpha e_{D,i}(0) &= -\frac{2d_r}{y_i - y_r} + \frac{2d_r}{y_{rr} - y_r}, \ \alpha = m \\
\partial^2_{\alpha^2} e_{D,i}(0) &= \frac{2}{y_i - y_r} - \frac{4d_i}{(y_i - y_r)^2}, \ \alpha = 1, ..., m-2 \\
\partial^2_{\alpha^2} e_{D,i}(0) &= -\frac{2}{y_{rr} - y_r} + \frac{4d_{rr}}{(y_{rr} - y_r)^2}, \ \alpha = m-1 \\
\partial^2_{\alpha^2} e_{D,i}(0) &= -\frac{2}{y_i - y_r} - \frac{4d_r}{(y_i - y_r)^2} + \frac{2}{y_{rr} - y_r} + \frac{4d_r}{(y_{rr} - y_r)^2}, \ \alpha = m
\end{aligned}$$

A.2 Direct Methods

A.2.1 Error Analysis of SOF TDoA

For the ease of notation, let $P = P_A^{\perp^T} W_D P_A^{\perp}$. Based on (4.75):

$$\partial_{e^T} b_c(0) = \frac{\underline{B}_D^T P \partial_{e^T} \underline{y}_D + \underline{y}_D^T P \partial_{e^T} \underline{B}_D}{\underline{B}_D^T P \underline{B}_D} - 2\frac{\underline{B}_D^T P \partial_{e^T} \underline{B}_D}{(\underline{B}_D^T P \underline{B}_D)^2} \underline{B}_D^T P \underline{y}_D \Big|_{\underline{e}=0} \tag{A.3}$$

$$\begin{aligned} \partial_{ee^T} b_c(0) &= \frac{\partial_e \underline{B}_D^T P \partial_{e^T} \underline{y}_D + \underline{B}_D^T P \partial_{ee^T} \underline{y}_D + \partial_e \underline{y}_D^T P \partial_{e^T} \underline{B}_D}{\underline{B}_D^T P \underline{B}_D} \Big|_{\underline{e}=0} \\ &- 2\frac{\partial_e \underline{B}_D^T P \underline{B}_D}{(\underline{B}_D^T P \underline{B}_D)^2} (\underline{B}_D^T P \partial_{e^T} \underline{y}_D + \underline{y}_D^T P \partial_{e^T} \underline{B}_D) \Big|_{\underline{e}=0} \\ &- 2\frac{\partial_e \underline{B}_D^T P \partial_{e^T} \underline{B}_D}{(\underline{B}_D^T P \underline{B}_D)^2} \underline{B}_D^T P \underline{y}_D + 4\frac{\partial_e \underline{B}_D^T P \underline{B}_D \underline{B}_D^T P \partial_{e^T} \underline{B}_D}{(\underline{B}_D^T P \underline{B}_D)^3} \underline{B}_D^T P \underline{y}_D \Big|_{\underline{e}=0} \\ &- 2\left[\frac{\partial_e \underline{B}_D^T P \underline{B}_D}{(\underline{B}_D^T P \underline{B}_D)^2} (\underline{B}_D^T P \partial_{e^T} \underline{y}_D + \underline{y}_D^T P \partial_{e^T} \underline{B}_D)\right]^T \Big|_{\underline{e}=0} \end{aligned} \tag{A.4}$$

where

$$\partial_{e^T} \underline{B}_D = \begin{bmatrix} 2 & & 0 & -2 \\ & \ddots & & \vdots \\ 0 & & 2 & -2 \end{bmatrix}_{(m-1)\times m}$$

$$\partial_{e^T} \underline{y}_D = \begin{bmatrix} 2d_1 + 2\underline{e}_1 & & 0 & -2d_r - 2\underline{e}_r \\ & \ddots & & \vdots \\ 0 & & 2d_{m-1} + 2\underline{e}_{m-1} & -2d_r - 2\underline{e}_r \end{bmatrix}_{(m-1)\times m}$$

$\partial_{ee^T} \underline{y}_D$ is a tensor of $m \times m \times (m-1)$, and for $k = 1, ..., m-1$

$$[\partial_{ee^T} \underline{y}_D]_{m\times m,k} = \begin{bmatrix} 0 & 0 & 0 \\ & 2 & \vdots \\ & 0 & 0 \\ 0 & 0 & -2 \end{bmatrix}_{m\times m} \quad \text{with } 2 \text{ at } (k,k),\ -2 \text{ at } (m,m)$$

For (4.78) and (4.79), we have

$$K = \mathsf{E}\left\{\underline{B}_D \underline{e}^T\right\} = \begin{bmatrix} 2\sigma_1^2 & & 0 & -2\sigma_r^2 \\ & \ddots & & \vdots \\ 0 & & 2\sigma_{m-1}^2 & -2\sigma_r^2 \end{bmatrix}_{(m-1)\times m} \tag{A.5}$$

$$\begin{aligned} J &= \mathsf{E}\left\{G \underline{e}_D \underline{\mu}_{\hat{b}_c}^T (F - GB_D)^T\right\} \\ &= G \begin{bmatrix} 2d_1\sigma_1^2 & & 0 & -2d_r\sigma_r^2 \\ & \ddots & & \vdots \\ 0 & & 2d_{m-1}\sigma_{m-1}^2 & -2d_r\sigma_r^2 \end{bmatrix} \partial_y M(y) (F - GB_D)^T \end{aligned} \tag{A.6}$$

and $\mathsf{E}\{\underline{e}_D \underline{e}_D^T\}$ can be found in (3.30).

A.2.2 Error Analysis of MOF TDoA

According to the Inverse Function Theorem [108], $J^{-1}[F(x)] = [J(F(x))]^{-1}$, thus, together with (4.80), we have:

$$\begin{aligned}
\partial_i \hat{\underline{x}} &= \left[2\partial_i \underline{D}_D^T W_D \underline{D}_D\right]^{-1}\left[\underline{D}_D^T W_D \underline{y}_D + C_D \hat{\underline{\kappa}}\right] \qquad \text{(A.7)}\\
&+ \left[\underline{D}_D^T W_D \underline{D}_D\right]^{-1}\left[\partial_i \underline{D}_D^T W_D \underline{y}_D + \underline{D}_D^T W_D \partial_i \underline{y}_D + C_D \partial_i \hat{\underline{\kappa}}\right]\\
\partial_{i^2}^2 \hat{\underline{x}} &= \left[2\partial_i \underline{D}_D^T W_D \partial_i \underline{D}_D\right]^{-1}\left[\underline{D}_D^T W_D \underline{y}_D + C_D \hat{\underline{\kappa}}\right] \qquad \text{(A.8)}\\
&+ 2\left[2\partial_i \underline{D}_D^T W_D \underline{D}_D\right]^{-1}\left[\partial_i \underline{D}_D^T W_D \underline{y}_D + \underline{D}_D^T W_D \partial_i \underline{y}_D + C_D \partial_i \hat{\underline{\kappa}}\right]\\
&+ \left[\underline{D}_D^T W_D \underline{D}_D\right]^{-1}\left[2\partial_i \underline{D}_D^T W_D \partial_i \underline{y}_D + \underline{D}_D^T W_D \partial_{i^2}^2 \underline{y}_D + C_D \partial_{i^2}^2 \hat{\underline{\kappa}}\right]
\end{aligned}$$

where $\partial_i \underline{D}_D|_{\underline{e}=0}$ is a $m \times (n+1)$ matrix containing all zero entries, except $[\partial_i \underline{D}_D]_{i,n+1}|_{\underline{e}=0} = 2$; $\partial_i \underline{y}_D\Big|_{\underline{e}=0}$ and $\partial_{i^2}^2 \underline{y}_D\Big|_{\underline{e}=0}$ are $m \times 1$ vectors containing all zero entries, except $[\partial_i \underline{y}_D]_i\Big|_{\underline{e}=0} = 2y_i$ and $[\partial_{i^2}^2 \underline{y}_D]_i\Big|_{\underline{e}=0} = 2$.

Moreover, based on (4.81)-(4.83):

$$\begin{aligned}
\partial_i \hat{\underline{\kappa}}^{\alpha,\beta} &= \frac{1}{2\underline{a}}\left(-\partial_i \underline{b} \pm \frac{-\partial_i \underline{b}}{\sqrt{-2\underline{b}-1}}\right) - \frac{\partial_i \underline{a}}{2\underline{a}^2}(-\underline{b} \pm \sqrt{-2\underline{b}-1}),\\
\partial_{i^2}^2 \hat{\underline{\kappa}}^{\alpha,\beta} &= \frac{1}{2\underline{a}}\left(-\partial_{i^2}^2 \underline{b} \pm \frac{-\partial_{i^2}^2 \underline{b}}{\sqrt{-2\underline{b}-1}} \pm \frac{-(\partial_i \underline{b})^2}{(-2\underline{b}-1)^{-3/2}}\right),\\
&- \frac{\partial_{i^2}^2 \underline{a} - 2(\partial_i \underline{a})^2}{2\underline{a}^2}(-\underline{b} \pm \sqrt{-2\underline{b}-1}) - \frac{\partial_i \underline{a}}{\underline{a}^2}\left(-\partial_i \underline{b} \pm \frac{-\partial_i \underline{b}}{\sqrt{-2\underline{b}-1}}\right)
\end{aligned}$$

with

$$\begin{aligned}
\partial_i \underline{a} &= 2C_D^T W_D \partial_i \underline{D}_D (\underline{D}_D^T W_D \underline{D}_D)^{-1}(\underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D C_D\\
&+ C_D^T W_D \underline{D}_D (2\partial_i \underline{D}_D^T W_D \underline{D}_D)^{-1}(\underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D C_D\\
&+ C_D^T W_D \underline{D}_D (\underline{D}_D^T W_D \underline{D}_D)^{-1}(2\partial_i \underline{D}_D^T W_D \underline{D}_D)^{-1} \partial_i \underline{D}_D^T W_D C_D\\
\partial_{i^2}^2 \underline{a} &= 2C_D^T W_D \partial_i \underline{D}_D (\underline{D}_D^T W_D \underline{D}_D)^{-1}(\underline{D}_D^T W_D \underline{D}_D)^{-1} \partial_i \underline{D}_D^T W_D C_D\\
&+ 4C_D^T W_D \partial_i \underline{D}_D (2\partial_i \underline{D}_D^T W_D \underline{D}_D)^{-1}(\underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D C_D\\
&+ 4C_D^T W_D \partial_i \underline{D}_D (\underline{D}_D^T W_D \underline{D}_D)^{-1}(2\partial_i \underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D C_D\\
&+ C_D^T W_D \underline{D}_D (2\partial_i \underline{D}_D^T W_D \partial_i \underline{D}_D)^{-1}(\underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D C_D\\
&+ 2C_D^T W_D \underline{D}_D (2\partial_i \underline{D}_D^T W_D \underline{D}_D)^{-1}(2\partial_i \underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D C_D\\
&+ C_D^T W_D \underline{D}_D (\underline{D}_D^T W_D \partial_i \underline{D}_D)^{-1}(2\partial_i \underline{D}_D^T W_D \partial_i \underline{D}_D)^{-1} \underline{D}_D^T W_D C_D
\end{aligned}$$

$$\begin{aligned}
\partial_i \underline{b} &= -4C_D^T W_D \partial_i \underline{D}_D (\underline{D}_D^T W_D \underline{D}_D)^{-1}(\underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D \underline{y}_D\\
&- 2C_D^T W_D \underline{D}_D (2\partial_i \underline{D}_D^T W_D \underline{D}_D)^{-1}(\underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D \underline{y}_D\\
&- 2C_D^T W_D \underline{D}_D (\underline{D}_D^T W_D \underline{D}_D)^{-1}(2\partial_i \underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D \underline{y}_D\\
&- 2C_D^T W_D \underline{D}_D (\underline{D}_D^T W_D \underline{D}_D)^{-1}(\underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D \partial_i \underline{y}_D
\end{aligned}$$

$$
\begin{aligned}
\partial_{i^2}^2 \underline{b} &= -4C_D^T W_D \partial_i \underline{D}_D (\underline{D}_D^T W_D \underline{D}_D)^{-1} (\underline{D}_D^T W_D \underline{D}_D)^{-1} \partial_i \underline{D}_D^T W_D \underline{y}_D \\
&- 4C_D^T W_D \partial_i \underline{D}_D (\underline{D}_D^T W_D \underline{D}_D)^{-1} (\underline{D}_D^T W_D \underline{D}_D)^{-1} \partial_i \underline{D}_D^T W_D \partial_i \underline{y}_D \\
&- 8C_D^T W_D \partial_i \underline{D}_D (2\partial_i \underline{D}_D^T W_D \underline{D}_D)^{-1} (\underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D \underline{y}_D \\
&- 8C_D^T W_D \partial_i \underline{D}_D (\underline{D}_D^T W_D \underline{D}_D)^{-1} (2\partial_i \underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D \underline{y}_D \\
&- 2C_D^T W_D \underline{D}_D (2\partial_i \underline{D}_D^T W_D \partial_i \underline{D}_D)^{-1} (\underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D \underline{y}_D \\
&- 4C_D^T W_D \underline{D}_D (2\partial_i \underline{D}_D^T W_D \underline{D}_D)^{-1} (2\partial_i \underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D \underline{y}_D \\
&- 4C_D^T W_D \underline{D}_D (2\partial_i \underline{D}_D^T W_D \underline{D}_D)^{-1} (\underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D \partial_i \underline{y}_D \\
&- 2C_D^T W_D \underline{D}_D (\underline{D}_D^T W_D \underline{D}_D)^{-1} (2\partial_i \underline{D}_D^T W_D \partial_i \underline{D}_D)^{-1} \underline{D}_D^T W_D \underline{y}_D \\
&- 4C_D^T W_D \underline{D}_D (\underline{D}_D^T W_D \underline{D}_D)^{-1} (2\partial_i \underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D \partial_i \underline{y}_D \\
&- 4C_D^T W_D \partial_i \underline{D}_D (\underline{D}_D^T W_D \underline{D}_D)^{-1} (\underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D \partial_i \underline{y}_D \\
&- 2C_D^T W_D \underline{D}_D (\underline{D}_D^T W_D \underline{D}_D)^{-1} (\underline{D}_D^T W_D \underline{D}_D)^{-1} \underline{D}_D^T W_D \partial_{i^2}^2 \underline{y}_D
\end{aligned}
$$

A.2.3 Error Analysis of MOF Type II TDoA

Based on (5.41), which yields, for $i = 1, ..., m$:

$$
\begin{aligned}
\partial_i \hat{\underline{x}}_D &= \left[2\partial_i \underline{F}_D^T W_D \underline{F}_D + \partial_i \hat{\underline{\lambda}}(L - ss^T)\right]^{-1} \left[\underline{F}_D^T W_D \underline{y}_D + \hat{\underline{\lambda}}(L \breve{\underline{x}}_r + s \hat{\underline{y}}_r)\right] \qquad \text{(A.9)} \\
&+ \left[\underline{F}_D^T W_D \underline{F}_D + \hat{\underline{\lambda}}(L - ss^T)\right]^{-1} \\
&\cdot \left[\partial_i \underline{F}_D^T W_D \underline{y}_D + \underline{F}_D^T W_D \partial_i \underline{y}_D + \partial_i \hat{\underline{\lambda}}(L \breve{\underline{x}}_r + s \hat{\underline{y}}_r) + \hat{\underline{\lambda}} s \partial_i \hat{\underline{y}}_r\right] \\
\partial_{i^2}^2 \hat{\underline{x}}_D &= \left[2\partial_i \underline{F}_D^T W_D \partial_i \underline{F}_D + \partial_{i^2}^2 \hat{\underline{\lambda}}(L - ss^T)\right]^{-1} \left[\underline{F}_D^T W_D \underline{y}_D + \hat{\underline{\lambda}}(L \breve{\underline{x}}_r + s \hat{\underline{y}}_r)\right] \qquad \text{(A.10)} \\
&+ 2\left[2\partial_i \underline{F}_D^T W_D \underline{F}_D + \partial_i \hat{\underline{\lambda}}(L - ss^T)\right]^{-1} \\
&\cdot \left[\partial_i \underline{F}_D^T W_D \underline{y}_D + \underline{F}_D^T W_D \partial_i \underline{y}_D + \partial_i \hat{\underline{\lambda}}(L \breve{\underline{x}}_r + s \hat{\underline{y}}_r) + \hat{\underline{\lambda}} s \partial_i \hat{\underline{y}}_r\right] \\
&+ \left[\underline{F}_D^T W_D \underline{F}_D + \hat{\underline{\lambda}}(L - ss^T)\right]^{-1} \\
&\cdot \left[2\partial_i \underline{F}_D^T W_D \partial_i \underline{y}_D + \underline{F}_D^T W_D \partial_{i^2}^2 \underline{y}_D + \partial_{i^2}^2 \hat{\underline{\lambda}}(L \breve{\underline{x}}_r + s \hat{\underline{y}}_r) + 2\partial_i \hat{\underline{\lambda}} s \partial_i \hat{\underline{y}}_r + \hat{\underline{\lambda}} s \partial_{i^2}^2 \hat{\underline{y}}_r\right]
\end{aligned}
$$

where $\partial_i \underline{F}_D|_{\underline{e}=0}$ is a $m \times (n+1)$ matrix containing all zero entries, except $[\partial_i \underline{F}_D]_{i,n+1}|_{\underline{e}=0} = 2$, for $i \neq r$ and $[\partial_i \underline{F}_D]_{r,n+1}|_{\underline{e}=0} = -2$; $\partial_i \underline{y}_D\Big|_{\underline{e}=0}$ and $\partial_{i^2}^2 \underline{y}_D\Big|_{\underline{e}=0}$ are $(m-1) \times 1$ vectors containing all zero entries, except $[\partial_i \underline{y}_D]_i\Big|_{\underline{e}=0} = 2y_i$ and $[\partial_{i^2}^2 \underline{y}_D]_i\Big|_{\underline{e}=0} = 2$ for $i \neq r$ and $[\partial_i \underline{y}_D]_r\Big|_{\underline{e}=0} = -2y_r$ and $[\partial_{i^2}^2 \underline{y}_D]_r\Big|_{\underline{e}=0} = -2$.

Moreover, based on (5.21)-(5.23), $\hat{\underline{\lambda}}$ and $\hat{\underline{y}}_r$ contain only $\underline{e}_r$. Thus, for $i \neq r$, $\partial_i \hat{\underline{\lambda}} = 0$ and

$\partial_i \underline{\hat{y}}_r = 0$, and for $i = r$:

$$\begin{aligned}
\partial_i \underline{\hat{\lambda}}^{\alpha} &= \partial^2_{i^2} \underline{\hat{\lambda}}^{\alpha} = 0, \\
\partial_i \underline{\hat{\lambda}}^{\beta} &= -2wy_r - w\left(y_r^2 + \frac{2w''}{w}\right)^{\frac{1}{2}} - wy_r^2\left(y_r^2 + \frac{2w''}{w}\right)^{-\frac{1}{2}} \\
\partial^2_{i^2} \underline{\hat{\lambda}}^{\beta} &= -2w - 3wy_r\left(y_r^2 + \frac{2w''}{w}\right)^{-\frac{1}{2}} + wy_r^3\left(y_r^2 + \frac{2w''}{w}\right)^{-\frac{3}{2}} \\
\partial_i \underline{\hat{\lambda}}^{\gamma} &= -2wy_r + w\left(y_r^2 + \frac{2w''}{w}\right)^{\frac{1}{2}} + wy_r^2\left(y_r^2 + \frac{2w''}{w}\right)^{-\frac{1}{2}} \\
\partial^2_{i^2} \underline{\hat{\lambda}}^{\gamma} &= -2w + 3wy_r\left(y_r^2 + \frac{2w''}{w}\right)^{-\frac{1}{2}} - wy_r^3\left(y_r^2 + \frac{2w''}{w}\right)^{-\frac{3}{2}}
\end{aligned}$$

and

$$\begin{aligned}
\partial_i \underline{\hat{y}}_r^{\alpha} &= 1 \;\; \partial^2_{i^2} \underline{\hat{y}}_r^{\alpha} = 0, \\
\partial_i \underline{\hat{y}}_r^{\beta} &= \frac{1}{2}\left[-1 + y_r\left(y_r^2 + \frac{2w''}{w}\right)^{-\frac{1}{2}}\right] \\
\partial^2_{i^2} \underline{\hat{y}}_r^{\beta} &= \frac{1}{2}\left[\left(y_r^2 + \frac{2w''}{w}\right)^{-\frac{1}{2}} - y_r^2\left(y_r^2 + \frac{2w''}{w}\right)^{-\frac{3}{2}}\right] \\
\partial_i \underline{\hat{y}}_r^{\gamma} &= \frac{1}{2}\left[-1 - y_r\left(y_r^2 + \frac{2w''}{w}\right)^{-\frac{1}{2}}\right] \\
\partial^2_{i^2} \underline{\hat{y}}_r^{\gamma} &= \frac{1}{2}\left[-\left(y_r^2 + \frac{2w''}{w}\right)^{-\frac{1}{2}} + y_r^2\left(y_r^2 + \frac{2w''}{w}\right)^{-\frac{3}{2}}\right]
\end{aligned}$$

A.3 Iterative Methods

A.3.1 EVD Problem of Constrained Iterative Methods

Using (4.95), we have,

$$\begin{aligned}
\partial_\lambda \hat{x}_D &= [\partial_\lambda(\underline{F}_D^T W_D \underline{F}_D + L\lambda)^{-1}](\underline{F}_D^T W_D \underline{y}_D - \frac{1}{2}\lambda s) \\
&+ (\underline{F}_D^T W_D \underline{F}_D + L\lambda)^{-1}(\underline{F}_D^T W_D \underline{y}_D - \frac{1}{2}s),
\end{aligned} \tag{A.11}$$

which can be elaborated by partitioning the following two matrices as:

$$\begin{aligned}
\underline{F}_D^T W_D \underline{F}_D &= \begin{bmatrix} A_{(n+1)\times(n+1)} & b_{(n+1)\times 1} \\ b^T_{1\times(n+1)} & c_{1\times 1} \end{bmatrix}_{(n+2)\times(n+2)}, \\
L &= \begin{bmatrix} L'_{(n+1)\times(n+1)} & 0_{(n+1)\times 1} \\ 0_{1\times(n+1)} & 0_{1\times 1} \end{bmatrix}_{(n+2)\times(n+2)},
\end{aligned} \tag{A.12}$$

the expression for $(\underline{F}_D^T W_D \underline{F}_D + L\lambda)^{-1}$ can then be derived based on (4.44) as:

$$(\underline{F}_D^T W_D \underline{F}_D + L\lambda)^{-1} = \begin{bmatrix} (A - L'\lambda + bc^{-1}b^T)^{-1} & -(A - L'\lambda - bc^{-1}b^T)^{-1}bc^{-1} \\ -c^{-1}b^T(A - L'\lambda - bc^{-1}b^T)^{-1} & c^{-1} + c^{-1}b^T(A - L'\lambda - bc^{-1}b^T)^{-1}bc^{-1} \end{bmatrix} \tag{A.13}$$

According to the Inverse Function Theorem [108], $J^{-1}[F(x)] = [J(F(x))]^{-1}$, with $J(.)$ denoting the Jacobian, thus,

$$\partial_\lambda (\underline{F}_D^T W_D \underline{F}_D + L\lambda)^{-1} = \begin{bmatrix} -L'^{-1} & L'^{-1}bc^{-1} \\ c^{-1}b^T L'^{-1} & -c^{-1}b^T L'^{-1}bc^{-1} \end{bmatrix} \tag{A.14}$$

which completes the details of (A.11).

A.3.2 Error Analysis for 1DI Method

In TDoA cases, $z = b_c$ and the objective function to minimize is,

$$F_{1D}(b_c) = \left\| \underline{y} - P(b_c, \underline{e}) \right\|_W^2, \tag{A.15}$$

where $P_i(b_c, \underline{e}) = \|E\underline{y}_D - E\underline{B}_D b_c - x_i\| + b_c$, and the problem is equivalent to solving

$$\partial_{b_c} P(\hat{\underline{b}}_c, \underline{e})^T W [\underline{y} - P(\hat{\underline{b}}_c, \underline{e})] = 0. \tag{A.16}$$

The terms $\partial_y M(y)$ and $\partial^2_{yy^T} M(y)$ can be calculated by first Taylorizing (A.16) with respect to $\hat{\underline{b}}_c$ at the true value b_c,

$$\begin{aligned} 0 &= \partial_{b_c} P(\hat{\underline{b}}_c, \underline{e})^T W [\underline{y} - P(\hat{\underline{b}}_c, \underline{e})] = \partial_{b_c} P(b_c, \underline{e})^T W [\underline{y} - P(b_c, \underline{e})] \\ &+ \left[\partial^2_{b_c^2} P(b_c, \underline{e})^T W [\underline{y} - P(b_c, \underline{e})] - \partial_{b_c} P(b_c, \underline{e})^T W \partial_{b_c} P(b_c, \underline{e}) \right] (\hat{\underline{b}}_c - b_c) \\ &+ \frac{1}{2} \Big\{ \partial^3_{b_c^3} P(b_c, \underline{e})^T W [\underline{y} - P(b_c, \underline{e})] - \partial^2_{b_c^2} P(b_c, \underline{e})^T W \partial_{b_c} P(b_c, \underline{e}) \\ &- \partial^2_{b_c^2} P(b_c, \underline{e})^T W \partial_{b_c} P(b_c, \underline{e}) - \partial_{b_c} P(b_c, \underline{e})^T W \partial^2_{b_c^2} P(b_c, \underline{e}) \Big\} (\hat{\underline{b}}_c - b_c)^2 + \cdots \end{aligned} \tag{A.17}$$

Then Taylorize $P(b_c, \underline{e})$ with respect to $\underline{e}$ at $E\{\underline{e}\} = 0$, which gives,

$$\begin{aligned} P_i(b_c, \underline{e}) &= P_i(b_c, 0) + \partial_e P_i(b_c, 0)^T \underline{e} + \frac{1}{2} \underline{e}^T \partial^2_{ee} P_i(b_c, 0) \underline{e} + \cdots \\ &= A_i(b_c) + R_i(b_c)\underline{e} + \frac{1}{2} \underline{e}^T S_i(b_c) \underline{e} + \cdots \end{aligned} \tag{A.18}$$

where we denote

$$[R_i]_{1\times m} = \partial_e P_i(b_c, 0)^T = \frac{(Ey_D - EB_D b_c - x_i)^T \left(E\partial_{e^T} \underline{y}_D - Eb_c \partial_{e^T} \underline{B}_D \right) \Big|_{\underline{e}=0}}{\|Ey_D - EB_D b_c - x_i\|}$$

$$[\partial_{e^T} \underline{y}_D - b_c \partial_{e^T} \underline{B}_D]_{(m-1)\times m} = \begin{bmatrix} 2d_1 + 2\underline{e}_1 & & 0 & -2d_r - 2\underline{e}_r \\ & \ddots & & \vdots \\ 0 & & 2d_{m-1} + 2\underline{e}_{m-1} & -2d_r - 2\underline{e}_r \end{bmatrix}$$

$$
\begin{aligned}
[S_i]_{m\times m} &= \partial^2_{ee} P_i(b_c, 0) \\
&= \frac{(E\partial_{e^T}\underline{y}_D - Eb_c\partial_{e^T}\underline{B}_D)^T(E\partial_{e^T}\underline{y}_D - Eb_c\partial_{e^T}\underline{B}_D)\Big|_{\underline{e}=0} + H}{\|Ey_D - EB_Db_c - x_i\|} \\
&- \frac{(E\partial_{e^T}\underline{y}_D - Eb_c\partial_{e^T}\underline{B}_D)^T(Ey_D - EB_Db_c - x_i)\Big|_{\underline{e}=0}}{\|Ey_D - EB_Db_c - x_i\|^3} \\
&\cdot \quad (Ey_D - EB_Db_c - x_i)^T(E\partial_{e^T}\underline{y}_D - Eb_c\partial_{e^T}\underline{B}_D)\Big|_{\underline{e}=0} \\
H_{m\times m} &= (Ey_D - Eb_cB_D - x_i)^T E\partial^2_{ee}(\underline{y}_D - b_c\underline{B}_D)\Big|_{\underline{e}=0} \\
&= 2diag\left\{\left[(Ey_D - Eb_cB_D - x_i)^T E,\ -\sum_{j=1}^{n-1}[(Ey_D - Eb_cB_D - x_i)^T E]_j\right]\right\}
\end{aligned}
$$

for simplicity.

Since (A.17) is equal to zero for all $\underline{e}$, we can collect terms with the same order and set them to zero. Plugging (5.68) and (A.18) into (A.17) and collect all the first order terms gives

$$
\partial_y M(y)^T = \left(\partial_{b_c}A(b_c)^T W \partial_{b_c}A(b_c)\right)^{-1}\partial_{b_c}A(b_c)^T W\,(I - R)\,, \tag{A.19}
$$

and collecting the second order terms gives us

$$
\begin{aligned}
\partial^2_{yy^T}M(y) &= \left(\partial_{b_c}A(b_c)^T W \partial_{b_c}A(b_c)\right)^{-1}\left\{-\sum_{i=1}^{m}\left\{[\partial_{b_c}A(b_c)^T W]_i\partial_{b_c}S_i\right\}\right. \\
&+ 2\partial_{b_c}R^T W(I - R) + 2(I - R)^T W\partial^2_{b_c^2}A(b_c)\partial_y M(y)^T \\
&- \left. 4R^T W\partial_{b_c}A(b_c)\partial_y M(y)^T - \partial_y M(y)\left[3\partial_{b_c}A(b_c)^T W\partial^2_{b_c^2}A(b_c)\right]\partial_y M(y)^T\right\}.
\end{aligned}
$$

Furthermore,

$$
K = \mathsf{E}\left\{\underline{B}_D\underline{e}^T\right\} = \begin{bmatrix} -2\sigma_1^2 & & 0 & 2\sigma_r^2 \\ & \ddots & & \vdots \\ 0 & & -2\sigma_{m-1}^2 & 2\sigma_r^2 \end{bmatrix}_{(m-1)\times m} \tag{A.20}
$$

$$
\begin{aligned}
J &= \mathsf{E}\left\{G\underline{e}_D\underline{\mu}^T_{\hat{b}_c}(F - GB_D)^T\right\} \\
&= G\begin{bmatrix} 2d_1\sigma_1^2 & & 0 & -2d_r\sigma_r^2 \\ & \ddots & & \vdots \\ 0 & & 2d_{m-1}\sigma_{m-1}^2 & -2d_r\sigma_r^2 \end{bmatrix}\partial_y M(y)(F - GB_D)^T
\end{aligned} \tag{A.21}
$$

Note that the third and higher order terms of $\underline{e}$ are ignored.

A.3.3 Flop Count for 1DI Method

A.3.3.1 TDoA

In each iteration, the common terms to update for both methods are d^0 and $A(x^0) = P(b_c^0, e)$. To update a single element d_i^0,

$$d_i^0 = \underbrace{\sqrt{\underbrace{(x_u^0 - x_i)^T \underbrace{(x_u^0 - x_i)}_{n \text{ flops}}}_{n+2n-1 \text{ flops}}}}_{\text{Total } n+2n-1+1 \text{ flops}}, \tag{A.22}$$

$3n$ flops are required. Therefore, for m-D vector d^0, $3mn$ flops are required.

The term $A(x^0) = P(b_c^0, e)$ can be updated based on d^0 and b_c^0 , which requires m flops.

1DI: The calculations in each iteration of the 1DI method include updating the vector $\partial_{b_c} P(b_c^0, e)$ and Eq. (5.61). For a single element $\partial_{b_c} P_i(b_c^0, e)$,

$$\partial_{b_c} P_i(b_c^0, e) = \underbrace{\frac{\overbrace{(Ey_L - x_i)^T E\delta y_r + \overbrace{(E\delta y_r)^T E\delta y_r b_c^0}^{1 \text{ flop}}}^{1+1 \text{ flops}}}{d_i^0} + 1}_{\text{Total } 1+1+1+1 \text{ flops}}, \tag{A.23}$$

only 4 flops are required. Note that the calculations for the constant terms $(Ey_L - x_i)^T E\delta y_r$ and $(E\delta y_r)^T E\delta y_r$ have been done in the first iteration and do not need to be repeated again. Therefore, $4m$ flops are needed to update $\partial_{b_c} P(b_c^0, e)$.

Once $P(b_c^0, e)$ and $\partial_{b_c} P(b_c^0, e)$ are updated, the flops required in Eq. (5.61) include, $(2m-1)m$ for $\partial_{b_c} P^T(b_c^0, e) \cdot W$, m for $y - P(b_c^0, e)$, $2m-1$ for $\partial_{b_c} P^T(b_c^0, e) W \cdot (y - d^0)$, and $2m-1$ for $\partial_{b_c} P^T(b_c^0, e) W \cdot \partial_{b_c} P(b_c^0, e)$. Then 1 flop is used to calculate $(\partial_{b_c} P^T(b_c^0, e) W \partial_{b_c} P(b_c^0, e))^{-1}$, 1 flop to multiply it with $\partial_{b_c} P^T(b_c^0, e) W (y - d^0)$, and 1 flop to add b_c^0. The stopping criteria (5.62) requires the evaluation of $\|\hat{b}_c - b_c^0\|^2$, and since $\hat{b}_c - b_c^0$ has already been evaluated in previous steps, only 1 flop is required. Finally, the update of $\underline{\hat{x}}_u = E\underline{y}_L + E\delta y_r \hat{b}_c$, for the evaluation of d^0, requires 2n flops.

Therefore, in each iteration, the number of required flops for the 1DI method, including the common terms, reads

$$\text{flops}_{1DI} = 2m^2 + 3mn + 9m + 2n + 2. \tag{A.24}$$

If W is simply a diagonal matrix, with unequal diagonal entries, only m flops are required to evaluate $\partial_{b_c} P^T(b_c^0, e) \cdot W$ rather than $(2m-1)m$, and the total number of flops becomes

$$\text{flops}'_{1DI} = 3mn + 11m + 2n + 2. \tag{A.25}$$

Gauss-Newton: In each iteration, the calculations for the Gauss-Newton method include updating part of the $m \times (n+1)$ matrix A and the equation

$$(A^T W A)(\hat{x}_u - x_u^0) = A^T W (y - A(x^0)). \tag{A.26}$$

The expression for a single element $A_{ij}, i = 1, ..., m$ and $j = 1, ..., n$ is

$$A_{ij} = \frac{x_{uj} - x_{ij}}{d_i^0}, \tag{A.27}$$

which requires 2 flops. Note that the last column of A contains only constants. For the whole matrix A, $2mn$ flops are needed.

To obtain $\hat{x}$ from (A.26), the explicit computation of the inversion of $A^T W A$ is usually avoided by first applying a triangular decomposition: $A^T W A = LL^T$, e.g. Cholesky decomposition, which requires $n^3/3$ flops [92]. Denoting $r = A^T W (y - A(x^0))$, one has

$$L\hat{z} = r \text{ and } L^T \hat{x} = \hat{z}. \tag{A.28}$$

Since L is lower triangular, the first entry of $\hat{z}$ can be computed as $\hat{z}_1 = r_1/L_{11}$. This result can be used to compute the second entry as $\hat{z}_2 = (r_2 - L_{21}\hat{z}_1)/L_{22}$. In this way one continues to obtain $\hat{z}$. For a $n+1$-D vector $\hat{z}$, the required flops is $(n+1)^2$. To obtain $\hat{x}$ from $\hat{z}$, the same number of flops is required.

Hence, to calculate $\hat{x}_u$ one needs, $n(2m-1)m$ for $A^T \cdot W$, m for $y - A(x^0)$, $(n+1)(2m-1)$ for $A^T W \cdot (y - A(x^0))$, $n(2m-1)(n+1)$ for $A^T W \cdot A$, $(n+1)^3/3$ for Cholesky decomposition, $2(n+1)^2$ for $\hat{x} - x^0$ and $n+1$ for $\hat{x}_u$. The stopping criteria requires the evaluation of $\|\hat{x} - x^0\|^2$ and since $\hat{x} - x^0$ has already been calculated in previous steps, the number of required flops is $2n+1$.

The total number of flops, including the common term, read

$$\text{flops}_{GN} = n^3/3 + 2m^2 n + 2mn^2 + 2n^2 + 7mn + 4m + 6n + 4/3. \tag{A.29}$$

If W is simply a diagonal matrix, with unequal diagonal entries, only nm flops are required to evaluate $A^T \cdot W$ rather than $n(2m-1)m$, and the total number of flops becomes

$$\text{flops}'_{GN} = n^3/3 + 2mn^2 + 2n^2 + 9mn + 4m + 6n + 4/3. \tag{A.30}$$

A.3.3.2 ToA/RSS

In each iteration, the common term to update is the m-D vector $A(x^0) = P(b_c^0, e) = d^0$, and $3mn$ flops are required.

1DI: Compared to the TDoA cases, the difference is only the vector $\partial_{b_c} P(b_c^0, e)$, which now requires only $3m$ flops.

Therefore, in each iteration, the number of required flops for the 1DI method, including the common term, read

$$\text{flops}_{1DI} = 2m^2 + 3mn + 7m + 2n + 2, \tag{A.31}$$

If W is diagonal the total number of flops becomes

$$\text{flops}'_{1DI} = 3mn + 9m + 2n + 2. \tag{A.32}$$

Gauss-Newton: To update A, $2mn$ flops are needed, the same as for TDoA cases. To calculate $\hat{x}_u$, $n(2m-1)m$ is needed for $A^T \cdot W$, m for $y-d^0$, $n(2m-1)$ for $A^T W \cdot (y-d^0)$, $n(2m-1)n$ for $A^T W \cdot A$, $(n)^3/3$ for Cholesky decomposition, $2n^2$ for $\hat{x} - x^0$, and n for $\hat{x}$. The stopping criteria requires $2n-1$.

The total number of flops, including the common term, read

$$\text{flops}_{GN} = n^3/3 + 2m^2n + 2mn^2 + n^2 + 6mn + m + 2n - 1. \tag{A.33}$$

If W is diagonal the total number of flops becomes

$$\text{flops}'_{GN} \;=\; n^3/3 + 2mn^2 + n^2 + 8mn + m + 2n - 1. \tag{A.34}$$

A.4 NLoS Identification and Mitigation

A.4.1 Upper and Lower Bounds of $T_{k,j}$ in Classification Type I

Concerning two hypotheses H_j and H_k, $j \neq k$, the following notations will be used:

$X = [x_u^T, \nabla_S^T]^T$	the parameter vector that contains the common unknown parameters of H_j and H_k, i.e., $X \in \Phi_j \cap \Phi_k$
∇_S	the common NLoS bias vector, i.e., $\nabla_S \in (\Phi_j \cap \Phi_k) \backslash \Phi_0$
C_S	the state matrix corresponding to ∇_S
$\nabla_{D,j}$	the bias vector containing non-common biases under H_j compared to H_k, i.e., $\nabla_{D,j} \in \Phi_j \backslash (\Phi_j \cap \Phi_k)$
$\nabla_{D,k}$	the bias vector containing non-common biases under H_k compared to H_j, i.e., $\nabla_{D,k} \in \Phi_k \backslash (\Phi_j \cap \Phi_k)$
$C_{D,j}$	the state matrix corresponds to $\nabla_{D,j}$
$C_{D,k}$	the state matrix corresponds to $\nabla_{D,k}$

Suppose $m = 5$, $j = 5$ and $k = 11$, then $S_j = [0,0,1,0,1]^T$, $S_k = [0,1,0,1,1]^T$ and the above terms can be evaluated as:

$$\begin{aligned}
X &= [x_u^T, \mu_5] \\
\nabla_S &= \mu_5 \\
\nabla_{D,j} &= \mu_3 \\
\nabla_{D,k} &= [\mu_2, \mu_4]^T \\
C_S &= [0,0,0,0,1,0,0,0,0,0]^T \\
C_{D,j} &= [0,0,1,0,0,0,0,0,0,0]^T \\
C_{D,k} &= \begin{bmatrix} 0,1,0,0,0,0,0,0,0,0 \\ 0,0,0,1,0,0,0,0,0,0 \end{bmatrix}^T
\end{aligned}$$

According to (6.59), assuming H_k true, the calculation of γ_k only involves the hypotheses H_j's for $j \in \{j \mid \|S_j - S_k\| = 1\}$. Hence, S_j and S_k are different in only 1 link state. Without loss of generality, we assume that S_k contains one extra NLoS link than S_j, and therefore, we have $C_{D,j} = 0$.

Then, with the above new notation, the approximations in (6.63)-(6.65) can be rewritten as:

$$\begin{aligned}
\hat{X}_j^T &\approx X + \left(A_j^T Q_j^{-1} A_j\right)^{-1} A_j^T Q_j^{-1}[y_{\mathsf{all}} - F_j(X)] && \text{(A.35)}\\
\hat{y}_{\mathsf{all},j} &= F_j(\hat{X}_j) \approx F_j(X) + A_j(\hat{X}_j - X) && \text{(A.36)}\\
&= F_j(X) + A_j \left(A_j^T Q_j^{-1} A_j\right)^{-1} A_j^T Q_j^{-1}[y_{\mathsf{all}} - F_j(X)] \\
\hat{e}_j &= y_{\mathsf{all}} - \hat{y}_{\mathsf{all},j} && \text{(A.37)}\\
&\approx \left(I - A_j \left(A_j^T Q_j^{-1} A_j\right)^{-1} A_j^T Q_j^{-1}\right)[y_{\mathsf{all}} - F_j(X)]
\end{aligned}$$

where $A_j = \partial_{X^T} F_j(X)$, and X is the true value. And for H_k, we have

$$\begin{aligned}
&[\hat{X}_k^T, \hat{\nabla}_{D,k}]^T && \text{(A.38)}\\
\approx\ & [X^T, 0]^T + \left(\begin{bmatrix} A_k^T \\ C_{D,k}^T \end{bmatrix} Q_k^{-1}[A_k, C_{D,k}]\right)^{-1} \begin{bmatrix} A_k^T \\ C_{D,k}^T \end{bmatrix} Q_k^{-1}[y_{\mathsf{all}} - F_k(X)] \\
\hat{y}_{\mathsf{all},k} &= F_k(\hat{X}_k, C_{D,k}\hat{\nabla}_{D,k}) \approx F_k(X) + \begin{bmatrix} A_k^T \\ C_{D,k}^T \end{bmatrix} [(\hat{X}_k - X)^T, \hat{\nabla}_{D,k}]^T && \text{(A.39)}\\
&= [A_k, C_{D,k}] \left(\begin{bmatrix} A_k^T \\ C_{D,k}^T \end{bmatrix} Q_k^{-1}[A_k, C_{D,k}]\right)^{-1} \begin{bmatrix} A_k^T \\ C_{D,k}^T \end{bmatrix} Q_k^{-1}[y_{\mathsf{all}} - F_k(X)] \\
\hat{e}_k &= y_{\mathsf{all}} - \hat{y}_{\mathsf{all},k} && \text{(A.40)}\\
&\approx \left(I - [A_k, C_{D,k}] \left(\begin{bmatrix} A_k^T \\ C_{D,k}^T \end{bmatrix} Q_k^{-1}[A_k, C_{D,k}]\right)^{-1} \begin{bmatrix} A_k^T \\ C_{D,k}^T \end{bmatrix} Q_k^{-1}\right)[y_{\mathsf{all}} - F_k(X)]
\end{aligned}$$

with $A_k = \partial_{X^T} F_k(X)$. Please note that $F_k(X)$ is obtained by setting $C_{D,k} = 0$ and $\nabla_k = 0$ in $F_k(x_u, C_k\nabla_k)$. Moreover, we have: $A_k = GA_j$, G a diagonal matrix with all diagonal entries equal to 1, except the $i+m$-th diagonal entry equal to $\frac{n_{\mathsf{NLoS}}}{n_{\mathsf{LoS}}}$:

$$G = \begin{bmatrix} 1 & & \cdots & & 0 \\ & \ddots & & & \\ & & \overbrace{\frac{n_{\mathsf{NLoS}}}{n_{\mathsf{LoS}}}}^{i\text{-th}} & & \\ \vdots & & & & \vdots \\ & & & \ddots & \\ 0 & & \cdots & & 1 \end{bmatrix}_{2m\times 2m}. \qquad \text{(A.41)}$$

where we assume that the i-th is the extra NLoS link under H_k, as compared to H_j

Recall from (6.62), the test variable:

$$\begin{aligned}
T_{k,j} &= \hat{e}_j^T Q_j^{-1} \hat{e}_j - \hat{e}_k^T Q_k^{-1} \hat{e}_k && \text{(A.42)}\\
&= (\hat{e}_j - G^{-1}\hat{e}_k)^T Q^{-1} (\hat{e}_j - G^{-1}\hat{e}_k) + 2(\hat{e}_j - G^{-1}\hat{e}_k)^T Q^{-1} \hat{e}_k \\
&+ \hat{e}_j^T (Q_j^{-1} - Q^{-1}) \hat{e}_j
\end{aligned}$$

where $Q^{-1} = G^T Q_k^{-1} G$.

With $\hat{e}_j = y_{\text{all}} - \hat{y}_{\text{all},j}$, $\hat{e}_k = y_{\text{all}} - \hat{y}_{\text{all},k}$, we have:

$$\begin{aligned} T_{k,j} &= \left((1-G^{-1})y_{\text{all}} + G^{-1}\hat{y}_{\text{all},k} - \hat{y}_{\text{all},j}\right)^T Q^{-1} \left((1-G^{-1})y_{\text{all}} + G^{-1}\hat{y}_{\text{all},k} - \hat{y}_{\text{all},j}\right) \\ &- 2\left((1-G^{-1})y_{\text{all}} + G^{-1}\hat{y}_{\text{all},k} - \hat{y}_{\text{all},j}\right)^T Q^{-1}G^{-1}\hat{e}_k + \hat{e}_j^T(Q_j^{-1} - Q^{-1})\hat{e}_j \end{aligned} \quad \text{(A.43)}$$

For H_j, (A.35) can be rewritten as:

$$A_j^T Q_j^{-1} A_j(\hat{X}_j - X) \approx A_j^T Q_j^{-1}[y_{\text{all}} - F_j(X)] \quad \text{(A.44)}$$

and we have for H_k,

$$\begin{aligned} &\begin{bmatrix} A_k^T Q_k^{-1} A_k & A_k^T Q_k^{-1} C_{D,k} \\ C_{D,k}^T Q_k^{-1} A_k & C_{D,k}^T Q_k^{-1} C_{D,k} \end{bmatrix} \begin{bmatrix} \hat{X}_k - X \\ \hat{\nabla}_{D,k} \end{bmatrix} = \begin{bmatrix} A_k^T Q_k^{-1}[y_{\text{all}} - F_k(X)] \\ C_{D,k}^T Q_k^{-1}[y_{\text{all}} - F_k(X)] \end{bmatrix} \\ \Leftrightarrow\ &\begin{bmatrix} A_j^T Q^{-1} A_j & A_j^T Q^{-1} C_{D,k} \\ C_{D,k}^T Q^{-1} A_j & C_{D,k}^T Q^{-1} C_{D,k} \end{bmatrix} \begin{bmatrix} \hat{X}_k - X \\ \hat{\nabla}_{D,k} \end{bmatrix} = \begin{bmatrix} A_j^T Q^{-1} G^{-1}[y_{\text{all}} - F_k(X)] \\ C_{D,k}^T Q^{-1} G^{-1}[y_{\text{all}} - F_k(X)] \end{bmatrix} \\ =\ &\begin{bmatrix} A_j^T Q^{-1} A_j(\hat{X}_j - X) + A_j^T Q^{-1}(G^{-1} - I)y_{\text{all}} - A_j^T(Q^{-1} - Q_j^{-1})\hat{e}_j \\ C_{D,k}^T Q^{-1} G^{-1}[y_{\text{all}} - F_k(X)] \end{bmatrix} \end{aligned} \quad \text{(A.45)}$$

Pre-multiplication with $(A_j^T Q^{-1} A_j)^{-1}$ on the first line of (A.45) gives:

$$\hat{X}_k - \hat{X}_j = (A_j^T Q^{-1} A_j)^{-1}(A_j^T Q^{-1} C_{D,k}\hat{\nabla}_{D,k} + A_j^T Q^{-1}(G^{-1} - I)y_{\text{all}} - A_j^T(Q^{-1} - Q_j^{-1})\hat{e}_j) \quad \text{(A.46)}$$

Then, we have:

$$\begin{aligned} \hat{e}_j - G^{-1}\hat{e}_k &= (1-G^{-1})y_{\text{all}} + A_j(\hat{X}_k - \hat{X}_j) + C_{D,k}\hat{\nabla}_{D,k} \quad \text{(A.47)} \\ &= P_{A_j,Q}^{\perp}(1-G^{-1})y_{\text{all}} + P_{A_j,Q}^{\perp} C_{D,k}\hat{\nabla}_{D,k} + P_{A_j,Q} Q(Q^{-1} - Q_j^{-1})\hat{e}_j \end{aligned}$$

where $P_{A_j,Q}^{\perp} = I - A_j(A_j^T Q^{-1} A_j)^{-1} A_j^T Q^{-1} = I - P_{A_j,Q}$.

The expression for $T_{k,j}$ then becomes:

$$\begin{aligned} & T_{k,j} \\ =\ & \|P_{A_j,Q}^{\perp}(1-G^{-1})y_{\text{all}} + P_{A_j,Q}^{\perp} C_{D,k}\hat{\nabla}_{D,k} + P_{A_j,Q} Q(Q^{-1} - Q_j^{-1})\hat{e}_j\|_{Q^{-1}}^2 \\ -\ & 2\left(P_{A_j,Q}^{\perp}(1-G^{-1})y_{\text{all}} + P_{A_j,Q}^{\perp} C_{D,k}\hat{\nabla}_{D,k} + P_{A_j,Q} Q(Q^{-1} - Q_j^{-1})\hat{e}_j\right)^T Q^{-1}G^{-1}\hat{e}_k \\ +\ & \hat{e}_j^T(Q_j^{-1} - Q^{-1})\hat{e}_j \end{aligned} \quad \text{(A.48)}$$

It can be observed from (A.48) that the difficulties of deriving distribution of $T_{k,j}$ are caused by the following two issues:

1. $A_j \neq A_k$, or equivalently $G \neq I$,

2. $Q \neq Q_j$.

To bypass the first barrier, firstly we introduce a new unknown: $P_{D,k} = F_k(X) - F_j(X) = (n_{\mathsf{NLoS}} - n_{\mathsf{LoS}})\frac{A_i(x_u)}{d_0}$, with i-th link the extra NLoS link under H_k, as compared to H_j. Then the estimates under H_k should be obtained as:

$$\begin{aligned}[\hat{X}_k, \hat{\nabla}_{D,k}, \hat{P}_{D,k}] &= \arg \min_{X,\nabla_{D,k},P_{D,k}} \|y_{\mathsf{all}} - F_j(X) - C_{D,k}\nabla_{D,k} - l_{D,k}P_{D,k}\|^2_{Q_k^{-1}} \\ &\quad \text{subject to } P_{D,k} = F_k(X) - F_j(X),\end{aligned} \tag{A.49}$$

where $l_{D,k}$ is a $2m \times 1$ vector with all zero entries except the $m + i$-th entry equal to 1. Secondly, denoting $L_{D,k} = [C_{D,k}, l_{D,k}]$, $\Delta_{D,k} = [\nabla_{D,k}, P_{D,k}]^T$ and ignoring the relation $P_{D,k} = (n_{\mathsf{NLoS}} - n_{\mathsf{LoS}}) \log_{10} \frac{A_i(x_u)}{d_0}$, hence having $P_{D,k}$ as an additional independent unknown, the above solution can be approximated as:

$$\begin{bmatrix} A_j^T Q_k^{-1} A_j & A_j^T Q_k^{-1} L_{D,k} \\ L_{D,k}^T Q_k^{-1} A_j & L_{D,k}^T Q_k^{-1} L_{D,k} \end{bmatrix} \begin{bmatrix} \hat{X}_k - X \\ \hat{\Delta}_{D,k} \end{bmatrix} \approx \begin{bmatrix} A_j^T Q_k^{-1}[y_{\mathsf{all}} - F_j(X)] \\ L_{D,k}^T Q_k^{-1}[y_{\mathsf{all}} - F_j(X)] \end{bmatrix} \tag{A.50}$$

Following similar derivations from (A.44)-(A.48), the approximated expression for $T_{k,j}$ can be obtained by replacing G with I, $C_{D,k}$ with $L_{D,k}$, $\hat{\nabla}_{D,k}$ with $\hat{\Delta}_{D,k}$, and it holds:

$$\begin{aligned} T_{k,j} &= \|P^{\perp}_{A_j,Q_k} L_{D,k}\hat{\Delta}_{D,k} + P_{A_j,Q_k} Q(Q_k^{-1} - Q_j^{-1})\hat{e}_j\|^2_{Q_k^{-1}} \\ &+ \hat{e}_j^T (Q_j^{-1} - Q_k^{-1})\hat{e}_j \end{aligned} \tag{A.51}$$

where we also make use of the fact that:

$$(\hat{e}_j - \hat{e}_k)^T Q_k^{-1} \hat{e}_k = (\hat{y}_{\mathsf{all},j} - \hat{y}_{\mathsf{all},k})^T Q_k^{-1} \hat{e}_k = 0, \tag{A.52}$$

since $\hat{y}_{\mathsf{all},k}$ now lies in the space $\mathcal{R}([A_j, L_{D,k}])$, $\hat{y}_{\mathsf{all},j}$ lies in $\mathcal{R}([A_j]) \in \mathcal{R}([A_j, L_{D,k}])$ and therefore, the difference $\hat{y}_{\mathsf{all},j} - \hat{y}_{\mathsf{all},k}$ also lies in $\mathcal{R}([A_j, L_{D,k}])$, which is Q_k^{-1}-orthogonal to $\hat{e}_k$.

Now with (A.51), the nuisance terms are the ones with $Q_k^{-1} - Q_j^{-1}$, which is non-zero in general. In the following, an upper- and a lower- bounds on $P\left(T_{k,j} > k_{k,j}\right)$, for $j \in \{j \mid \|S_j - S_k\| = 1\}$, will be derived, where the term $f(Q_k^{-1} - Q_j^{-1})$ becomes zero.

Upper Bound of γ_k with $Q_j = Q_k = Q_0$ For the upper bound, we replace both Q_j and Q_k in (A.51) with Q_0, and we have:

$$T^{\mathsf{UB}}_{k,j} = \hat{\Delta}^T_{D,k} L^T_{D,k} Q_0^{-1} P^{\perp}_{A_j,Q_0} L_{D,k} \hat{\Delta}_{D,k} \tag{A.53}$$

$$k^{\mathsf{UB}}_{k,j} = \ln\left(\frac{\det(Q_0)}{\det(Q_0)}\right) = 0 \tag{A.54}$$

where we made use of $(P^{\perp}_{A_j,Q_0})^T Q_0^{-1} = Q_0^{-1} P^{\perp}_{A_j,Q_0}$ and $P^{\perp}_{A_j,Q_0} P^{\perp}_{A_j,Q_0} = P^{\perp}_{A_j,Q_0}$.

The motivation of setting $Q_j = Q_k = Q_0$ is that, suppose the mean values of NLoS ToA range measurements and RSS measurements are kept unchanged, but the variances are reduced to as in LoS cases (Q_0), then it would be easier to identify LoS and NLoS measurements since the two PDFs are more separated.

The distribution of $T_{k,j}^{\mathsf{UB}}$ can be derived based on (A.53) as follows. Applying a Gaussian elimination on (A.50) by multiplying the following matrix on both sides:

$$\begin{bmatrix} I_a & 0 \\ -L_{D,k}Q_0^{-1}(A_j^T Q_0^{-1} A_j)^{-1} & I_b \end{bmatrix} \tag{A.55}$$

we obtain:

$$\hat{\Delta}_{D,k}^{\mathsf{UB}} = (L_{D,k}^T Q_0^{-1} P_{A_j,Q_0}^{\perp} L_{D,k})^{-1} L_{D,k}^T Q_0^{-1} P_{A_j,Q_0}^{\perp} [y_{\mathsf{all}} - F_j(X)]. \tag{A.56}$$

Please note that the estimates with the super script UB are assumed to be obtained when the variance matrix of the models in all hypotheses are equal to Q_0. In practice, the final estimates obtained using Classification Type I will be used instead.

Based on the above equation, the variance of $\hat{\Delta}_{D,k}^{\mathsf{UB}}$ can be derived based on the variance propagation law [89]:

$$Q_{\hat{\Delta}_{D,k}\hat{\Delta}_{D,k}}^{\mathsf{UB}} = (L_{D,k}^T Q_0^{-1} P_{A_j,Q_0}^{\perp} L_{D,k})^{-1}. \tag{A.57}$$

Hence, (A.53) can be further written as:

$$T_{k,j}^{\mathsf{UB}} \;=\; [\hat{\Delta}_{D,k}^{\mathsf{UB}}]^T [Q_{\hat{\Delta}_{D,k}\hat{\Delta}_{D,k}}^{\mathsf{UB}}]^{-1} \hat{\Delta}_{D,k}^{\mathsf{UB}}. \tag{A.58}$$

The mean of $\hat{\Delta}_{D,k}^{\mathsf{UB}}$ is different under H_j and H_k. Under H_j, one has $\mathsf{E}\{\hat{\Delta}_{D,k}^{\mathsf{UB}}\} = 0$ and under H_k, one has $\mathsf{E}\{\hat{\Delta}_{D,k}^{\mathsf{UB}}\} = \Delta_{D,k} = [\nabla_{D,k}^{\mathsf{UB}}, \mathsf{P}_{D,k}^{\mathsf{UB}}]^T$. Please note that we have again ignored the second- and higher- order terms in the Taylor expansion of $\hat{\Delta}_{D,k}^{\mathsf{UB}}$ at $[\nabla_{D,k}^{\mathsf{UB}}, \mathsf{P}_{D,k}^{\mathsf{UB}}]^T$. The variance matrices of $\hat{\Delta}_{D,k}^{\mathsf{UB}}$ under both hypotheses are equal to $Q_{\hat{\Delta}_{D,k}\hat{\Delta}_{D,k}}^{\mathsf{UB}}$.

Thus, given the fact that y_{all} follows Normal distributions under both H_j and H_k, the distribution of $\hat{\Delta}_{D,k}^{\mathsf{UB}}$ can be summarized as:

$$H_j : \hat{\Delta}_{D,k}^{\mathsf{UB}} \sim \mathcal{N}(0, Q_{\hat{\Delta}_{D,k}\hat{\Delta}_{D,k}}^{\mathsf{UB}}) \text{ and } H_k : \hat{\Delta}_{D,k}^{\mathsf{UB}} \sim \mathcal{N}(\Delta_{D,k}^{\mathsf{UB}}, Q_{\hat{\Delta}_{D,k}\hat{\Delta}_{D,k}}^{\mathsf{UB}}) \tag{A.59}$$

so that the quadratic form $T_{k,j}^{\mathsf{UB}} = [\hat{\Delta}_{D,k}^{\mathsf{UB}}]^T [Q_{\hat{\Delta}_{D,k}\hat{\Delta}_{D,k}}^{\mathsf{UB}}]^{-1} \hat{\Delta}_{D,k}^{\mathsf{UB}}$ is Chi-square distributed:

$$H_j : \underline{T}_{k,j}^{\mathsf{UB}} \sim \chi^2(q_D, 0) \text{ and } H_k : \underline{T}_{k,j}^{\mathsf{UB}} \sim \chi^2(q_{D,k}, \lambda^{\mathsf{UB}}) \tag{A.60}$$

where $\lambda^{\mathsf{UB}} = [\Delta_{D,k}^{\mathsf{UB}}]^T [Q_{\hat{\Delta}_{D,k}\hat{\Delta}_{D,k}}^{\mathsf{UB}}]^{-1} \Delta_{D,k}^{\mathsf{UB}}$ and $q_{D,k}$ denotes the length of the vector $\Delta_{D,k}$, which is equal to 2.

With $\Phi_j \cap \Phi_k = \Phi_j$, the probability of $T_{k,j}^{\mathsf{UB}} > k_{k,j}^{\mathsf{UB}}$, when H_k is true, can be calculated as:

$$P^{\mathsf{UB}}\left(T_{k,j}^{\mathsf{UB}} > k_{k,j}^{\mathsf{UB}}\right), \text{ with } \underline{T}_{k,j}^{\mathsf{UB}} \sim \chi^2(q_{D,k}, \lambda^{\mathsf{UB}}). \tag{A.61}$$

With $\Phi_j \cap \Phi_k = \Phi_k$, the probability of $T_{k,j}^{\mathsf{UB}} > k_{k,j}^{\mathsf{UB}}$, when H_k is true, actually equals to the probability of $T_{k,j}^{\mathsf{UB}} < k_{k,j}^{\mathsf{UB}}$, when H_j true and $\Phi_j \cap \Phi_k = \Phi_j$:

$$P^{\mathsf{UB}}\left(T_{k,j}^{\mathsf{UB}} < k_{k,j}^{\mathsf{UB}}\right), \text{ with } \underline{T}_{k,j}^{\mathsf{UB}} \sim \chi^2(q_{D,k}, 0) \tag{A.62}$$

Please note that in both (A.61) and (A.62) it holds that $j \in \{j \mid \|S_j - S_k\| = 1\}$.

Thus, we can calculate the probability of correct decision when H_k is true as:

$$\begin{aligned}\gamma_k^{\mathsf{UB}} = \prod_{j\in\{j\,|\,\|S_j-S_k\|=1\}} & \left[a\, P^{\mathsf{UB}}\left(T_{k,j}^{\mathsf{UB}} > k_{k,j}^{\mathsf{UB}} | \underline{T}_{k,j}^{\mathsf{UB}} \sim \chi^2(q_{D,k}, \lambda^{\mathsf{UB}})\right)\right. \\ + & \left.(1-a)\, P^{\mathsf{UB}}\left(T_{k,j}^{\mathsf{UB}} < k_{k,j}^{\mathsf{UB}} | \underline{T}_{k,j}^{\mathsf{UB}} \sim \chi^2(q_{D,k}, 0)\right)\right]\end{aligned} \tag{A.63}$$

where $k_{k,j}^{\mathsf{UB}} = 0$, see (A.54), and a is a controlling factor:

$$a = \begin{cases} 1, & \Phi_j \cap \Phi_k = \Phi_j \\ 0, & \Phi_j \cap \Phi_k = \Phi_k \end{cases} \tag{A.64}$$

Lower Bound of γ_k with $Q_j = Q_k = Q_{2^m-1}$ Here, the lower bound is derived by replacing both Q_j and Q_k in (A.51) with Q_{2^m-1}. On the opposite side of the reasoning for the upper bound, the motivation here is that, suppose the mean values of LoS ToA range measurements and RSS measurements are kept unchanged, but the variances are increased to as in NLoS cases (Q_{2^m-1}), then it would be more difficult to identify LoS and NLoS measurements since the two PDFs are less separated.

The expression of $T_{k,j}^{\mathsf{LB}}$ can be given as:

$$T_{k,j}^{\mathsf{LB}} = \hat{\Delta}_{D,k}^T L_{D,k}^T Q_{2^m-1}^{-1} P_{A_j,2^m-1}^{\perp} L_{D,k} \hat{\Delta}_{D,k} \tag{A.65}$$

$$k_{k,j}^{\mathsf{LB}} = \ln\left(\frac{\det(Q_{2^m-1})}{\det(Q_{2^m-1})}\right) = 0, \tag{A.66}$$

and

$$H_j : \underline{T}_{k,j}^{\mathsf{LB}} \sim \chi^2(q_{D,k}, 0) \text{ and } H_k : \underline{T}_{k,j}^{\mathsf{LB}} \sim \chi^2(q_{D,k}, \lambda^{\mathsf{LB}}) \tag{A.67}$$

where $\lambda^{\mathsf{LB}} = [\Delta_{D,k}^{\mathsf{LB}}]^T [Q_{\hat{\Delta}_{D,k}\hat{\Delta}_{D,k}}^{\mathsf{LB}}]^{-1} \Delta_{D,k}^{\mathsf{LB}}$ and $Q_{\hat{\Delta}_{D,k}\hat{\Delta}_{D,k}}^{\mathsf{LB}} = (L_{D,k}^T Q_{2^m-1}^{-1} P_{A,2^m-1}^{\perp} L_{D,k})^{-1}$. The estimates with the super script LB are assumed to be obtained by setting the variance matrix of the corresponding model to Q_{2^m-1}. In practice, the final estimates obtained using Classification Type I will be used instead.

The corresponding lower bound for γ_k is calculated as:

$$\begin{aligned}\gamma_k^{\mathsf{LB}} = \prod_{j\in\{j\,|\,\|S_j-S_k\|=1\}} & \left[a\, P^{\mathsf{LB}}\left(T_{k,j}^{\mathsf{LB}} > k_{k,j}^{\mathsf{LB}} | \underline{T}_{k,j}^{\mathsf{LB}} \sim \chi^2(q_{D,k}, \lambda^{\mathsf{LB}})\right)\right. \\ + & \left.(1-a)\, P^{\mathsf{LB}}\left(T_{k,j}^{\mathsf{LB}} < k_{k,j}^{\mathsf{LB}} | \underline{T}_{k,j}^{\mathsf{LB}} \sim \chi^2(q_{D,k}, 0)\right)\right]\end{aligned} \tag{A.68}$$

where $k_{k,j}^{\mathsf{LB}} = 0$, see (A.66), and a is a controlling factor defined in (A.64).

To this end, the derivations for approximated γ_k, an upper- and a lower- bounds have been presented, based on which, the probability of incorrect decision when H_k is true, can be given as $\beta_k = 1 - \gamma_k$, and the two bounds as $\beta_k^{\mathsf{UB}} = 1 - \gamma_k^{\mathsf{LB}}$ and $\beta_k^{\mathsf{LB}} = 1 - \gamma_k^{\mathsf{UB}}$.

A.4.2 Upper and Lower Bounds of $T_{k,j}$ in Classification Type II

Recall from (6.88), with $\hat{e}_j \approx P^{\perp}_{A_j,Q_j}[y_{\mathsf{all}} - C_j\nabla_j - F_j(x_u)]$, $\hat{e}_k \approx P^{\perp}_{A_k,Q_k}[y_{\mathsf{all}} - C_k\nabla_k - F_k(x_u)]$, the test variable $T_{k,j}$ can be further written as:

$$\begin{aligned} & T_{k,j} \\ = \; & \|P^{\perp}_{A_j,Q_j}[y_{\mathsf{all}} - C_j\nabla_j - F_j(x_u)] - P^{\perp}_{A_k,Q_k}[y_{\mathsf{all}} - C_k\nabla_k - F_k(x_u)]\|^2_{Q_k^{-1}} \\ - \; & 2(P^{\perp}_{A_j,Q_j}[y_{\mathsf{all}} - C_j\nabla_j - F_j(x_u)] - P^{\perp}_{A_k,Q_k}[y_{\mathsf{all}} - C_k\nabla_k - F_k(x_u)])^T Q_k^{-1}\hat{e}_k \\ + \; & \hat{e}_j^T(Q_j^{-1} - Q_k^{-1})\hat{e}_j, \end{aligned} \tag{A.69}$$

where $A_k = \partial_{x_u^T} F_k(x_u) = GA_j$, and G can be found in (A.41).

Similar as in case of Classification Type I, the derivation for a simple expression for the distribution of $\underline{T}_{k,j}$ is hindered by the following two factors: 1) $G \neq I$ and 2) $Q_j \neq Q_k$.

Again, we firstly treat $P_{D,k} = F_k(x_u) - F_j(x_u) = (n_{\mathsf{NLoS}} - n_{\mathsf{LoS}})\frac{A_i(x_u)}{d_0}$ as an additional unknown, with i-th link the extra NLoS link under H_k, as compared to H_j. And the estimates under H_k should be obtained as:

$$\begin{aligned} [\hat{x}_{u,k}, \hat{P}_{D,k}]^T \; = \; & \arg\min_{x_u, P_{D,k}} \|y_{\mathsf{all}} - F_j(x_u) - C_k\nabla_k - L_{D,k}P_{D,k}\|^2_{Q_k^{-1}}, \\ & \text{subject to } P_{D,k} = F_k(x_u) - F_j(x_u) \end{aligned} \tag{A.70}$$

where $L_{D,k}$ is a $2m \times 1$ vector with all zero entries except the $m+i$-th entry equal to 1.

Secondly, ignoring the relation $P_{D,k} = (n_{\mathsf{NLoS}} - n_{\mathsf{LoS}})\log_{10}\frac{A_i(x_u)}{d_0}$, the above solution can be approximated as:

$$\begin{aligned} & \begin{bmatrix} A_j^T Q_k^{-1} A_j & A_j^T Q_k^{-1} L_{D,k} \\ L_{D,k}^T Q_k^{-1} A_j & L_{D,k}^T Q_k^{-1} L_{D,k} \end{bmatrix} \begin{bmatrix} \hat{x}_{u,k} - x_u \\ \hat{P}_{D,k} \end{bmatrix} \\ \approx \; & \begin{bmatrix} A_j^T Q_k^{-1}[y_{\mathsf{all}} - F_j(x_u) - C_k\nabla_k] \\ L_{D,k}^T Q_k^{-1}[y_{\mathsf{all}} - F_j(x_u) - C_k\nabla_k] \end{bmatrix}. \end{aligned} \tag{A.71}$$

$$\Rightarrow \hat{x}_{u,k} \approx P_{A_j,Q_k}[y_{\mathsf{all}} - F_j(x_u) - C_k\nabla_k - A_j^T Q_k^{-1} L_{D,k}\hat{P}_{D,k}] \tag{A.72}$$

$$\hat{P}_{D,k} \approx (L_{D,k}^T Q_k^{-1} P^{\perp}_{A_k,Q_k} L_{D,k})^{-1} L_{D,k}^T Q_k^{-1} P^{\perp}_{A_j,Q_k}[y_{\mathsf{all}} - F_j(x_u) - C_k\nabla_k] \tag{A.73}$$

and we have, accordingly,

$$\hat{e}_k \approx P^{\perp}_{A_j,Q_k}[y_{\mathsf{all}} - F_j(x_u) - C_k\nabla_k - L_{D,k}\hat{P}_{D,k}]. \tag{A.74}$$

Based on (A.69) and (A.74), the idea of an upper- and a lower- bounds can also be applied for Classification Type II, which are introduced next.

Upper Bound with $Q_j = Q_k = Q_0$ Similar as for Classification Type I, the upper bound for Type II is derived by replacing the variance matrices in all hypotheses with Q_0.

Now, plugging (A.74) into (A.69), $T^{\mathsf{UB}}_{k,j}$ can be written as:

$$T^{\mathsf{UB}}_{k,j} = (C_k\nabla_k - C_j\nabla_j + L_{D,k}\hat{P}^{\mathsf{UB}}_{D,k})^T Q_0^{-1} P^{\perp}_{A_j,Q_0}(C_k\nabla_k - C_j\nabla_j + L_{D,k}\hat{P}^{\mathsf{UB}}_{D,k}) \tag{A.75}$$

where we also make use of the fact that:

$$(\hat{e}_j - \hat{e}_k)^T Q_k^{-1} \hat{e}_k = (\hat{y}_{\mathsf{all},j} - \hat{y}_{\mathsf{all},k})^T Q_k^{-1} \hat{e}_k = 0, \tag{A.76}$$

see (A.52).

Assuming i-th link the extra NLoS link under H_k, as compared to H_j, we denote $C_k \nabla_k - C_j \nabla_j = R_{D,k} \mu_i$, with $R_{D,k}$ the i-th column of $C_{D,k}$ and μ_i the NLoS bias in the i-th ToA range measurement, then (A.75) can be rewritten as:

$$\begin{aligned} T_{k,j}^{\mathsf{UB}} &= u(\hat{P}_{D,k}^{\mathsf{UB}})^2 + 2v\mu_i \hat{P}_{D,k}^{\mathsf{UB}} + w\mu_i^2 \\ &= u(\hat{P}_{D,k}^{\mathsf{UB}} + \frac{v}{u}\mu_i)^2 + (w - \frac{v^2}{u})\mu_i^2 \end{aligned} \tag{A.77}$$

where

$$\begin{aligned} u &= L_{D,k}^T Q_0^{-1} P_{A_k,Q_0}^{\perp} L_{D,k} \\ v &= R_{D,k}^T Q_0^{-1} P_{A_k,Q_0}^{\perp} L_{D,k} \\ w &= R_{D,k}^T Q_0^{-1} P_{A_k,Q_0}^{\perp} R_{D,k} \end{aligned}$$

The decision in (6.84) can be rewritten as:

$$\text{accept } H_k \text{ if} \quad \underline{T}_{k,j}^{\mathsf{UB}} = u(\hat{P}_{D,k}^{\mathsf{UB}} + \frac{v}{u}\mu_i)^2 > k_{k,j}^{\mathsf{UB}} = (\frac{v^2}{u} - w)\mu_i^2 \tag{A.78}$$

Based on (A.73), the distribution of $\hat{P}_{D,k}^{\mathsf{UB}}$ can be approximated as:

$$\underline{\hat{P}}_{D,k}^{\mathsf{UB}} \sim \begin{cases} \mathcal{N}(0, (L_{D,k}^T Q_0^{-1} P_{A_k,Q_0}^{\perp} L_{D,k})^{-1}), & \text{under } H_j \\ \mathcal{N}(P_{D,k}^{\mathsf{UB}}, (L_{D,k}^T Q_0^{-1} P_{A_k,Q_0}^{\perp} L_{D,k})^{-1}), & \text{under } H_k \end{cases} \tag{A.79}$$

with $P_{D,k}^{\mathsf{UB}} = F_k(x_u) - F_j(x_u)$.

The distributions of $\underline{T}_{k,j}^{\mathsf{UB}}$ under H_j and H_k can then be given as:

$$H_j : \underline{T}_{k,j}^{\mathsf{UB}} \sim \chi^2(q_{D,k}, 0) \text{ and } H_k : \underline{T}_{k,j}^{\mathsf{UB}} \sim \chi^2(q_{D,k}, \lambda^{\mathsf{UB}}) \tag{A.80}$$

where $q_{D,k}$ denotes the length of the vector $P_{D,k}^{\mathsf{UB}}$, which is equal to 1, and $\lambda^{\mathsf{UB}} = u(P_{D,k}^{\mathsf{UB}} + \frac{v}{u}\mu_i)^2$.

With $\Phi_j \cap \Phi_k = \Phi_j$, the probability of $T_{k,j}^{\mathsf{UB}} > k_{k,j}^{\mathsf{UB}}$, when H_k is true, can be calculated as:

$$P^{\mathsf{UB}}\left(T_{k,j}^{\mathsf{UB}} > k_{k,j}^{\mathsf{UB}}\right), \text{ with } \underline{T}_{k,j}^{\mathsf{UB}} \sim \chi^2(q_{D,k}, \lambda^{\mathsf{UB}}). \tag{A.81}$$

With $\Phi_j \cap \Phi_k = \Phi_k$, the probability of $T_{k,j}^{\mathsf{UB}} > k_{k,j}^{\mathsf{UB}}$, when H_k is true, actually equals to the probability of $T_{k,j}^{\mathsf{UB}} < k_{k,j}^{\mathsf{UB}}$, when H_j true and $\Phi_j \cap \Phi_k = \Phi_j$:

$$P^{\mathsf{UB}}\left(T_{k,j}^{\mathsf{UB}} < k_{k,j}^{\mathsf{UB}}\right), \text{ with } \underline{T}_{k,j}^{\mathsf{UB}} \sim \chi^2(q_{D,k}, 0) \tag{A.82}$$

Please note that in both (A.81) and (A.82) it holds that $j \in \{j \mid \|S_j - S_k\| = 1\}$.

Thus, we can calculate the probability of correct decision when H_k is true as:

$$\begin{aligned} \gamma_k^{\mathsf{UB}} = \prod_{j \in \{j \mid \|S_j - S_k\| = 1\}} & \left[a\, P^{\mathsf{UB}}\left(T_{k,j}^{\mathsf{UB}} > k_{k,j}^{\mathsf{UB}} | \underline{T}_{k,j}^{\mathsf{UB}} \sim \chi^2(q_{D,k}, \lambda^{\mathsf{UB}})\right)\right. \\ & + \left. (1-a)\, P^{\mathsf{UB}}\left(T_{k,j}^{\mathsf{UB}} < k_{k,j}^{\mathsf{UB}} | \underline{T}_{k,j}^{\mathsf{UB}} \sim \chi^2(q_{D,k}, 0)\right)\right] \end{aligned} \tag{A.83}$$

where $k_{k,j}^{\mathsf{UB}} = (\frac{v^2}{u} - w)\mu_i^2$, see (A.78), and a is a controlling factor in (A.64).

Lower Bound of γ_k with $Q_j = Q_k = Q_{2^m-1}$ The derivation for the lower bound can be done in a similar way as for the upper bound. The only difference here is that the variance matrices in all hypotheses are replaced with Q_{2^m-1}.

The decision in (6.84) can be rewritten as:

In accordance with (A.78)-(A.83), we have:

$$\text{accept } H_k \text{ if} \quad \underline{T}_{k,j}^{\mathsf{LB}} = u(\hat{P}_{D,k}^{\mathsf{LB}} + \frac{v}{u}\mu_i)^2 > k_{k,j}^{\mathsf{LB}} = (\frac{v^2}{u} - w)\mu_i^2 \tag{A.84}$$

where

$$\begin{aligned} u &= L_{D,k}^T Q_{2^m-1}^{-1} P_{A_k,Q_{2^m-1}}^{\perp} L_{D,k} \\ v &= R_{D,k}^T Q_{2^m-1}^{-1} P_{A_k,Q_{2^m-1}}^{\perp} L_{D,k} \\ w &= R_{D,k}^T Q_{2^m-1}^{-1} P_{A_k,Q_{2^m-1}}^{\perp} R_{D,k} \end{aligned}$$

The distributions of $\underline{T}_{k,j}^{\mathsf{LB}}$ under H_j and H_k can then be given as:

$$H_j : \underline{T}_{k,j}^{\mathsf{LB}} \sim \chi^2(q_{D,k}, \lambda_j^{\mathsf{LB}}) \text{ and } H_k : \underline{T}_{k,j}^{\mathsf{LB}} \sim \chi^2(q_{D,k}, \lambda_k^{\mathsf{LB}}) \tag{A.85}$$

where $q_{D,k} = 1$, and $\lambda^{\mathsf{LB}} = u(P_{D,k}^{\mathsf{LB}} + \frac{v}{u}\mu_i)^2$.

Thus, we can calculate the probability of correct decision when H_k is true as:

$$\begin{aligned} \gamma_k^{\mathsf{LB}} = \prod_{j \in \{j \,|\, \|S_j - S_k\| = 1\}} & \left[a\, P^{\mathsf{LB}} \left(T_{k,j}^{\mathsf{LB}} > k_{k,j^{\mathsf{LB}}} | \underline{T}_{k,j}^{\mathsf{LB}} \sim \chi^2(q_{D,k}, \lambda^{\mathsf{LB}}) \right) \right. \\ + & \left. (1-a)\, P^{\mathsf{LB}} \left(T_{k,j}^{\mathsf{LB}} < k_{k,j}^{\mathsf{LB}} | \underline{T}_{k,j}^{\mathsf{LB}} \sim \chi^2(q_{D,k}, 0) \right) \right] \end{aligned} \tag{A.86}$$

where $k_{k,j}^{\mathsf{LB}} = (\frac{v^2}{u} - w)\mu_i^2$, and a is a controlling factor in (A.64).

A.4.3 Flop Count for NLoS Schemes

A.4.3.1 Classification Type I

The flop count for the Classification Type I can be done in a similar way as in Appendix A.3.3.

Considering the clock offset free case using the Gauss-Newton method, assuming that there are q NLoS biases to be estimated, then in each iteration, we need to update d^0 (A.22) and $F(x_u^0, \nabla^0)$. To update a single element d_i^0, $3mn$ flops are required. And the term $F(x^0, \nabla^0)$ can be updated based on d^0 and ∇^0, which requires q flops.

In each iteration, the calculations for the Gauss-Newton method include updating part of the $2m \times (n+q)$ matrix A and the equation

$$(A^T W A)(\begin{bmatrix} \hat{x}_u \\ \hat{\nabla} \end{bmatrix} - \begin{bmatrix} \hat{x}_u^0 \\ \hat{\nabla}^0 \end{bmatrix}) = A^T W (y - F(x_u^0, \nabla^0)). \tag{A.87}$$

The expression for a single element A_{ij} is

$$A_{ij} = \begin{cases} \frac{x_{uj}-x_{ij}}{d_i^0}, & i = 1, ..., m \text{ and } j = 1, ..., n \\ \frac{1}{d_i^0} \cdot \frac{x_{uj}-x_{ij}}{d_i^0}, & i = m+1, ..., 2m \text{ and } j = 1, ..., n \end{cases}, \tag{A.88}$$

which requires 2+1 flops. Note that the columns of A with $j > n$ contains only constants, and n is the dimension of x_u. For the whole matrix A, $3mn$ flops are needed.

Furthermore, to calculate $\hat{x}_u, \hat{\nabla}$, one needs, $6mn$ for $A^T \cdot W$ (note the structure of W in (6.49)), $2m$ for $y_{\text{all}} - F(x_u^0, \nabla^0)$, $(n+q)(4m-1)$ for $A^T W \cdot (y_{\text{all}} - F(x_u^0, \nabla^0))$, $n(4m-1)(n+q)$ for $A^T W \cdot A$, $(n+q)^3/3$ for Cholesky decomposition, $2(n+q)^2$ for $\begin{bmatrix} \hat{x}_u \\ \hat{\nabla} \end{bmatrix} - \begin{bmatrix} \hat{x}_u^0 \\ \hat{\nabla}^0 \end{bmatrix}$ and $n+q$ for $\begin{bmatrix} \hat{x}_u \\ \hat{\nabla} \end{bmatrix}$. The stopping criteria requires the evaluation of $\| \begin{bmatrix} \hat{x}_u \\ \hat{\nabla} \end{bmatrix} - \begin{bmatrix} \hat{x}_u^0 \\ \hat{\nabla}^0 \end{bmatrix} \|^2$ and since $\begin{bmatrix} \hat{x}_u \\ \hat{\nabla} \end{bmatrix} - \begin{bmatrix} \hat{x}_u^0 \\ \hat{\nabla}^0 \end{bmatrix}$ has already been calculated in previous steps, the number of required flops is $2(n+q)+1$.

The total number of flops read:

$$\begin{aligned} \text{flops}_{CTI} &= q^3/3 + (n+2)q^2 + (n^2 + 4mn + 3n + 4m)q \\ &+ n^3/3 + (4m+1)n^2 + 13mn + 4n + 1. \end{aligned} \tag{A.89}$$

Bibliography

[1] Pratap Misra and Per Enge. *Global Positioning System: Signals, Measurements, and Performances*. Ganga-Jamuna Press, second edition, 2006.

[2] S. Gezici, Zhi 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. *Signal Processing Magazine, IEEE*, 22(4):70–84, July 2005.

[3] David S. Chiu and Kyle P. O'Keefe. DGPS+UWB: Seamless outdoor-to-indoor positioning. *GPS World*, 20(3):32–38, March 2009.

[4] K Krumvieda, C. Cloman, E. Olson, J. Thomas, W. Kober, P. Madhani, and P. Axelrad. A complete IF software GPS receiver. In *Proceedings of The Institute of Navigation's ION GPS-2001*, pages 789–829, Salt Lake City, Sep 2001.

[5] E.D. Kaplan and C.J. Hegarty. *Understanding GPS: Principles and Applications*. Artech House, Boston MA, 2006.

[6] D. Odijk and C. C. J. M. Tiberius. Assessing the accuracy of high sensitivity GPS receiver for location based services. In *Proc. NaviTec '06*, Noordwijk, The Netherlands, December 5–9, 2006.

[7] Maria-Gabriella Di Benedetto and Guerino Giancola. *Understanding Ultra Wide Band Radio Fundamentals*. Professional Technical Reference, Prentice Hall, 1995.

[8] I. Oppermann, M. Hämäläinen, and J. Iinatti. *UWB: Theory and Applications*. UK: Wiley, 2004.

[9] J.H. Reed. *An Introduction to Ultra Wideband Communication Systems*. Prentice Hall, Upper Saddl Ridge River NJ, 2005.

[10] Liuqing Yang and G. B. Giannakis. Ultra-wideband communications: an idea whose time has come. *IEEE Signal Process. Mag.*, 21(6):26–54, November 2004.

[11] Federal Communications Commission. Revision of part 15 of the commission's rules regarding ultra-wideband transmission systems, 2002, available at: hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-02-48A1.pdf.

[12] Electronic Communications Committee. Ecc decision of 24 March 2006 amended 6 July 2007 at Constanta on the harmonised conditions for devices using ultra-wideband (UWB) technology in bands below 10.6 GHz. Jul 2007, available at:

http://www.ero.dk/documentation/docs/doc98/official/Word/ECCDEC0604.DOC ?frames=0.

[13] R. Zetik, R. Thomi, and I. Sachs. Ultra-wideband real-time channel sounder design and application. In *COST273 Temporary Document TD(03) 201*, Prague, Czech Republic, Sep 2003.

[14] R. Zctik, J. Snchs, and P. Peyerl. UWB radar: Distance and positioning measurements. In *Proceedings of The International Conference on Electromagnetics in Advanced Applications ICEAA03*, Torino, Italy, Sep 2003.

[15] G.R. Opshaug and P. Enge. GPS and UWB for indoor navigation. In *GPS Conference*, Salt Lake City, Sep 2001.

[16] N. Patwari, J.N. Ash, S. Kyperountas, III Hero, A.O., R.L. Moses, and N.S. Correal. Locating the nodes: cooperative localization in wireless sensor networks. *Signal Processing Magazine, IEEE*, 22(4):54–69, July 2005.

[17] Y. Zhao. Standardization of mobile phone positioning for 3G systems. *Communication Magazine, IEEE*, 40:108–116, July 2002.

[18] S. Kim, T. Pals, R.A. Iltis, and H. Lee. CDMA multipath channel estimation using generalized successive interference cancellation algorithm for radiolocation. In *Proc. 36th Conf. Information Sciences and Systems*, Mar 2002.

[19] D.D. McCrady, L. Doyle, H. Forstrom, T. Dempsy, and M. Martorana. Mobile ranging with low accuracy clocks. *IEEE Trans. Microwave Theory Tech.*, 48(6):951957, Jun 2000.

[20] R. Fleming and C. Kushner. Low-power, miniature, distributed position location and communication devices using ultra-wideband, nonsinusoidal communication technology. In *Aetherwire Inc., ARPA Contract J-FBI-94-058, Tech. Rep.*, Jul 1995.

[21] J. Y. Lee and R. A. Scholtz. Ranging in a dense multipath environment using an UWB radio link. *IEEE J. Sel. Areas Commun.*, 20(9):1677–1683, December 2002.

[22] B. Ludden and L. Lopes. Cellular based location technologies for UMTS: a comparison between IPDL and TA-IPDL. In *IEEE Spring Vehicular Technology Conf.*, 2000.

[23] J. Winter and C. Wengerter. High resolution estimation of the time of arrival for GSM location. In *IEEE Spring Vehicular Technology Conf.*, 2000.

[24] M. McGuire, K.N. Plataniotis, and A.N. Venetsanopoulos. Location of mobile terminals using time measurements and survey points. *Vehicular Technology, IEEE Transactions on*, 52(4):999 – 1011, July 2003.

[25] F. Sivrikaya and B. Yener. Time synchronization in sensor networks: A survey. In *IEEE Network*, volume 18, Jul-Aug 2004.

[26] Cliff Randell and Henk Muller. Low cost indoor positioning system. In *Proc. Ubiquitous Computing*, pages 42–48, Atlanta Georgia, September 2001.

[27] E.G. Larsson. Cramer-rao bound analysis of distributed positioning in sensor networks. In *IEEE Signal Processing Lett.*, volume 11, pages 334–337, Mar 2004.

[28] T. Gigl, G. J. M. Janssen, V. Dizdarevic, K. Witrisal, and Z. Irahhauten. Analysis of a UWB indoor positioning system based on received signal strength. In *Proc. IEEE WPNC 2007*, pages 97–101, Hannover, Germany, March 2007.

[29] Yihong Qi and H. Kobayashi. On relation among time delay and signal strength based geolocation methods. In *Proc. IEEE GLOBECOM 2003*, volume 7, pages 4079–4083, December 2003.

[30] G. Bellusci, G. J. M. Janssen, J. Yan, and C. C. J. M. Tiberius. Novel ultra wideband low complexity ranging using different channel statistics. In *Proc. IEEE Int. Wireless Commun. Networking Conf. (WCNC)*, pages 290–295, Las Vegas, NV, March 31/April 3, 2008.

[31] J.A. Dabin, Nan Ni Haimovich, A.M. Niver, and H. E. Grebel. The effects of antenna directivity on path loss and multipath propagation in UWB indoor wireless channels. In *Proc. IEEE Ultra Wideband Systems and Technologies Conference*, pages 305–309, Nov 2003.

[32] T. Wakabayashi and H. Matsui. Omni-directional characteristics over frequency range for UWB of modified planar antenna with an elliptical element on the dielectric substrate. In *Proc. IEEE International Wireless Communication Systems Symposium, ISWCS*, pages 463–467, Tuscany, Sep 2009.

[33] Mikkel Baun Kjærgaard. Indoor positioning with radio location fingerprinting. *PhD Thesis*, 2010, University of Pittsburgh, School of Information Science.

[34] Kamol Kaemarungsi. Design of indoor positioning systems based on location fingerprinting technique. *PhD Thesis*, 2005, University of Aarhus, Faculty of Science.

[35] Hui Liu, H. Darabi, P. Banerjee, and Jing Liu. Survey of wireless indoor positioning techniques and systems. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(6):1067–1080, Nov. 2007.

[36] G. Bellusci, G. J. M. Janssen, J. Yan, and C. C. J. M. Tiberius. Model of the distance and bandwidth dependency of TOA-based UWB ranging error. In *Proc. IEEE Int. Conf. on Ultra-Wideband (ICUWB)*, pages 193–196, Hannover, Germany, September 10–12, 2008.

[37] Z. Irahhauten, G. Bellusci, G. J. M. Janssen, and C. C. J. M. Tiberius. Investigation of UWB ranging in dense indoor multipath environments. In *Proc. IEEE Int. Conf. Commun. Systems (ICCS)*, pages 1–5, Singapore, October 2006.

[38] G. Bellusci, G.J.M. Janssen, J. Yan, and C.C.J.M. Tiberius. A low-complexity uwb receiver concept for TOA based UWB ranging. In *Proc. IEEE ICUWB*, pages 618–623, Vancouver, 2009.

[39] S. M. Kay. *Fundamentals of Statistical Signal Processing: EstimationTheory*. Prentice-Hall International Editions, Englewood Cliffs, NJ, 1993.

[40] Yihong Qi, H. Suda, and H. Kobayashi. On time-of-arrival positioning in a multipath environment. In *Vehicular Technology Conference, 2004. VTC2004-Fall. 2004 IEEE 60th*, volume 5, pages 3540–3544, Sept. 2004.

[41] P. J. G. Teunissen. Nonlinear least-squares. *Manuscripta Geodaetica*, 15(3):137–150, 1990.

[42] R. O. Schmidt. A new approach to geometry of range difference location. *IEEE Trans. Aerosp. Electron. Syst.*, 8:821–835, 1972.

[43] S. Bancroft. An algebraic solution for the GPS equations. *IEEE Trans. Aerosp. Electron. Syst.*, 21(7):56–59, 1985.

[44] H. C. Schau and A. Z. Robinson. Passive source localization employing intersecting spherical surfaces from time-of-arrival differences. *IEEE Trans. Acoust., Speech, Signal Process.*, ASSP-35:1223–1225, August 1987.

[45] J. O. Smith and J. S. Abel. The spherical interpolation method of source localization. *IEEE J. Ocean. Eng.*, 12:246–252, January 1987.

[46] J. O. Smith and J. S. Abel. Closed-form least-squares source location estimation from range-difference measurements. *IEEE Trans. Acoust., Speech, Signal Process.*, ASSP-35:1661–1669, August 1987.

[47] B. Friedlander. A passive location algorithm and its accuracy analysis. *IEEE J. Ocean. Eng.*, pages 234–245, 1987.

[48] L.O. Krause. A direct solution to GPS-type navigation equations. *IEEE Trans. Aerosp. Electron. Syst.*, 23(2):225–232, 1987.

[49] B. Fang. Simple solutions for hyperbolic and related position fixes. *IEEE Trans. Aerosp. Electron. Syst.*, pages 734–753, 1990.

[50] J. Chaffee and J. Abel. On the exact solutions of pseudorange equations. *IEEE Trans. Aerosp. Electron. Syst.*, 30(4):1021–1030, 1991.

[51] K. Ho and Y. Chan. Solution and performance analysis of geolocation by TDOA. *IEEE Trans. Aerosp. Electron. Syst.*, 29:1311–1322, 1993.

[52] Y. Chan and K. Ho. A simple and efficient estimator for hyperbolic location. *IEEE Trans. Signal Process.*, 29:1905–1915, 1994.

[53] J. L. Leva. An alternative closed-form solution to the GPS pseudo-range equations. *IEEE Trans. Aerosp. Electron. Syst.*, 32(4):1430–1439, 1996.

[54] Alfred Kleusberg. Analytical GPS navigation solution. In *Grafarend EW., Krumm FW. and Schwarze VS. (eds) Geodesy -the Challenge of the 3rd Millennium*, pages 93–96, Heidelberg, 2003.

[55] J. Beutel. Geolocation in a picoradio environment. Master's thesis, UC Berkeley, 1999.

[56] J. J. Caffery. A new approach to the geometry of TOA location. In *Proc. IEEE Vehic. Technol. Conf. (VTC)*, volume 4, pages 1943–1949, Boston, MA, September 2000.

[57] Z. Li, W. Trappe, Y. Zhang, and B. Nath. Robust statistical methods for securing wireless localization in sensor networks. In *Proc. IEEE Int. Symposium on Info. Process. in Sensor Networks (IPSN)*, pages 91–98, Los Angeles, CA, April 15, 2005.

[58] S. Venkatesh and R. M. Buehrer. A linear programming approach to NLOS error mitigation in sensor networks. In *Proc. IEEE Int. Symposium on Info. Process. in Sensor Networks (IPSN)*, pages 301–308, Nashville, TN, April 2006.

[59] I. Guvenc, S. Gezici, F. Watanabe, and H. Inamura. Enhancements to linear least squares localization through reference selection and ML estimation. In *Proc. IEEE Int. Conf. Wireless Commun. Networking (WCNC)*, pages 284–289, Las Vegas, NV, March 31/April 3, 2008.

[60] J. Yan, C. C. J. M. Tiberius, G. Bellusci, and G. J. M. Janssen. A novel non-iterative localization solution. In *IEEE GLOBECOM 09*, Honolulu, November 30/December 4, 2009.

[61] Yiteng Huang and Jacob Benesty. Real-time passive source localization: A practical linear-correction least-squares approach. *IEEE Trans. Speech Audio Process.*, 9(8):943–955, November 2001.

[62] K. W. Cheung, H. C. So, W.-K. Ma, and Y. T. Chan. Least squares algorithms for time-of-arrival-based mobile location. *IEEE Trans. Signal Process.*, 52(4):1121–1128, April 2004.

[63] J. Yan and G. Bellusci. Low-complexity ultra-wideband indoor positioning. In *Proc. ION Global Navigation Satellite Systems (GNSS)*, pages 1013–1024, Savannah, September 22–25, 2009.

[64] J. Yan, C. C. J. M. Tiberius, P. J. G. Teunissen, G. Bellusci, and G. J. M. Janssen. A framework for low complexity least-squares localization with high accuracy. *IEEE Trans. Signal Process.*, 57(12), 2009.

[65] Jeffrey Hightower and Gaetano Borriello. Location systems for ubiquitous computing. *Computer*, 34(8):57–66, 2001.

[66] K. Pahlavan, Xinrong Li, and J.P. Makela. Indoor geolocation science and technology. *Communications Magazine, IEEE*, 40(2):112–118, Feb 2002.

[67] M. Vossiek, L. Wiebking, P. Gulden, J. Wieghardt, C. Hoffmann, and P. Heide. Wireless local positioning. *Microwave Magazine, IEEE*, 4(4):77–86, Dec. 2003.

[68] A.H. Sayed, A. Tarighat, and N. Khajehnouri. Network-based wireless location: challenges faced in developing techniques for accurate wireless location information. *Signal Processing Magazine, IEEE*, 22(4):24–40, July 2005.

[69] F. Gustafsson and F. Gunnarsson. Mobile positioning using wireless networks: possibilities and fundamental limitations based on available wireless network measurements. *Signal Processing Magazine, IEEE*, 22(4):41–53, July 2005.

[70] Guolin Sun, Jie Chen, Wei Guo, and K.J.R. Liu. Signal processing techniques in network-aided positioning: a survey of state-of-the-art positioning designs. *Signal Processing Magazine, IEEE*, 22(4):12–23, July 2005.

[71] G. Bellusci, G. J. M. Janssen, J. Yan, and C. C. J. M. Tiberius. Low complexity ultra-wideband ranging in indoor multipath environments. In *Proc. IEEE/ION Position Location and Navigation Symposium (PLANS)*, Monterey, May 6–8, 2008.

[72] Bardia Alavi and Kaveh Pahlavan. Modeling of the distance error for indoor geolocation. In *Proc. IEEE Int. Conf. Wireless Commun. Networking (WCNC)*, pages 668–672, March 2003.

[73] Y. Qi and H. Kobayashi. On relation among time delay and signal strength based geolocation methods. In *Proc. IEEE Global Telecommunications Conf. (GLOBECOM'03)*, volume 7, pages 4079–4083, San Francisco, CA, December 2003.

[74] G. Bellusci, J. Yan, G. J. M. Janssen, and C. C. J.M. Tiberius. An Ultra-Wideband positioning demonstrator using audio signals. In *Proc. 4th Workshop on Positioning, Navigation and Communication*, pages 71–76, Hannover, March 2007.

[75] J. Yan, C. C. J. M. Tiberius, G. Bellusci, and G. J. M. Janssen. Analyzing non-linearity effect for indoor positioning using an acoustic ultra-wideband system. In *IEEE Workshop on Positioning, Navigation and Communication (WPNC)*, pages 95 – 101, Hannover, Mar 2008.

[76] N. Alsindi, B. Alavi, and K. Pahlavan. Spatial characteristics of UWB TOA-based ranging in indoor multipath environments. In *Proc. IEEE (PIMRC)*, pages 1–6, September 2007.

[77] Chia-Chin Chong, Fujio Watanabe, and title = Proc. IEEE (WCNC) year = 2007 month = Mar pages = 1559-1564 M.Z. Win, title = Spatial characteristics of UWB TOA-based ranging in indoor multipath environments.

[78] S.-S. Woo, H.-R. You, and J.-S. Koh. The NLOS mitigation technique for position location using IS-95 CDMA networks. In *IEEE VTC*, pages 2556–2560, Sep, 2000.

[79] Li Cong and Weihua Zhuang. Nonline-of-sight error mitigation in mobile location. *Wireless Communications, IEEE Transactions on*, 4(2):560–573, March 2005.

[80] K. Yu and Y.J. Guo. Statistical NLOS identification based on AOA, TOA, and signal strength. *Vehicular Technology, IEEE Transactions on*, 58(1):274–286, Jan. 2009.

[81] H. Hashemi. The indoor radio propagation channel. *Proceedings of the IEEE*, 81(7):943–968, Jul 1993.

[82] T.S. Rappaport. *Wireless Communications: Principles and Practice*. Prentice-Hall, Englewood Cliffs, NJ, 1996.

[83] G.D. Durgin. *Space-Time Wireless Channels*. Prentice Hall, Englewood Cliffs, NJ, 2002.

[84] J. Yan, C. C. J. M. Tiberius, G. Bellusci, and G. J. M. Janssen. Feasibility of Gauss-Newton method for indoor positioning. In *Proc. IEEE/ION Position Location and Navigation Symposium (PLANS)*, Monterey, May 6–8, 2008.

[85] A. G. Stove. Linear FMCW radar techniques. In *IEE Proceedings-F, Radar and Signal Processing*, volume 139, pages 343–350, October 1992.

[86] O. Cramer. The variation of the specific heat ratio and the speed of sound in air with temperature, pressure, humidity, and CO2 concentration. *Journal of the Acoustical Society of America*, 93(5):25102516, May 1993.

[87] P. J. G. Teunissen. Nonlinear inversion of geodetic and geophysical data: diagnosing nonlinearity. *Lecture notes in Earth Science*, (29):241–264, 1990.

[88] P. J. G. Teunissen. A note on the use of Gauss' formula in nonlinear geodetic adjustments. *Statistics and Decisions*, (2):455–466, 1984.

[89] P. J. G. Teunissen, D. G. Simons, and C. C. J. M. Tiberius. *PROBABILITY AND OBSERVATION THEORY*. Lecture notes AE2CE01, Faculty of Aerospace Engineering Delft University of Technology, December 2009.

[90] P. J. G. Teunissen. First and second moments of non-linear least-squares estimators. *Bulletin Geodesique*, 63(3):253–262, 1989.

[91] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2):164–168, 1944.

[92] G. H. Golub and C. F. van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, Md, 1989.

[93] A. Rabbachin, I. Oppermann, and B. Denis. ML time-of-arrival estimation based on low complexity UWB energy detection. In *Ultra-Wideband, The 2006 IEEE 2006 International Conference on*, pages 599 –604, 2006.

[94] J. Schroeder, S. Galler, K. Kyamakya, and K. Jobmann. NLOS detection algorithms for Ultra-Wideband localization. In *Positioning, Navigation and Communication, 2007. WPNC '07. 4th Workshop on*, pages 159 –166, 2007.

[95] Ismail Guvenc, Chia-Chin Chong, Fujio Watanabe, and Hiroshi Inamura. NLOS identification and weighted least-squares localization for UWB systems using multipath channel statistics. In *EURASIP Journal on Advances in Signal Processing*, volume 2008, 2008.

[96] S. Venkatraman and Jr. Caffery, J. Statistical approach to non-line-of-sight BS identification. In *Wireless Personal Multimedia Communications, 2002. The 5th International Symposium on*, volume 1, pages 296 – 300, 27-30 2002.

[97] H. Shimizu, H. Masui, M. Ishii, and K. Sakawa. LOS and NLOS path-loss and delay characteristics at 3.35 GHz in a residential environment. In *Antennas and Propagation Society International Symposium, 2000. IEEE*, volume 2, pages 1142 –1145, 2000.

[98] F. Benedetto, G. Giunta, A. Toscano, and L. Vegni. Dynamic LOS/NLOS statistical discrimination of wireless mobile channels. In *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pages 3071 –3075, 2007.

[99] C. Tepedelenlioglu, A. Abdi, and G.B. Giannakis. The Ricean K factor: estimation and performance analysis. *Wireless Communications, IEEE Transactions on*, 2(4):799 – 810, July 2003.

[100] J. Borras, P. Hatrack, and N.B. Mandayam. Decision theoretic framework for NLOS identification. In *Vehicular Technology Conference, 1998. VTC 98. 48th IEEE*, volume 2, pages 1583 –1587, 1998.

[101] K. Yu and Y.J. Guo. Statistical NLOS identification based on AOA, TOA, and signal strength. *Vehicular Technology, IEEE Transactions on*, 58(1):274 –286, Jan. 2009.

[102] Pi-Chun Chen. A non-line-of-sight error mitigation algorithm in location estimation. In *Wireless Communications and Networking Conference, 1999. WCNC. 1999 IEEE*, pages 316–320, 1999.

[103] L. Cong and Weihua Zhuang. Non-line-of-sight error mitigation in TDOA mobile location. In *Global Telecommunications Conference, 2001. GLOBECOM '01. IEEE*, volume 1, pages 680 –684, 2001.

[104] M.P. Wylie and J. Holtzman. The non-line-of-sight problem in mobile location estimation. In *Universal Personal Communications, 1996. Record., 1996 5th IEEE International Conference on*, volume 2, pages 827 –831, 1996.

[105] T. Mayer. Abhandlung über die umwalzung des monds um seine axe und die scheinbare bewegung der mondsflecten. In *Kosmogr. Nachr. Samml. auf das Jahr 1748*, pages 52–183, 1750.

[106] P. J. G. Teunissen. Towards a least-squares framework for adjusting and testing of both functional and stochastic models. *LGR/MGP report, Delft University of Technology, Faculty of Civil Engineering and Geosciences, Department of MGP*, pages 16–18, 1988.

[107] Giovanni Bellusci. NLoS detection using combined RSS and ToA. *Personal Communication*, 2009, Delft University of Technology.

[108] Walter Rudin. *Principles of mathematical analysis*. McGraw-Hill Book Co., New York, 1976.

Samenvatting

Algoritmen voor binnenshuis-plaatsbepalingssystemen met Ultra-Wideband signalen

Plaatsbepalingssystemen en -technieken zijn de afgelopen jaar meer en meer in de belangstelling komen te staan, in het bijzonder met het beschikbaar komen van satellietnavigatie-technologie (GPS), en ontwikkelingen op het gebied van navigatie in gebouwen. Het werk in dit proefschrift is uitgevoerd in het kader van het onderzoeksproject "HERE - binnenshuis-plaatsbepaling gebaseerd op Ultra-Wideband (UWB) radiosignalen", dat zich richt op de ontwikkeling van een alternatieve oplossing voor het binnenshuis-plaatsbepalingsprobleem, omdat satellietnavigatie in deze omgeving in het algemeen niet goed presteert, vanwege sterke signaalverzwakking en intense zogenaamde multipad effecten (die optreden door veelvuldige reflecties van het signaal). Het project richt zich specifiek op technologie gebaseerd op afstandmeting. waarbij twee stappen zijn te onderscheiden: 1) het meten van de afstand middels het tijdstip van aankomst van het signaal (Time-of-Arrival, ToA), het tijdsverschil tussen de aankomst van twee signalen (Time Difference of Arrival, TDoA), en de ontvangen signaal sterkte (Received Signal Strength, RSS), 2) de feitelijke plaatsbepaling op basis van de gemeten afstanden. Dit vierjarig onderzoeksproject is uitgevoerd door een team van twee promovendi en twee begeleiders. Als één van de twee promovendi, heeft de auteur van dit proefschrift zich gericht op de plaatsbepalingsaspecten van het project, en de belangrijkste bevindingen worden hieronder samengevat.

Het gros van de plaatsbepalingssystemen, zoals GPS satellietnavigatie, gebruikt zogenaamde 'iterative descent' (ID) methoden om het (niet-lineair) plaatsbepalingsprobleem in kleinstekwadraten zin op te lossen. Het toepassen van dergelijke methoden bij binnenshuis-plaatsbepaling kent de volgende drie bezwaren:

- Het is veel moeilijker om een goede benaderde waarde (voor de oplossing) te verkrijgen, welke van cruciaal belang is voor de ID methoden om te starten en naar de correcte oplossing te convergeren, en om sneller te convergeren. Voor aardse toepassingen van satellietnavigatie kan een goede benaderde waarde heel eenvoudig verkregen worden door het middelpunt van de aarde te kiezen - een eventuele andere (lokale) oplossing ligt doorgaans ver weg in de ruimte. Voor binnenshuis-plaatsbepaling ligt dit niet zo eenvoudig, daar van de positie van de gebruiker, ten opzichte van de basis-stations (zenders), weinig bekend is. In dit proefschrift stellen we voor om zogenaamde directe methoden te gebruiken om een benaderde waarde te bepalen, ofwel simpelweg het geometrisch middelpunt van de ontvangen zenders te nemen. In dit proefschrift wordt een gedetailleerde analyse gemaakt van de bestaande directe

methoden.

- De niet-lineaire kleinstekwadraten schatter kent een inherent systematische fout, zelfs wanneer de afstandmetingen zonder dergelijke fouten zijn, doordat de verwachting van de hogere orde termen in de uitdrukking van de uiteindelijke schatter niet gelijk aan nul is. Deze fout is in het algemeen verwaarloosbaar bij satellietnavigatie, met enorme afstanden tussen satelliet en gebruiker, maar kan problematisch zijn bij plaatsbepaling binnenshuis, omdat hier de afstanden veel korter zijn. De fout ten gevolge van niet-lineariteit is geanalyseerd, en een procedure is voorgesteld om de significantie van de fout te toetsen. Validatie is hierbij uitgevoerd op basis van metingen met een akoestisch UWB systeem.

- Het iteratieve karakter van de zogenaamde ID methoden kan tot een te zware rekenbelasting leiden voor binnenshuis toepassingen, die doorgaans kleine, energiezuinige systemen vragen. Er is gezocht naar mogelijkheden om de rekenbelasting van de traditionele ID methoden te beperken, en een nieuwe aanpak voor plaatsbepaling is voorgesteld. Het meer-dimensionale niet-lineaire plaatsbepalingsprobleem wordt daarbij eerst getransformeerd naar een lagere dimensie en dan iteratief opgelost. In drie-dimensionale plaatsbepalingssystemen kan hiermee een reductie van 67% op de rekenbelasting in iedere iteratiestap behaald worden. En niettemin kunnen op deze manier nauwkeurige plaatsbepalingsresultaten behaald worden, in het bijzonder met afstandmetingen gebaseerd op tijdverschil (TDoA).

In het algemeen bestaan er geen strikt directe (dus niet iteratieve) kleinstekwadraten oplossingen voor niet-lineaire problemen. Echter, door een aantal aannamen of vereenvoudigingen te maken, kunnen wel directe kleinstekwadraten plaatsbepalings-algoritmen ontwikkeld worden. In de literatuur is een groot aantal directe methoden beschreven, in domeinen zoals radar-, luchtvaart- en ruimtevaarttechniek, oceaan- en maritieme techniek, (akoestische) signaalverwerking en draadloze communicatie. Een aantal van deze methoden blijkt, verbazingwekkend, identiek te zijn, hoewel de afleidingen doorgaans zeer verschillend zijn. Een gedetailleerde studie en zuivere classificatie van de methoden helpt om te komen tot een beter begrip, en dat is één van de centrale bijdragen van dit proefschrift. Dit kan onderzoekers en ontwikkelaars helpen om de juiste keuze te maken voor hun toepassing. Directe methoden leveren eenvoudig te berekenen schattingen, maar zijn suboptimaal als gevolg van de gemaakte vereenvoudigingen. Gebaseerd op de studie van bestaande directe methoden, is een nieuwe niet-iteratieve methode ontwikkeld om de plaatsbepalings-nauwkeurigheid te verbeteren. Een theoretisch bewijs wordt gegeven dat deze methode een betere schatter levert, in de zin dat ze overeenkomt met een gelijke, of kleinere waarde voor de oorspronkelijke kleinstekwadraten doelfunctie. Dit wordt bereikt door twee vergelijkbare modellen te gebruiken, die beide de benodigde modellerings-condities volledig in rekening brengen. Het niet-iteratieve karakter van de methode maakt het algoritme aantrekkelijk voor goedkope, energiezuinige toepassingen.

Het is welbekend dat het afwezig zijn van directe zichtlijnen tussen zenders en ontvanger (Non Line-of-Sight, NLoS) ten behoeve van voortplanting van het radiosignaal, één van de grote nadelige effecten is op de plaatsbepalings-nauwkeurigheid. De detectie, en het teniet doen van dergelijke effecten is een belangrijk onderwerp in binnenshuis-plaatsbepaling. Een overzicht wordt gegeven van bestaande manieren voor detectie van geen-zichtlijn sit-

uaties en het teniet doen van de nadelige effecten ervan, en een viertal nieuwe aanpakken wordt beschreven, welke gebaseerd zijn op systematische hypothesetoetsing van plaatsbepalingsmodellen. De idee is om de informatie van afstandmetingen gebaseerd op tijd en signaalsterkte te combineren, en tegelijkertijd het feit te benutten dat alle verzamelde metingen betrekking hebben op dezelfde onbekende positie. De vier voorgestelde aanpakken gebruiken allen de combinatie van tijd- en signaalsterkte-metingen, omdat 1) tijdmetingen doorgaans erg nauwkeurig zijn vergeleken met signaalsterkte-metingen, 2) de (statistische) verdelingen van signaalsterkte-metingen bij direct-zicht, en bij geen-direct-zicht sterk van elkaar verschillen. De rekenbelasting van de aanpakken vermindert als er meer vereenvoudigingen gemaakt worden, en de prestaties lopen dan in het algemeen ook terug, behalve in het geval waar de meeste zender-ontvanger-verbindingen geen zichtlijn kennen. Validatie-resultaten laten zien dat, met de volle UWB signaal bandbreedte van 7.5 GHz, een correcte beslissing tussen zicht en geen-zicht behaald kan worden in 99% van de gevallen.

Curriculum Vitae

Junlin Yan was born in Shanghai, China, in 1981. In 2003, he received his B.E. degree in Communication Engineering from Fudan University, Shanghai. He moved to Gothenburg, Sweden, in 2004 to study at the Chalmers University of Technology. In 2006, he received his M.Sc. degree with honor in Digital Communication Systems and Technology from Chalmers University of Technology, Sweden. In the same year, he joined the Mathematical Geodesy and Positioning group, faculty Aerospace Engineering, Delft University of Technology as a Ph.D. student. His doctoral research project was "HERE: indoor positioning based on UWB radio signals", and his work focuses on positioning algorithms for indoor applications.

Acknowledgments

This thesis could not have been accomplished without the assistance of many people, to whom I would like to express my gratitude.

First of all, I am heartily thankful to my supervisor Dr. Christian Tiberius, who led me into the world of positioning. His constant scientific guidance and support were essential to me during the entire period of my research. He gave me a lot of freedom to make decisions. He was very generous with his praises when I made progress, because of which my confidence and self-esteem were boosted. He was full of forgiveness and pardon when I made mistake, from which I learned what is mercy and generosity. I would like to thank my promotor, Prof. Peter Teunissen, who offered me the opportunity to pursue my Ph.D. degree at the MGP group in the first place and gave me many valuable advices on non-linear least-squares estimation. I am also grateful to my co-supervisor Dr. Gerard Janssen, for his support, many valuable comments, and helping me finalize this thesis. Many thanks to Giovanni Bellusci, colleague Ph.D. candidate of the "HERE" project, from whom I learned a lot during all the valuable discussions and cooperations.

The work would not have been so pleasant without many colleagues in the MGP group. I express my appreciation to Ria Scholtes, File Koot-Stomp, and Rebeca Domingo, for their professional and prompt administrative supports. I express my sincere appreciation to my former and current roommates: Dr. Alireza Amiri, Anh Quan Le, Yongjin Moon, Benoît Muth, Peter de Bakker and Dr. Yahya Memarzadeh, for all those research- and life-related helps that they have offered me. I also greatly appreciate Gabriele Giorgi and Peter Buist for their friendship. Furthermore, I would like to thank all members of the MGP group at TU-Delft for the pleasant working environment.

I express my thanks to all the committee members for spending time on my thesis review and for their important comments.

I would like to thank the Technical Foundation (STW) for the financial support.

Last but not the least, I would like to manifest my deep gratitude to my parents, Liegen and Jinye, for their endless support and sacrificial love, and my beloved girl friend Juan, who is the source of joy and the spiritual pillar of mine.

Junlin Yan
September 2010
Düsseldorf, Germany

Notation and Symbols

Mathematical and Statistical Notation and Operators

$\underline{(.)}$	vector of random variables
$\hat{(.)}$	vector of estimates
$\underline{\hat{(.)}}$	vector of estimators
$\underline{(.)}_i$	i-th value of a vector or the vector representing the i-th link
$(.)_r$	reference vector
$(.)_{i,j}$	value of a matrix on i-th row and j-th column or j-th row/element matrix/vector $(.)_i$
$(.)_{i,j,k}$	value of a tensor on i-th row, j-th column of the k-th matrix or value of the matrix $(.)_i$ on j-th row and k-th column
$\perp$	orthogonal complement (is orthogonal to)
$(.)^T$	transpose of a matrix
$(.)^{-1}$	inverse of a matrix
$\partial_x y$	first order partial derivative of y, w.r.t. x
$\partial^2_{xx^T} y$	second order partial derivative of y, w.r.t. x
$\|.\|$	norm of a vector
$\prod(.)$	production
$\sum(.)$	summation
$\mathcal{N}(\mu, Q)$	normal distribution with mean μ and variance matrix Q
$\chi^2(n, \lambda)$	chi-square distribution with degrees of freedom n and non-central parameter λ
μ	mean or bias vector
σ	standard deviation
λ	Lagrange multiplier
$\ln$	natural logarithm
$\log_{10}$	base 10 logarithm
$\triangleq$	by definition equal to
C{.}	mathematical covariance operator
D{.}	mathematical dispersion operator
det(.)	determinant
dim(.)	dimension of a matrix or vector
E{.}	mathematical expectation operator
$f_{\underline{y}}(y)$	probability density function of the random vector $\underline{y}$
$f_{\underline{y}\|\underline{S}}(y_{\text{all}}\|S)$	conditional probability density function of the random vector $\underline{y}$, given state vector $\underline{S} = S$

$o(.)$	Landau order term
$P_{(.)}$	projection matrix
$P_{(.)}^{\perp}$	orthogonal matrix
$P(.)$	probability
$\mathbb{R}^m$	real Euclidean space of dimension m
rank$(.)$	rank of a matrix (independent columns or rows of a matrix)
trace$(.)$	sum of elements on the main diagonal of a square matrix

Symbols

β_k	probability of incorrect decision when H_k is true
Δ	range difference
γ_k	probability of correct decision when H_k is true
κ	kurtosis, or unknown non-linear term of x, used by direct methods
∇_k	the NLoS bias vector of the k-th state
Φ_j	parameter space put forward by the j-th hypothesis H_j
τ_c	clock misalignment between a transmitter and a receiver
τ_{MED}	mean excess delay
τ_{RMS-DS}	root mean squared delay spread
τ_{ToA}	time of arrival
τ_{ToT}	time of transmission
(θ, ϕ)	angle coordinates in a spherical system
A	design matrix
$A(x)$	mapping from unknown x to measurement y
B	bandwidth
b_c	bias in pseudorange due to clock misalignment
b_s	constant shift added to b_c in MOF methods
c	speed of light
c_s	speed of sound
d	distance vector
d_{ss}	system scale, defined as the mean value of the distances between the transmitters and the geometric center of the transmitters
d_{rt}	distance between the receiver and the geometric center of the transmitters
e	error vector
$F(.)$	objective function
$f(.)$	constraint function used by direct methods
f_H	upper frequency of the -10 dB emission point
f_L	lower frequency of the -10 dB emission point
f_s	the sampling frequency
GL	generalized likelihood
H	humidity or hypothesis
$h(t)$	channel impulse response
$h_{(.)}$	kernel function width

k_n	normal curvature
$K_{(.)}$	kernel function
$L(z)$	function that transforms x to a 1-dimensional unknown z
m	dimension of the measurement vector y
$M(y)$	mapping from measurement y to unknown x
n	pathloss exponent, dimension of x, or normal unit vector
P	received power
C_k	the state matrix of size $2m \times q_k$
q_k	number of NLoS links
Rx	receiver
R	rotation matrix
s	shadowing effect
$\underline{S}_{m\times 1}$	the random state vector, $P(\underline{S} = S_k) = \frac{1}{2^m}$, with $k = 0, 1, ..., 2^m - 1$
$[S_k]_{m\times 1}$	possible realizations of the state vector, with 0 entries representing LoS links and 1 entries representing NLoS ones
Tx	transmitter
T	temperature, transformation matrix, transpose operator, or decision variable
W	weight matrix
x	unknown vector, $x = x_u$ with ToA/RSS, $x = [x_u^T, b_c]^T$ with TDoA
x_u	unknown Cartesian coordinates of the user
x_s	constant shift added to x_s in MOF methods
y	measurement vector

Acronyms

AoA	Angle of Arrival
AWGN	Additive White Gaussian Noise
CDMA	Code Division Multiple Access
CEPT	European Conference of Postal and Telecommunications Administrations
CIR	Channel Impulse Response
CM	Confidence Metric
CRLB	Cramer-Rao Lower Bound
CT I	Classification Type I
CT II	Classification Type II
ECC	Electronic Communications Committee
EEMCS	Electrical Engineering, Mathematics and Computer Sciences
EIRP	Equivalent Isotropically Radiated Power
EVD	Eigenvalue Decomposition
FB	Full Bandwidth
FCC	Federal Communications Commission
FMCW	Continuous Wave Frequency Modulation
GL	Generalized Likelihood
GLRT	Generalized Likelihood Ratio Test
GNSS	Global Navigation Satellite Systems
GPS	Global Positioning System
ID	Iterative Descent
IFFT	Inverse Fast Fourier Transform
LB	Lower Bound
LoS	Line of Sight
MAP	Maximum A Posterior
MED	Mean Excess Delay
MGP	Mathematical Geodesy and Positioning
ML	Maximum Likelihood
MOF	Multiple Objective Function
MOF II	Multiple Objective Function Method Type II
MSE	Mean Squared Error
PDF	Probability Distribution Function

RB	Resolution Bandwidth
RMS-DS	Root Mean Square Delay Spread
RMSE	Root Mean Squared error
RSS	Received Signal Strength
SLRT	Simple Likelihood Ratio Test
SNR	Signal-to-Noise Ratio
SOF	Single Objective Function
ST I	Simple likelihood ratio test Type I
ST II	Simple likelihood ratio test Type II
STD	STandard Deviation
TDoA	Time Difference of Arrival
TDMA	Time Division Multiple Access
TLS	Transformed Least-Squares
ToA	Time of Arrival
ToT	Time of Transmission
UB	Upper Bound
UWB	Ultra-WideBand