TECHNISCHE
UNIVERSITÄT
DARMSTADT

Wireless Network Localization via Cooperation

Vom Fachbereich 18
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

von

Di Jin, M.Sc.

Referent:	Prof. Dr.-Ing. Abdelhak M. Zoubir
Korreferent:	Prof. Dr. Hing Cheung So
Korreferent:	Prof. Dr.-Ing. Feng Yin

Darmstadt, 2020

Jin, Di — *Wireless Network Localization via Cooperation*
Darmstadt, Technischen Universität Darmstadt
Jahr der Veröffentlichung der Dissertation auf TUprints: 2021
URN: urn:nbn:de:tuda-tuprints-196540
Tag der mündlichen Prüfung: 17.12.2020

Veröffentlicht unter CC BY-SA 4.0 International
`https://creativecommons.org/licenses/`

Acknowledgments

I sincerely thank Prof. Abdelhak Zoubir for opening my journey to the field of statistical signal processing, for encouraging me to pursue a PhD, and for supporting me during my doctoral studies. To me, his excellent and fascinating lectures are one of most invaluable parts of my academic life. I also thank Prof. Hing Cheung So for being my co-supervisor and providing valuable feedback on my work.

Special thanks to my co-supervisor Prof. Feng Yin. Thank you for always offering me your guidance and support, both professionally and personally, for making my doctoral studies as pleasurable as possible, and for hosting me at the Chinese University of Hong Kong, Shenzhen.

I also wish to thank Prof. Fredrik Gustafsson and Carsten Fritsche for the pleasant collaboration. I further thank Prof. Marius Pesavento, Prof. Jürgen Adamy and Prof. Klaus Hofmann for being on my committee.

I wish to express my appreciation to all of my current and former SPG colleagues: Nevine Demitri, Mark Ryan Leonard, Dominik Reinhard, Martin Gölz, Christian Weiß, Jürgen Hahn, Adrian Šošić, Ann-Kathrin Seifert, Simon Rosenkranz, Michael Fauß, Christian Debes, Michael Muma, Mahmoud El-Hindi, Huiping Huang, Jasin Machkour, Afief Dias Pambudi, Christian Schroth, Aylin Tastan, Ziliang Qiao, Sergey Sukhanov, Amare Kassaw, Freweyni Teklehaymanot, Tim Schäck, Wenjun Zeng, Patricia Binder, Lala Khadidja Hamaidi, Sahar Khawatmi, Wassim Suleiman, Sara Al-Sayed, Gökhan Gül, Michael Leigsnering,Tai Fei, Hauke Fath, Prof. Roy Howard. Thank you all for providing a friendly environment, for offering me reliable support at work, for sharing so many scientific or funny talks. Special thanks to Renate Koschella for her constant support and comfort throughout all these years.

I wish to thank my parents Baolin Jin and Jie Huang, my parents-in-law Jianqi Huang and Xiuling Song, and my brother Hongluan Jin. Thank you for your unconditional support and always being there for me.

Thank you, Xin, for standing beside me with your unconditional support. Words would never say how grateful I am to you. Thank you, my little girl, Serena, for bringing me so much joy and surprise in every way. Thank you, my baby boy, Erik, for your company at all times and for not-kicking me during the defense.

Darmstadt, September 2021

Kurzfassung

Diese Dissertation leistet einen Beitrag zum Themengebiet der Lokalisierung in drahtlosen Netzwerken, indem zwei Klassen von kooperativen Lokalisierungsmethoden für gemischte Lokalisierungsumgebungen mit und ohne direkte Sichtverbindung (LOS/NLOS) bereitgestellt werden. Als aus der praktischen Anwendung stammende Einschränkungen wird hier angenommen, dass weder NLOS-Identifikation, noch experimentelle Messungen durchführbar sind. Abhängig vom Umfang des Vorwissens über das statistische Messmodell, werden zwei Klassen kooperativer Lokalisierungsmethoden entwickelt. Zwei wichtige Beiträge sind zum einen Methoden, die eine zufriedenstellende Lokalisierungsgenauigkeit bieten, während das benötigte Wissen über das statistische Messmodell reduziert wird, zum anderen Methoden, die eine deutlich verbesserte Lokalisierungsleistung bieten, ohne eine gute Initialisierung zu benötigen.

Der erste Teil der Dissertation behandelt die kooperative Lokalisierung in homogenen, gemischen LOS/NLOS Umgebungen, wobei der Pfadverlust als Hauptmodellparameter unbekannt ist. Der gewählte Ansatz basiert darauf, die Position und den Pfadverlust als Zufallsvariablen zu modellieren und die marginale, posteriore Verteilung der unbekannten Parameter zu inferieren, aus denen Positionsschätzung und Unsicherheitsinformation gewonnen werden können. Dies wird mit Hilfe von *message passing* Methoden erreicht, bei denen zwei Mengen von Funktionen, die als *messages* und *beliefs* bezeichnet werden, iterativ aktualisiert werden. Durch die Kombination von variabler Diskretisierung und Monte-Carlo-basierten numerischen Approximationsschemata erhält man zwei Mengen von Funktionen. Eine solche numerische Vorgehensweise erlaubt, die Regel zur Verbesserung der Nachrichten approximativ zu implementieren und somit den Rechenaufwand zu limitieren. Zusätzlich werden für Netzwerke mit Low-End-Sensoren, die nur quantisierte RSS-Messungen liefern, *message passing* Methoden und deren parametrische Varianten mit geringer Komplexität abgeleitet.

Im zweiten Teil der Dissertation wird der allgemeinere Fall behandelt, dass kein statistisches Wissen über die LOS-/NLOS-Messfehler vorhanden ist, während präzisere Entfernungsmessungen zur Verfügung stehen. Diese reagieren recht empfindlich auf NLOS-Ausbreitung. Der Bias-Term jeder Abstandsmessung wird als unbekannter Parameter behandelt und es wird gezeigt, dass die Bias-Parameter in LOS-dominierenden Szenarien dünnbesetzt sind. Diese Eigenschaft wird dann ausgenutzt, indem zu den konventionellen Kostenfunktionen ein zusätzlicher Bestandteil hinzugefügt wird, der einen dünnbesetzten Bias-Vektor fördert. Dies führt zu einer generischen, *sparsity-promoting* regularisierten Methode. Durch Beschränkung der Kostenfunktion wird eine alternative, generische, *bound-constrained* regularisierte Methode entwickelt. Die Kostenfunk-

tionen dieser beiden generischen Methoden werden so definiert, dass sie relaxiert werden können, um mithilfe zweier semi-definite program (SDP)-basierter Methoden das globale Optimum zu erreichen. Es wird theoretisch gezeigt, dass diese beiden SDPs unter bestimmten Bedingungen in dem Sinne gleichwertig sind, dass sie sich die gleiche optimale Lösung teilen. Eine große Herausforderung der beiden vorgeschlagenen SDPs liegt in der Wahl eines geeigneten Regularisierungsparameters. Es wird eine effiziente, datengetriebene Methode entwickelt, um den Regularisierungsparameter zu erhalten, welche auf der speziellen Struktur des *bound-contrained* regularisierten SDPs basiert. Schließlich werden numerische Ergebnisse, basierend auf synthetischen und experimentellen Daten, vorgestellt. Es wird gezeigt, dass die vorgeschlagenen SDP Methoden insgesamt gute Lokalisierungsergebnisse liefern.

Abstract

This dissertation details two classes of cooperative localization methods for wireless networks in mixed line-of-sight and non-line-of-sight (LOS/NLOS) environments. The classes of methods depend on the amount of prior knowledge available. The methods used for both classes are based on the assumptions in practical localization environments that neither NLOS identification nor experimental campaigns are affordable. Two major contributions are, first, in methods that provide satisfactory localization accuracy whilst relaxing the requirement on statistical knowledge about the measurement model. Second, in methods that provide significantly improved localization performance without the requirement of good initialization.

In the first half of the dissertation, cooperative localization using received signal strength (RSS) measurements in homogeneous mixed LOS/NLOS environments is considered for the case where the key model parameter, the path loss exponent, is unknown. The approach taken is to model the positions and the path loss exponent as random variables and to utilize a Bayesian framework. The goal is to infer the marginal posterior distribution of each unknown parameter, from which a position estimate, as well as its uncertainty information, can be obtained. This is achieved by using message passing methods in which two sets of functions, referred to as *messages* and *beliefs*, are iteratively updated. By combining variable discretization and Monte-Carlo-based numerical approximation schemes, two sets of functions are obtained. Such a numerical strategy allows the message updating rule to be implemented approximately while keeping the computational complexity affordable. Additionally, for networks with low-end sensors that only provide quantized RSS measurements, message passing algorithms and their parametric variants of low complexity are derived.

The second part of the thesis considers the more general case where statistical knowledge of the LOS/NLOS measurement errors is completely unknown, and range measurements, which are believed to be more accurate but quite sensitive to NLOS propagation, are available. The bias associated with each range measurement is modeled as an unknown parameter, and it is shown that bias parameters possess a sparsity property in LOS-heavy scenarios. This sparsity is exploited by introducing a sparsity-promoting term in the conventional cost functions, giving rise to a generic sparsity-promoting regularized formulation. By bounding the cost function, an alternative generic bound-constrained regularized formulation is developed. To ensure global optimality, the cost functions in these two generic formulations are specified so that they can be conveniently solved as two semi-definite programs (SDPs). It is theoretically shown, for certain conditions, that these two SDPs are equivalent in the sense that they share the

same optimal solution. A major challenge of these two SDPs lies in the selection of an appropriate regularization parameter. An efficient data-driven strategy is developed to determine the regularization parameter and this is based on the special structure of the bound-constrained regularized SDP. Finally, numerical results, based on both synthetic and experimental data, are detailed. It is shown that the devised SDP approach provides overall good localization performance.

Contents

Chapter 1
Introduction

The advances in microelectromechanical system (MEMS) technology and wireless communications underpin the recent development of different types of wireless networks, such as wireless sensor networks, ultra-wideband wireless networks, Internet-of-Things, etc. An indication of the growth and importance of wireless sensor networks can be found in their market value: estimated to be USD 47 billion in 2019 and is expected to reach USD 124 billion by 2025 [WSN].

Location-awareness is one essential and enabling feature of many wireless applications. In some contexts, for example with environmental monitoring, wireless nodes have to be deployed over a large area and this is only feasible if they are distributed, for example, from an airplane. Naturally, with such distribution, there is a high degree of randomness associated with the location of each wireless node. In such a context, knowledge of the position of each node is of critical importance as it allows the information from the wireless nodes to be collated in a manner to facilitate high level decision making. Despite the advances of global navigation satellite systems, such as the Global Positioning System (GPS) and BeiDou Navigation Satellite System, it is unrealistic to equip all wireless nodes with GPS/Beidou receivers which are very expensive and extremely power-consuming. Therefore, satellite independent localization solutions have to be developed for such application scenarios.

The focus of this thesis is the development of satellite independent localization methods for these networks. The context is as described above where a network has part of the following characteristics: First, a network which comprises of hundreds of wireless nodes. Second, a network with little or no associated infrastructure, consistent with only a small number of reference nodes whose positions are known. Third, wireless nodes which have a very limited power supply such that the power of the transmitted signal, and hence communication range, from an individual wireless node is limited. Fourth, wireless nodes which have limited processing speed, storage capacity and communication bandwidth. The first two features characterize different types of wireless networks; while the last two pertain in particular to wireless sensor networks. The last two characteristics are consistent with wireless nodes that can only provide position related measurements that have low hardware requirements. More importantly, conventional localization methods adopted in cellular networks and global navigation satellite systems, are not suitable or even not straightforwardly applicable to wireless network applications. This necessitates development of custom localization methods.

Specifically, this thesis details cooperative localization methods for wireless networks, under the considerations of several limitations in real-world scenarios, such as mixed line-of-sight and non-line-of-sight (LOS/NLOS) propagation environments, the absence of full statistical knowledge of LOS/NLOS measurement errors, no training datasets and no offline calibration.

Among different position related measurements, received signal strength (RSS) has gained significant attention due to its ubiquitous use in radio frequency signal systems. For example, an RSS indicator (RSSI) has been encoded in the IEEE 802.15.4 standards [WG815]. Although an RSS measurement provides rather coarse distance information, it can be exploited to enable low-cost, simple and opportunistic localization systems, without the need for additional hardware. This is the motivation for the development of cooperative localization algorithms using RSS measurements which is the focus of the first part of the thesis. To be close to practical scenarios, path loss exponent, the key parameter in the classical log-distance path loss propagation model, is considered to be unknown.

The second half of the thesis considers a more general localization scenario, that of cooperative localization using range measurements in mixed LOS/NLOS environments, in which statistical knowledge about the measurement model is completely unknown. In contrast to RSS measurements, which are generally available in various radio based wireless networks, accurate range measurements are mainly provided in wireless networks with hardware of moderate to high quality, such as ultra wideband wireless networks and some wireless sensor networks.

1.1 Cooperative Localization Methods

In conventional localization situations, a specific target node is usually connected to several reference nodes. At a specific time instance, there is usually available a position related measurement for each of the connections between the target node and the reference nodes. Taking time-of-arrival (TOA) measurements, which is one type of position related measurements, as an example, in a two dimensional context, three TOA measurements, from three distinct reference nodes, are required to estimate the position of a single target node without ambiguity. Such localization is known as non-cooperative localization, has a long history and has been adopted in many contexts including global navigation satellite systems.

To ensure that each target node is connected to at least three reference nodes, it is required that the reference nodes must be densely distributed. In general, this requirement is not met in many wireless networks, primarily due to the relatively low level of network infrastructure (reference nodes). The limitation inherent in a wireless network necessitates an alternative localization paradigm and one with potential is that of *cooperative localization* where the position related measurements between target nodes, called *agents*, are used along side the position related measurements available from the reference nodes, called *anchors*. Research, over the past two decades, has shown that cooperative localization can lead to significantly improved localization accuracy.

A variety of cooperative localization methods have been developed. Table 1.1 provides an overview of the five dominant types of methods. First, maximum-likelihood-estimation-based methods are used when the localization measurements are known to follow a specific statistical model (e.g., Gaussian or log-normal) and the parameters of the statistical model are known in advance. One attractive advantage of estimators arising from this approach is that their variance asymptotically approaches the Cramér Rao bound, at least for the high signal to noise ratio case. Representative examples include those published by Patwari *et al.* in [PHP+03a], Li in [Li07] and Simonetto *et al.* in [SL14]. However, there are two major challenges with the use of maximum likelihood estimation. First, the optimization problem that arises with such estimation has many local optima and an initialization close to the global optimum is required for good localization performance. Second, the assumption that the underlying statistical model is known in advance is often unrealistic. When measurements deviate from the assumed model, the localization performance degrades with the use of maximum likelihood estimation.

Second, least-squares-estimation-based methods are used when the statistical model associated with the localization measurements is unknown. The positions are estimated by minimizing the sum of the squared distance residuals. A typical cooperative localization method based on least-squares estimation is Algorithm 1 detailed by Wymeersch *et al.* in [WLW09]. These methods are attractive due to their simplicity, since little, or no, statistical knowledge is required in advance. However, similar to maximum likelihood estimators, the localization accuracy of these estimators can be affected by the existence of many local optima. Further, even when some statistical knowledge is *a priori* available, it is difficult to integrate this information into this approach.

Third, in multi-dimensional-scaling (MDS)-based methods, both the agents and anchors are first treated as having unknown positions and their positions are concatenated to build a matrix. The position matrix that minimizes the objective function is then determined. The objective function is commonly a residual error function that

measures the difference between the pairwise dissimilarity and the position related measurements which may be relative ranges, pairwise RSS measurements [CPH06] or even just connectivity information [SRZF04]. The estimated position matrix is then transformed, including rotation, reflection, and translation, so that it is a best fit with the known anchor positions. The advantages and disadvantages of MDS methods are similar to that of least-squares-estimation-based methods.

Fourth, mathematical-programming-based methods avoid the problem associated with local optima by converting the non-convex optimization problem to a convex problem via convex relaxation. Well-known examples include second-order cone programming (SOCP) [Tse07] and semi-definite programming (SDP) [BLWY06]. The advantage of such mathematical-programming-based methods is that initialization is not a critical issue since there are no local optima in convex optimization problems. In addition, there exist several off-the-shelf toolboxes (e.g., SeDuMi [Stu99]) to solve such convex problems. However, as with least-squares-estimation-based methods, statistical knowledge cannot be fully exploited to improve the localization performance.

Fifth, message passing methods treat the positions as random variables with the marginal posterior distribution of each position being calculated or approximated by using message passing methods [IFMW05, WLW09]. Such methods are different from those mentioned above in that a posterior probability density function is provided for each agent location, while only a point estimate is provided by the methods mentioned above. In general, this is advantageous but comes at the cost of significantly increased computational complexity. Additionally, the statistical model of measurement error is required in message passing methods.

1.2 Research Contributions

In this thesis, several advances with respect to cooperative localization methods for wireless networks, operating under practical real-world constraints, are detailed. In particular, mixed LOS/NLOS environments with only partial or even no statistical knowledge about the measurement model are considered. It is assumed that neither NLOS identification nor datasets from experimental campaigns conducted in advance are available. In practice, good initialization is required to avoid local optima in cooperative localization methods and methods that avoid local optima solutions are detailed in [JYF$^+$16, JYF$^+$20b, JYF$^+$20a, JYZS21].

Advances include, first, when severely quantized RSS measurements are utilized it is shown in [JYF$^+$16] that message passing algorithms are effective for localization with

Methods	**Advantages**	**Disadvantages**
Maximum likelihood	- Asymptotic efficiency	- Local optima - Statistical model required
Least-squares	- Simplicity - Statistical information not required	- Local optima - Difficult to integrate statistical information
Multi-dimensional scaling	- Simplicity - Statistical information not required	- Local optima - Difficult to integrate statistical information
Mathematical programming	- Global optima - Statistical information not required	- Difficult to integrate statistical information
Message passing	- Marginal posterior of each position	- Computational complexity - Statistical model required

Table 1.1: Advantages and disadvantages of representative cooperative localization methods.

satisfactory localization accuracy. Two versions of message passing algorithms are developed: a particle-based algorithm and a parametric-based one. It is shown, in a typical scenario, that the latter is computationally more efficient than the former with the trade-off being slightly degraded localization accuracy.

Second, in [JYF+20b], the problem of cooperative localization is generalized to the scenario where RSS measurements are available but the path loss exponent is unknown. Both positions and the path loss exponent are treated as random variables and a Bayesian framework is developed so that jointly estimating positions and the path loss exponent becomes a probabilistic inference problem. For this probabilistic inference problem, message passing methods are developed in which the marginal posterior of each unknown variable is approximately calculated. For mathematical tractability, variable discretization and Monte-Carlo-based numerical approximation mechanisms are combined. To reduce the computational complexity, an auxiliary importance sampler is proposed for the belief update. This auxiliary importance sampler has a complexity order which scales linearly with the number of samples. Moreover, a novel strategy is developed for sampling from a normalized likelihood function, which plays an important role in particle-based message passing methods and mathematically interprets and corrects an existing heuristic sampling strategy.

Third, in [JYF+20a, JYZS21], a more general localization scenario is considered in which statistical knowledge of the LOS/NLOS measurement errors is completely un-

known but range measurements, which are generally more accurate but quite sensitive to NLOS propagation, are available. For this case the bias associated with each range measurement is treated as an unknown parameter and it is shown that the ranging bias are sparse in LOS-dominating environments. By exploiting this sparsity property, the localization problem can be formulated as two generic regularized optimization problems. Further modifications allow these two generic problems to be specified, giving rise to two semi-definite programs which share equivalent optimal solutions under certain conditions [JYZS21]. It is shown that an appropriate regularization parameter plays a crucial role for accurate localization. Owing to the special structure of the bound-constrained regularized SDP, an efficient data-driven method is proposed to determine the regularization parameter.

1.3 Thesis Structure

In Chapter 2, necessary background information is detailed. First, position related measurements as well as sources of error in such measurements are explained. Second, fundamental concepts of inference on graphical models are reviewed, followed by a short introduction to convex optimization. In Chapter 3, the problem of cooperative localization for the case of severely quantized RSS measurements is considered. For this case message passing methods are considered and are shown to provide satisfactory location estimates. In Chapter 4, cooperative localization is considered for the case of RSS measurements when the path loss exponent is unknown. The path loss exponent is treated as a random variable and the associated localization problem is formulated in a Bayesian framework. In Chapter 5, range measurements are considered for cooperative localization in more general mixed LOS/NLOS environments, where the statistical knowledge of the LOS/NLOS measurement error is completely unknown. Conclusions and potential for further research are detailed in Chapter 6.

1.4 Publications

The following publications have been produced during the period of doctoral candidacy.

Internationally Refereed Journal Articles

- F. Yin, C. Fritsche, D. Jin, F. Gustafsson, and A. M. Zoubir, "Cooperative Localization in WSNs Using Gaussian Mixture Modeling: Distributed ECM Algo-

rithms", *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1448–1463, 2015.

- L. Carlino, D. Jin, M. Muma, and A. M. Zoubir, "Robust Distributed Cooperative RSS-Based Localization for Directed Graphs in Mixed LoS/NLoS Environments", *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, pp. 1–20, 2019.

- D. Jin, F. Yin, C. Fritsche, F. Gustafsson, and A. M. Zoubir, "Bayesian Cooperative Localization Using Received Signal Strength With Unknown Path Loss Exponent: Message Passing Approaches", *IEEE Transactions on Signal Processing*, vol. 68, no. 1, pp. 1120–1135, 2020.

- D. Jin, F. Yin, A. M. Zoubir, and H. C. So, "Exploiting Sparsity of Ranging Biases for NLOS Mitigation", *IEEE Transactions on Signal Processing*, vol. 69, pp. 3782–3795, 2021.

Internationally Refereed Conference Papers

- D. Jin, F. Yin, C. Fritsche, F. Gustafsson, and A. M. Zoubir, "Efficient Cooperative Localization Algorithm In LOS/NLOS Environments", In *European Signal Processing Conference (EUSIPCO)*, Nice, France, 2015.

- D. Jin, F. Yin, C. Fritsche, F. Gustafsson, and A. M. Zoubir, "Cooperative Localization Based on Severely Quantized RSS Measurements in Wireless Sensor Network", In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, 2016.

- F. Scheidt, D. Jin, M. Muma, and A. M. Zoubir, "Fast and Accurate Cooperative Localization in Wireless Sensor Networks", In *European Signal Processing Conference (EUSIPCO)*, Budapest, Hungary, 2016.

- D. Jin, B. Boiadjieva, H. Backhaus, M. Fauß, T. Fu, A. Stroh, A. Klein and A. M. Zoubir, "Analysis of Activity States of Local Neuronal Microcircuits in Mouse Brain", In *European Signal Processing Conference (EUSIPCO)*, Florence, Italy, 2018.

- D. Jin, F. Yin, M. Fauß, M. Muma, and A. M. Zoubir, "Exploiting Sparsity for Robust Sensor Network Localization in Mixed LOS/NLOS Environments", In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020.

Other Contributions

- D. Jin, F. Yin, C. Fritsche, F. Gustafsson, and A. M. Zoubir, "Dithering in Quantized-RSS-Based Localization", in *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Cancun, Mexico, 2015.

Chapter 2

Fundamentals

In this chapter, some basic concepts that are relevant for the thesis are introduced. First, different types of position related measurements as well as their advantages, disadvantages and main sources of error are introduced in Section 2.1. Second, necessary background information on inference in graphical models is reviewed in Section 2.2, including graphical models and (loopy) belief propagation. Third, several convex optimization related concepts, including convex sets, conic linear programming and the Karush–Kuhn–Tucker conditions, are reviewed in Section 2.3. It is necessary to point out that only necessary background information that is closely related to the thesis is included in this chapter.

2.1 Position Related Measurements

The nature of the position localization system in a wireless network depends on the type of technology being used, with radio-frequency-based technology being the most common one. Examples for such systems include the global navigation satellite systems of the United States (the Global Positioning System), Russia (Global Navigation Satellite System), China (BeiDou Navigation Satellite System) and the European Union (Galileo System). Non radio-frequency-based systems include camera-based localization systems, infrared-based localization systems and (ultra) sound-based localization systems, to name a few. A thorough overview of different technologies is given in the habilitation thesis [Mau12]. In this thesis, the focus is mainly on radio-frequency-based localization methods.

With high performance radio-frequency-based systems, the position of a wireless node can be directly inferred from the incoming radio signal, e.g. [Wei11]. With such systems, position localization is directly based on the waveform and corresponds to a one stage process. In contrast, with the radio-frequency systems utilized in wireless networks, position localization is associated with two stages. In the first stage, position related metrics are extracted from the radio frequency signal. In the second stage, from the position related metrics, positions are estimated by using different localization algorithms. Consistent with the majority of existing literature, this dissertation focuses on the development of two-stage localization methods with the main focus being on

the second stage. In the following, different position related measurements as well as their sources of error are briefly introduced.

The first category is *propagation time related measurements*, including *time-of-arrival* (TOA), *round-trip propagation time* and *time-difference-of-arrival* (TDOA). These propagation time related measurements are inherently range related measurements, since the relative distance is the product of the propagation speed and the true propagation time along LOS propagation path. TOA measurement measures the difference between the sending time at the transmitter and the receiving time at the receiver. One requirement on the hardware is that local clocks at the transmitter and the receiver must be synchronized. This requirement is unrealistic for many applications, especially for wireless sensor networks with sensors of low cost. One alternative without the synchronization requirement is round-trip propagation time. As indicated by its name, it measures the propagation time in which one signal is propagated from a transmitter to a receiver, replied by the receiver and returned to the same transmitter. Since this round-trip time is calculated based on the same clock, there is no need for synchronization. However, the delay caused by the processing at the receiver is one major source of error for round-trip propagation time measurements. A second alternative avoiding the synchronization requirement is TDOA that measures the difference between the arrival time of the same signal transmitted by the same transmitter and received at two different receivers. Clocks at all receivers are required to be synchronized, but clocks at the transmitters are not required to be synchronized to the receivers. One localization system based on TDOA measurements is the Global Localization System, in which the clocks of all satellites are synchronized. The most widely used method for obtaining the propagation time related measurements is the generalized cross-correlation methods [KC76], also called the maximum likelihood estimator for TOA.

The second class of position related measurements is *received signal strength* (RSS), which is another type of range related measurements. RSS measures the power of received signal, i.e., squared magnitude of signal strength. Compared with the propagation time related measurements, RSS measurements are relatively unpredictable, and the accuracy of range estimates inferred from RSS measurements is comparatively low. However, RSS measurements are relatively inexpensive and easy to acquire, usually even without the need of additional hardware, which makes RSS measurements attractive for network localization.

In free space, received power $P_r(d)$ is related to the distance d through the Friis equation [Rap01]

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2}, \tag{2.1}$$

where P_t is the transmit power, G_t is the antenna gain of the transmitter, G_r is the antenna gain of the receiver, and λ is the wavelength of the transmitted signal. In real world, multipath and shadowing are two major environmental factors that largely affect the measured RSS. The existence of multipath may cause constructive or destructive addition of signals at the receiver, leading to frequency-selective fading. This type of fading can be eliminated by using a wideband method. Apart from the multipath effect, RSS measurements are influenced by the shadowing effect as well. For instance, one signal may be largely attenuated by obstructions (walls, trees, buildings, etc.). The shadowing effect can be modeled as a random effect influencing the RSS measurements.

One typical statistical model characterizing the ensemble mean received power [dBm] at distance d is given by

$$P_r(d) = P_0(d_0) - 10\alpha \log_{10}\left(\frac{d}{d_0}\right) + v, \tag{2.2}$$

where $P_0(d_0)$ is a reference power in dBm at the reference distance d_0, α denotes the path loss exponent, v is a zero-mean Gaussian random variable, and it represents the log-normal shadowing error. Here, [dBm] denotes that the power is in dB milliwatts units. The standard deviation of the received power, is in units of dB, and it may vary between 4 and 12 in practical environments [Rap01]. It is evident that this statistical model indicates a relation between the RSS measurement, $P_r(d)$, and the distance, d. From this perspective, RSS measurements can be seen as range related measurements as well.

Though multipath and shadowing effects impair the accuracy of range estimates, spatial RSS profile caused by the environmental randomness can be exploited for the purpose of localization. The corresponding methods are RSS-profile-based localization techniques, commonly called RSS-based fingerprinting [BP00]. In these techniques, RSS measurements are treated as certain features that characterize each individual position, and a signal strength map can be constructed for a specific area in advance. Once a wireless node collects several RSS measurements, the position of this node can be estimated by fingerprinting the RSS measurements and the RSS map.

The third category is *angle-of-arrival (AOA) measurements*. Complementary to the range related measurements, AOA measurements provide information about the direction to connected wireless nodes. There are two common categories of techniques for measuring AOA, including array processing techniques and phase interferometry [MFA07]. Measuring AOA typically requires an antenna array, which, however, significantly increases the hardware cost. Typical AOA measurement methods include maximum-likelihood-estimation-based algorithms and subspace-based algorithms, such

as multiple signal classification (MUSIC) and estimation of signal parameters by rotational invariance techniques (ESPRIT). For a detailed technical discussion on AOA measurement methods, see [RRW96, Sch86, RK89].

In practice, the wireless nodes are probably placed with unknown orientation, and the directivity of antennas may influence the AOA measurements. Apart from the directivity of antennas, shadowing and multipath reflections are two environmental factors impairing the accuracy of AOA measurements. An accurate AOA measurement requires a direct LOS path between the transmitter and the receiver. A multipath component or the blockage of the LOS path may lead to an AOA estimate of an entirely different angle.

An additional type of position related measurements is connectivity measurements which provide very coarse range related information. Examples of localization systems using connectivity information include [NN03, SRZF04, DpG01].

As briefly mentioned above, all these position related measurements may be affected by various factors. A summary of the major sources of error is given below.

- *Equipment related noise*: Typical equipment related noise includes thermal noise in the electronic circuits of the hardware and quantized readings. For propagation time related measurements, additional source of equipment noise is related to the local clock, including clock offset and clock drift. For RSS measurements, deviation or fluctuation of transmitted power is one further source of equipment noise. As for AOA measurements, antenna directionality is an additional error source.
- *Environmental factors*: Environmental factors include multipath propagation, shadowing and NLOS propagation. These environmental factors are one of the main challenges for accurate localization. Extensive research activities are conducted for the purpose of alleviating the adverse effects caused by the environmental influence.

2.2 Probabilistic Inference in Graphical Models

2.2.1 Graphical Models

In this section, several basic concepts on graphical models will be explained. A graphical model is a graphical representation of a joint probability distribution and it de-

scribes the statistical relationship between a large collection of random variables. A graph $G = (\mathcal{V}, \mathcal{E})$ consists of a set of nodes, whose set is denoted by $\mathcal{V}$, and a set of edges, whose set is denoted by $\mathcal{E} = \big\{(s,t)\big|s,t \in \mathcal{V}\big\}$. Random variables are represented as nodes and statistical dependence is represented by edges. A graph may be *directed*, in which case each edge has a specific direction and the corresponding model is often called a *Bayesian network*, or a graph may be *undirected*, in which case the model is generally called a *Markov random field* (MRF). In the following, we will focus on undirected graphs and introduce several graphical terminologies.

A simple example of an undirected graph is illustrated in Fig. 2.1-(a). The nodes that are connected to one node, e.g., node x_1, are called its *neighbors*. In this example, neighbors of node x_1 include node x_2 and node x_4. A *clique* is a collection of several nodes which satisfy the condition that there is an edge between each pair of nodes . A *maximal clique* is a clique that cannot be a subset of another clique, meaning that the inclusion of any other node will break the clique property of this maximal clique. The undirected graph in Fig. 2.1-(a) has the following three maximal cliques: $\{x_1, x_2, x_4\}$, $\{x_2, x_3, x_4\}$, $\{x_4, x_5\}$.

A fundamental concept in graphical models is the *conditional independence*. For undirected graphs, conditional independence is defined as follows: For three sets of nodes A, B and C, A and B are conditionally independent given set C if and only if C separates A and B. This means that removing all nodes in set C, there exists no path that connects any node from set A to any node from set B [Mur12]. For the undirected graph example in Fig. 2.1-(a), node x_1 and node x_3 are conditionally independent given nodes x_2 and x_4.

Now it comes to the question how is a graphical model related to its associated joint distribution. For an undirected graph, each maximal clique is associated with one potential function. This potential function is a non-negative function whose arguments are the nodes in the maximal clique. The product of all potential functions is then proportional to the joint distribution. A formal statement of this relation is given in *Hammersley-Clifford* theorem below.

Theorem 1 ***(Hammersley-Clifford)**: A positive distribution $p(\mathbf{x})$ satisfies the conditional independence property of an undirected graph if and only if $p(\mathbf{x})$ can be represented as a product of factors, that is*

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{c \in \mathcal{C}l} \psi_c(\mathbf{x}_c), \tag{2.3}$$

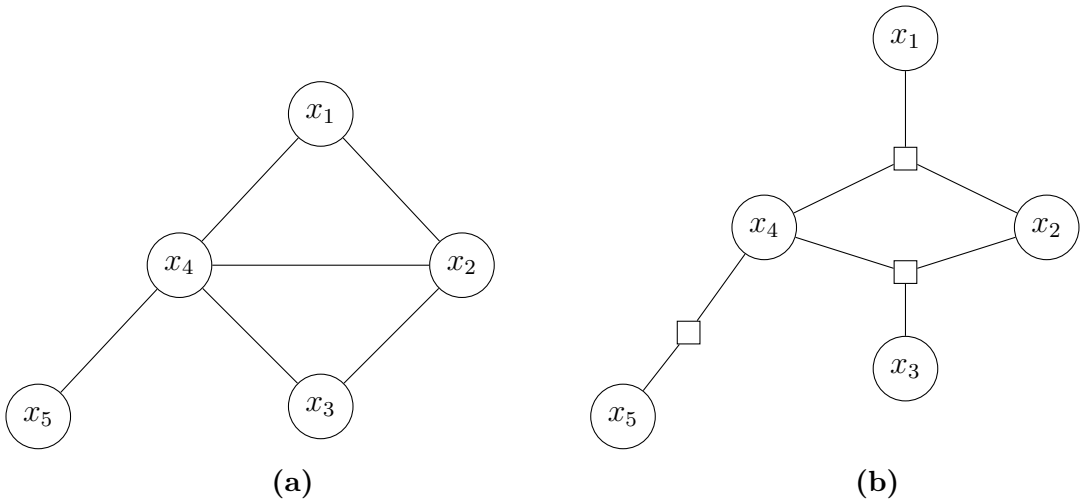


Figure 2.1: A simple example of (a) an undirected graph with the maximal cliques: $\{1, 2, 4\}$, $\{2, 3, 4\}$, $\{4, 5\}$, and (b) its corresponding factor graph.

where $\mathcal{C}l$ is the set of all maximal cliques of the graph, Z is the normalization parameter, $\mathbf{x}_c$ is the collection of all nodes in the maximal clique c, and $\psi_c(\cdot)$ is the potential function that is associated with the maximal clique c.

Based on this theorem, the joint distribution $p(x_1, x_2, x_3, x_4, x_5)$ associated with the undirected graph in Fig. 2.1-(a) must have the form of

$$p(x_1, x_2, x_3, x_4, x_5) = \psi_{1,2,4}(x_1, x_2, x_4)\psi_{2,3,4}(x_2, x_3, x_4)\psi_{4,5}(x_4, x_5). \tag{2.4}$$

Besides general directed and undirected graphs, an additional useful graphical representation are factor graphs. Both directed and undirected graphs can be represented using factor graphs. The advantage of factor graphs is that this graphical representation facilitates the derivation of certain probabilistic inference algorithms. A factor graph is an undirected bipartite graph with two kinds of nodes: variables and factors. Figure 2.1-(b) depicts the factor graph associated with the undirected graph in Fig. 2.1-(a). In this example, variable nodes are represented as circles and factor nodes are represented as squares. A variable node is connected with a factor node using an edge if and only if this variable node is an argument of the factor node. An edge cannot connect two variable nodes or two factors nodes.

2.2.2 Belief Propagation and Loopy Belief Propagation

The focus of this section is to explain an exact statistical inference algorithm, i.e., *belief propagation*, that provides the exact marginal posterior of each variable for general

graphs without loops. Belief propagation is originally proposed by Pearl in [Pea88]. To start, we first consider belief propagation, also called *sum product algorithm*, for a pairwise discrete MRF. Belief propagation for models with cliques of higher orders will be discussed later. For a pairwise MRF, the joint probability function $p(\mathbf{x})$ has the form of

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{s \in \mathcal{V}} \psi_s(x_s) \prod_{(s,t) \in \mathcal{E}} \psi_{s,t}(x_s, x_t), \tag{2.5}$$

where $\psi_s(x_s)$ is the local evidence of node x_s, and $\psi_{s,t}(x_s, x_t)$ is the potential associated with edge (s, t). The key idea of belief propagation is to update a set of messages according to a certain message-updating rule. Each variable (node) has its associated local belief, which depends on a certain collection of messages. After several steps of message updating, the local belief is then the exact marginal posterior of a variable.

To explain the message-updating procedure in belief propagation, consider the example in Fig. 2.2. Each edge is associated with two messages. Taking edge (s, t) as an example, there are messages from node x_s to node x_t, denoted by $m_{s \to t}(x_t)$, and messages from node x_t to node x_s, denoted by $m_{t \to s}(x_s)$. Each node has its associated local belief, denoted by $B_s(x_s)$ for node x_s. First, all messages are initialized to the all-one vector. Next, each node performs the same actions, i.e., belief update and message update, in parallel and in an iterative manner. Taking node x_s at the n-th iteration as an example, belief update and message update are introduced in the following. The local belief $B_s(x_s)$ is updated by collecting messages from all its neighbors according to

$$B_s^n(x_s) \propto \psi_s(x_s) \prod_{t \in \Gamma_s} m_{t \to s}^{n-1}(x_s), \tag{2.6}$$

where Γ_s denotes the set of node x_s's neighbors. Next, node x_s has to update the messages send to each of its neighbors according to

$$m_{s \to t}^n(x_t) = \sum_{x_s} \left(\psi_s(x_s) \psi_{s,t}(x_s, x_t) \prod_{u \in \Gamma_s \backslash t} m_{u \to s}^{n-1}(x_s) \right). \tag{2.7}$$

Here, $\Gamma_s \backslash t$ stands for the collection of node x_s's neighbors, excluding node x_t. In this message-updating step, all messages that node x_s receives, except $m_{t \to s}(x_s)$, contribute to $m_{s \to t}(x_t)$ in the way that the product of these incoming messages are passing through the pairwise potential, $\psi_{s,t}(x_s, x_t)$. For each of its neighbors, node x_s has to calculate one individual message, meaning that altogether node x_s has to update $|\Gamma_s|$ messages.

The belief update and message update are conducted iteratively. For a graph with diameter $D(N)$, after $D(N)$ iterations, $B_s(x_s)$ represents exactly the marginal posterior

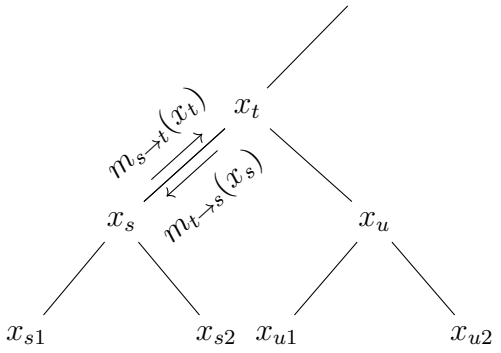


Figure 2.2: Message passing on an undirected pairwise tree.

distribution of x_s. A close inspection of the message-updating step in (2.7) indicates that it has an equivalent reformulation, that is

$$m^n_{s\to t}(x_t) = \sum_{x_s}\left(\psi_{s,t}(x_s, x_t)\frac{B^n_s(x_s)}{m^{n-1}_{t\to s}(x_s)}\right). \tag{2.8}$$

Lastly, we remark that the procedure above is the parallel version of belief propagation. There exists a serial version of belief propagation. For general graphs without loops, both versions provide equivalent results. The message-updating procedure of the parallel version is simpler than that of the serial version, while the serial version is the optimal approach for trees from the perspective of how many times messages are updated. For details on the serial version of belief propagation, see Chapter 20.2.1 in [Mur12].

Algorithm 1 Loopy belief propagation for a pairwise MRF (Algorithm 22.1 in [Mur12])

1: **Input**: node potentials $\psi_s(x_s)$, edge potentials $\psi_{s,t}(x_s, x_t)$
2: **Initialize**: messages $m_{s\to t}(x_t) = 1$ for all edges
3: **Initialize**: beliefs $B_s(x_s) = 1$ for all nodes
4: **Repeat**
5: update belief of each node according to (2.6)
6: calculate message on each edge according to (2.7)
7: **Until** no significant change of beliefs

One extension of belief propagation is loopy belief propagation that is an approximate inference algorithm for general graphs that may contain loops [Mur12]. The basic idea of loopy belief propagation is simple, that is, applying the belief propagation iteratively to the graph. More precisely, for an undirected graphical model with pairwise factors, beliefs and messages are updated iteratively according to (2.6) and (2.7) until convergence, if possible. A summary of loopy belief propagation algorithm is given in

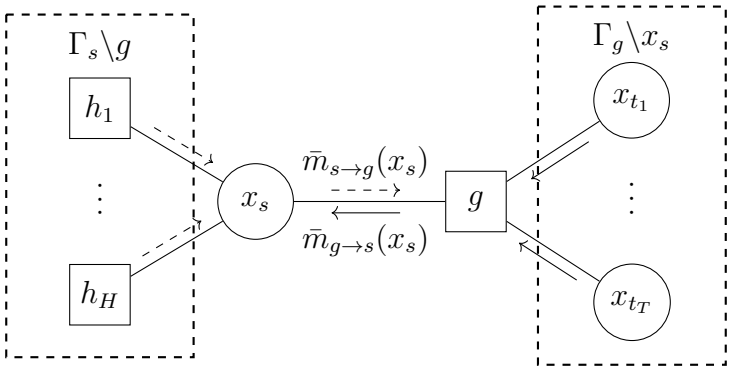


Figure 2.3: An example of a factor graph and belief propagation on it.

Algorithm 1. It should be noted that loopy belief propagation is not guaranteed to converge, and even with convergence local beliefs are not guaranteed to converge to the true marginal posteriors. However, loopy belief propagation can be used as an approximation scheme in loopy networks and it has shown its success in many engineering examples.

For graphs with cliques of high orders, it is convenient to derive belief propagation on a factor graph. There are two kinds of messages on a factor graph: messages from a variable to a factor and messages from a factor to a variable. Taking the factor graph in Fig. 2.3 as an example, the message from the variable node x_s to the factor g is given by

$$\bar{m}^n_{s\to g}(x_s) = \prod_{h\in\Gamma_s\backslash g} \bar{m}^{n-1}_{h\to s}(x_s). \tag{2.9}$$

Here, Γ_s denotes the set of all the factors that are connected to variable node x_s. The message passed from the factor g to the variable node x_s is obtained as

$$\bar{m}^n_{g\to s}(x_s) = \sum_{x_{t_1},\ldots,x_{t_T}} g(x_s, x_{t_1}, \ldots, x_{t_T}) \prod_{t_i\in\Gamma_g\backslash x_s} \bar{m}^n_{t_i\to g}(x_{t_i}). \tag{2.10}$$

Here, Γ_g denotes all the variable nodes that are connected to factor g. Collecting all messages from factors that are connected to variable node x_s, the local belief at x_s is obtained as

$$B^n(x_s) \propto \prod_{g\in\Gamma_s} \bar{m}^n_{g\to s}(x_s). \tag{2.11}$$

Such a message passing procedure iterates until convergence or until a certain termination condition is fulfilled. Upon convergence, the belief $B(x_s)$ is an approximation of the marginal posterior of x_s.

2.3 Convex Optimization

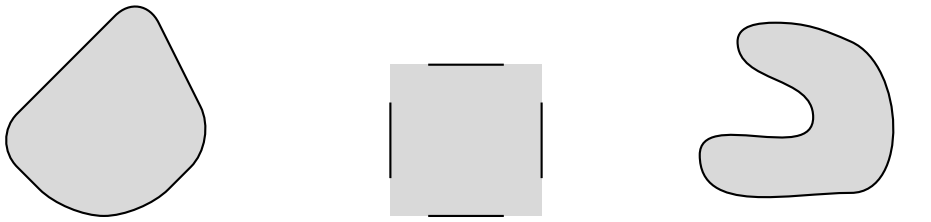
The theory of convex optimization is broad and a comprehensive treatment of all concepts is certainly out of the scope of the thesis. This section covers several key concepts that are relevant for the contributions in Chapter 5. For a comprehensive treatment on convex optimization, see the textbooks [BV04, Ber09, Dat05].

2.3.1 Convex Sets and Convex Cones

A set $\mathcal{C}$ is *convex* if for every x_1 and x_2 in $\mathcal{C}$, the line segment connecting x_1 and x_2 lies in $\mathcal{C}$. Mathematically speaking, this means that $\mathcal{C}$ is convex, if $\forall x_1, x_2 \in \mathcal{C}$ and $\forall l$ in the interval $[0, 1]$, we have

$$(1 - l)\, x_1 + l\, x_2 \in \mathcal{C}. \tag{2.12}$$

Several simple examples of convex and non-convex sets in $\mathbb{R}^2$ are given in Fig. 2.4.



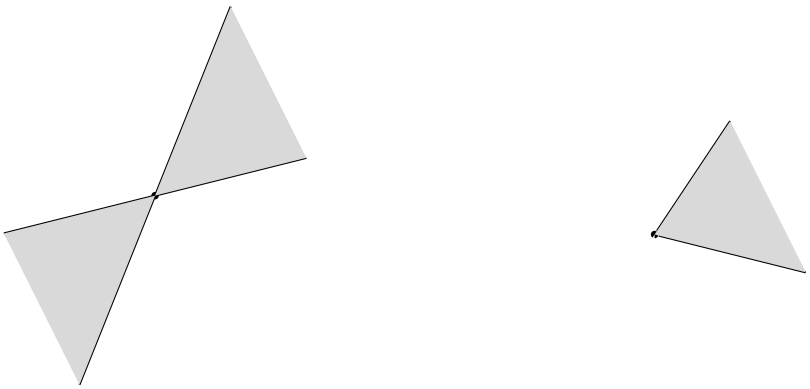
(a) convex **(b)** non-convex **(c)** non-convex

Figure 2.4: Simple examples of convex/non-convex sets in $\mathbb{R}^2$.

A set $\mathcal{K}$ is called a *cone*, if for every $x \in \mathcal{K}$ and every $l \geq 0$, it holds that $l\, x \in \mathcal{K}$. A set $\mathcal{K}$ is a *convex cone*, if it is a cone and convex. Equivalently speaking, $\mathcal{K}$ is a convex cone, if $l_1\, x_1 + l_2\, x_2 \in \mathcal{K}$ for any $x_1,\ x_2 \in \mathcal{K}$ and any $l_1,\ l_2 \geq 0$. Figure 2.5 illustrates two examples of (convex) cones in $\mathbb{R}^2$. Several important and commonly used convex cones are non-negative orthant $\mathbb{R}^n_+$, second-order cone $\mathcal{K} = \left\{[\mathbf{x}; y] \in \mathbb{R}^{n+1} \middle| \|\mathbf{x}\|_2 \leq y\right\}$ and positive semi-definite cone $\mathcal{S}^n_+$, which will be defined below.

The set of all symmetric $n \times n$ matrices is denoted by $\mathcal{S}^n$, i.e., $\mathcal{S}^n = \left\{\mathbf{X} \in \mathbb{R}^{n \times n} \middle| \mathbf{X} = \mathbf{X}^T\right\}$. Based on this, the set of all $n \times n$ symmetric positive semi-definite matrices is denoted by $\mathcal{S}^n_+$ and given by

$$\mathcal{S}^n_+ = \left\{\mathbf{X} \in \mathcal{S}^n \middle| \mathbf{x}^T \mathbf{X} \mathbf{x} \geq 0, \forall \mathbf{x} \in \mathbb{R}^n\right\}. \tag{2.13}$$



(a) a cone but not a convex cone **(b)** a convex cone

Figure 2.5: Simple examples of (convex) cones in $\mathbb{R}^2$.

As stated above, the set of symmetric positive semi-definite matrices $\mathcal{S}_+^n$ is a convex cone. To verify this statement, we have to show that for every $\mathbf{X}_1,\ \mathbf{X}_2 \in \mathcal{S}_+^n$ and every $l_1,\ l_2 \geq 0$, we have $l_1\,\mathbf{X}_1 + l_2\,\mathbf{X}_2 \in \mathcal{S}_+^n$. The symmetry of matrix $l_1\,\mathbf{X}_1 + l_2\,\mathbf{X}_2$ follows straightforwardly from the symmetry of matrices $\mathbf{X}_1$ and $\mathbf{X}_2$. The proof of the positive semi-definiteness can be done straightforwardly based on the definition of positive semi-definiteness [BV04]. More precisely, for any vector $\mathbf{x} \in \mathbb{R}^n$, if $\mathbf{X}_1,\ \mathbf{X}_2 \in \mathcal{S}_+^n$ and $l_1,\ l_2 \geq 0$, it holds that

$$\mathbf{x}^T\big(l_1\,\mathbf{X}_1 + l_2\,\mathbf{X}_2\big)\mathbf{x} = l_1\,\mathbf{x}^T\mathbf{X}_1\mathbf{x} + l_2\,\mathbf{x}^T\mathbf{X}_2\mathbf{x} \geq 0. \tag{2.14}$$

The symbol $\succeq$ is commonly used to represent the positive semi-definiteness, that is $\mathbf{X} \in \mathcal{S}_+^n$ is equivalent to $\mathbf{X} \succeq 0$, and $\mathbf{X}_1 \succeq \mathbf{X}_2$ means $\mathbf{X}_1 - \mathbf{X}_2 \succeq 0$.

Suppose that $\mathcal{K}$ is a cone. The set $\mathcal{K}^*$

$$\mathcal{K}^* = \big\{\mathbf{y}\big|\langle\mathbf{x},\mathbf{y}\rangle \geq 0,\ \forall\mathbf{x} \in \mathcal{K}\big\} \tag{2.15}$$

is called the *dual cone* of $\mathcal{K}$. Here, $\langle\cdot,\cdot\rangle$ denotes the inner product of both variables. A dual cone $\mathcal{K}^*$ must be a convex cone, regardless of whether $\mathcal{K}$ is convex or not. A cone $\mathcal{K}$ is called self-dual, if we have $\mathcal{K}^* = \mathcal{K}$. Two examples of self-dual cones are the non-negative orthant $\mathbb{R}_+^n$ and the positive semi-definite cone $\mathcal{S}_+^n$. For proof of this statement, see Chapter 2 in [BV04]. The concept of a dual cone is useful for the optimality conditions of a conic linear program, which will treated in the following two sections.

2.3.2 Conic Linear Programming

Conic linear programs are a typical class of convex optimization problems. The standard equality form of a conic linear program is given as follows:

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{x} \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{K}, \\ & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{2.16}$$

Here, $\mathcal{K}$ is a closed convex cone, $\mathbf{x} \in \mathbb{R}^n$ is the optimization variable, and $\mathbf{c} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$ are given by the problem. In a conic linear program, both the objective function and the constraint functions are linear. In general, $\mathcal{K}$ can be a product of several closed convex cones, for instance, the Cartesian product of non-negative orthants, second-order cones and positive semi-definite cones.

When $\mathcal{K}$ is $\mathcal{S}_+^n$, the positive semi-definite cone of $n \times n$ matrices, the conic linear program in (2.16) then becomes a *semi-definite program* (SDP):

$$\begin{aligned} \min \quad & \langle \mathbf{C}, \mathbf{X} \rangle \\ \text{subject to} \quad & \mathbf{X} \in \mathcal{S}_+^n, \\ & \langle \mathbf{A}_i, \mathbf{X} \rangle = b_i, \quad i = 1, \ldots, m, \end{aligned} \tag{2.17}$$

where $\mathbf{C}, \mathbf{A}_1, \ldots, \mathbf{A}_m \in \mathcal{S}^n$. Here, matrix $\mathbf{X}$ is the optimization variable and $\mathbf{C}, \mathbf{A}_1, \ldots, \mathbf{A}_m, b_1, \ldots, b_m$ are given by the problem. The representation in (2.17) is called the standard form of an SDP. An alternative form of an SDP is the inequality form, that is

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{x} \\ \text{subject to} \quad & x_1\mathbf{A}_1 + \ldots + x_n\mathbf{A}_n \succeq \mathbf{B}. \end{aligned} \tag{2.18}$$

Here, $\mathbf{x} \in \mathbb{R}^n$ is the optimization variable, and $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{A}_1, \ldots, \mathbf{A}_n$, $\mathbf{B} \in \mathcal{S}_+^n$ are given by the problem.

The standard form of an SDP will be used for the remainder of the thesis, and the inequality form is introduced simply for completeness. We remark that the inner-product $\langle \cdot, \cdot \rangle$ has an equivalent operation, that is

$$\langle \mathbf{C}, \mathbf{X} \rangle = \mathrm{tr}(\mathbf{C}\mathbf{X}^T) = \sum_{j=1}^{n}\sum_{i=1}^{n} C_{ij}X_{ij}, \tag{2.19}$$

where $\mathrm{tr}(\cdot)$ stands for the trace. Adopting this notation, the SDP in (2.17) can be equivalently re-written as

$$\begin{aligned} \min \quad & \mathrm{tr}\big(\mathbf{C}\ \mathbf{X}\big) \\ \text{subject to} \quad & \mathbf{X} \in \mathcal{S}_+^n, \\ & \mathrm{tr}\big(\mathbf{A}_i\ \mathbf{X}\big) = b_i, \quad i = 1, \ldots, m. \end{aligned} \tag{2.20}$$

Note that the transpose symbol $(\cdot)^T$ is neglected due to the symmetry of matrix $\mathbf{X}$.

Interior point methods are the most common generic methods to solve SDPs. Due to the robustness and efficiency of interior point methods in solving general SDPs, most off-the-shelf SDP solvers, e.g., SeDuMi, MOSEK, SDPT3, are based on interior point methods. For a detailed explanation of interior point methods, see Chapter 11 in [BV04]. Interior point methods are in fact second-order methods and need to process a large Hessian matrix in problems of a high dimension. This makes interior point methods become readily computationally prohibitive in high-dimensional problems. From the perspective of computational efficiency, first-order methods are more attractive than interior point methods, but at the cost of accuracy [OCPB16]. Moreover, there exist also other methods that are specially designed for solving certain SDPs, e.g., the augmented Lagrangian method.

2.3.3 Karush–Kuhn–Tucker Optimality Conditions

The *Lagrangian* of the conic linear program in (2.16) is given by

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\nu}, \boldsymbol{\lambda}) = \mathbf{c}^T\mathbf{x} + \boldsymbol{\nu}^T(\mathbf{b} - \mathbf{A}\mathbf{x}) - \boldsymbol{\lambda}^T\mathbf{x}. \tag{2.21}$$

Here, $\boldsymbol{\nu}$ is referred to as the *dual variables*, also known as the *Lagrange multiplier vector*, associated with the equality constraint, and $\boldsymbol{\lambda}$ is the dual variables associated with the conic constraint, $\mathbf{x} \in \mathcal{K}$.

The *Lagrange dual function*, or simply the dual function, $g(\boldsymbol{\nu}, \boldsymbol{\lambda})$ is defined as

$$\begin{aligned} g(\boldsymbol{\nu}, \boldsymbol{\lambda}) &= \inf_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\nu}, \boldsymbol{\lambda}) \\ &= \inf_{\mathbf{x}} \Big(\mathbf{c}^T\mathbf{x} + \boldsymbol{\nu}^T(\mathbf{b} - \mathbf{A}\mathbf{x}) - \boldsymbol{\lambda}^T\mathbf{x} \Big), \end{aligned} \tag{2.22}$$

which is the minimum of the Lagrangian $\mathcal{L}(\mathbf{x}, \boldsymbol{\nu}, \boldsymbol{\lambda})$ over $\mathbf{x}$. Based on the fact that the dual function is the pointwise infimum of a family of affine functions of $\big(\boldsymbol{\nu}, \boldsymbol{\lambda}\big)$, $g(\boldsymbol{\nu}, \boldsymbol{\lambda})$ is always convex. Notably, a dual function is always convex regardless of whether its associated primal problem is convex or not. A primal problem is simply the original problem, i.e., problem in (2.16). Note that, with a slight notation abuse, we denote the Lagrange dual function by $g(\boldsymbol{\nu}, \boldsymbol{\lambda})$ in this chapter and denote a sparsity-inducing function by $g(\cdot)$ in Chapter 5.

Let p^* be the optimal value of the objective function in the primal problem:

$$\begin{aligned} p^* = \min \quad & \mathbf{c}^T\mathbf{x} \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{K}, \\ & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{2.23}$$

For any $\boldsymbol{\lambda} \in \mathcal{K}^*$ and any $\boldsymbol{\nu}$, we have

$$g(\boldsymbol{\nu}, \boldsymbol{\lambda}) \leq p^*. \tag{2.24}$$

This means that the dual function $g(\boldsymbol{\nu}, \boldsymbol{\lambda})$ provides a lower bound on the optimal value p^*. The verification of this property is as follows. Let $\tilde{\mathbf{x}}$ be a feasible point of the primal problem in (2.16). The feasibility tells us that $\mathbf{A}\tilde{\mathbf{x}} = \mathbf{b}$ and $\boldsymbol{\lambda}^T\tilde{\mathbf{x}} \geq 0$, which is due to the definition of dual cone. Based on this fact, we have

$$\boldsymbol{\nu}^T(\mathbf{b} - \mathbf{A}\tilde{\mathbf{x}}) - \boldsymbol{\lambda}^T\tilde{\mathbf{x}} \leq 0. \tag{2.25}$$

Then, the following relation holds:

$$\mathcal{L}(\tilde{\mathbf{x}}, \boldsymbol{\nu}, \boldsymbol{\lambda}) = \mathbf{c}^T\tilde{\mathbf{x}} + \boldsymbol{\nu}^T(\mathbf{b} - \mathbf{A}\tilde{\mathbf{x}}) - \boldsymbol{\lambda}^T\tilde{\mathbf{x}} \leq \mathbf{c}^T\tilde{\mathbf{x}}. \tag{2.26}$$

Moreover, the definition of a dual function indicates

$$g(\boldsymbol{\nu}, \boldsymbol{\lambda}) = \inf_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\nu}, \boldsymbol{\lambda}) \leq \mathcal{L}(\tilde{\mathbf{x}}, \boldsymbol{\nu}, \boldsymbol{\lambda}) \leq \mathbf{c}^T\tilde{\mathbf{x}}. \tag{2.27}$$

The relation $g(\boldsymbol{\nu}, \boldsymbol{\lambda}) \leq \mathbf{c}^T\tilde{\mathbf{x}}$ in (2.27) holds for each feasible point $\tilde{\mathbf{x}}$. Therefore, the inequality $g(\boldsymbol{\nu}, \boldsymbol{\lambda}) \leq p^*$ is true.

Let d^* be defined as

$$d^* = \max_{\boldsymbol{\nu},\, \boldsymbol{\lambda} \in \mathcal{K}^*} g(\boldsymbol{\nu}, \boldsymbol{\lambda}). \tag{2.28}$$

We have

$$d^* \leq p^*, \tag{2.29}$$

which follows from the relation in (2.27). This property is called *weak duality*. It tells us that the duality gap, $p^* - d^*$, is always non-negative. The weak duality always holds, even when the primal problem is non-convex.

If the equality in (2.29) holds, i.e.,

$$d^* = p^*, \tag{2.30}$$

meaning that the duality gap is zero, *strong duality* holds. Different from weak duality, which always holds, strong duality does not hold in general. Strong duality holds only under certain conditions, called *constraint qualifications*. *Slater's condition* is a simple and common constraint qualification for convex problems. Slater's theorem states that strong duality holds when the primal problem is strictly feasible. Notably, Slater's condition can be relaxed to some extent when some of the inequality constraints are

affine functions. The inequality constraints of affine functions do not need to hold with strict inequality.

Based on the concepts above, we will introduce one set of optimality conditions for the conic linear program in (2.16). Assume that strong duality holds, i.e., $p^* = d^*$, and both the primal and dual are attained by one pair of primal and dual points $\left(\mathbf{x}^*, \boldsymbol{\nu}^*, \boldsymbol{\lambda}^*\right)$. Then, we have

$$\begin{aligned} p^* = \mathbf{c}^T\mathbf{x}^* = d^* &= g(\boldsymbol{\nu}^*, \boldsymbol{\lambda}^*) \\ &= \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\nu}^*, \boldsymbol{\lambda}^*) \\ &\leq \mathcal{L}(\mathbf{x}^*, \boldsymbol{\nu}^*, \boldsymbol{\lambda}^*) \\ &= \mathbf{c}^T\mathbf{x}^* + (\boldsymbol{\nu}^*)^T(\mathbf{b} - \mathbf{A}\mathbf{x}^*) - (\boldsymbol{\lambda}^*)^T\mathbf{x}^* \\ &\leq \mathbf{c}^T\mathbf{x}^* = p^*. \end{aligned} \tag{2.31}$$

Since the second term in the fourth line must be zero, it follows that $(\boldsymbol{\lambda}^*)^T\mathbf{x}^* = 0$, which is called *complementary slackness* condition. Moreover, when strong duality holds, the inequality sign in the third line becomes an equality sign, indicating that $\mathbf{x}^*$ minimizes the Lagrangian $\mathcal{L}(\mathbf{x}, \boldsymbol{\nu}^*, \boldsymbol{\lambda}^*)$. This means that the gradient of the Lagrangian with respect to $\mathbf{x}$, i.e. $\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}, \boldsymbol{\nu}^*, \boldsymbol{\lambda}^*)$, is equal to zero at $\mathbf{x}^*$, which is called *stationarity* condition. Then, collecting all conditions together, we have

$$\begin{aligned} &\mathbf{x}^* \in \mathcal{K}, \quad \mathbf{A}\mathbf{x}^* = \mathbf{b} && \text{(primal feasibility)} \\ &\boldsymbol{\lambda}^* \in \mathcal{K}^* && \text{(dual feasibility)} \\ &(\boldsymbol{\lambda}^*)^T\mathbf{x}^* = 0 && \text{(complementary slackness)} \\ &\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}, \boldsymbol{\nu}^*, \boldsymbol{\lambda}^*)\big|_{\mathbf{x}=\mathbf{x}^*} = \mathbf{c} - \mathbf{A}^T\boldsymbol{\nu}^* - \boldsymbol{\lambda}^* = 0 && \text{(stationarity)} \end{aligned} \tag{2.32}$$

The conditions listed in (2.32) are called the *Karush-Kuhn-Tucker* (KKT) conditions [Gha12]. For the conic linear program in (2.16), any primal/dual points $(\mathbf{x}, \boldsymbol{\nu}, \boldsymbol{\lambda})$ that fulfill the KKT conditions in (2.32) must be primal and dual optimal with zero duality gap. This means that for the problem in (2.16), the KKT conditions are sufficient for the points to be primal and dual optimal. Under Slater's condition, one pair of optimal primal/dual points must satisfy the KKT conditions in (2.32), meaning that the KKT conditions in (2.32) are necessary for the points to be primal and dual optimal. Putting these results together, we have the following statement: For a conic linear program that satisfies Slater's condition, the KKT conditions in (2.32) are necessary and sufficient for optimality.

Chapter 3

Quantized-RSS-Based Cooperative Localization via Message Passing

In this chapter, severely quantized RSS measurements are considered for cooperative localization in wireless sensor networks. This is motivated by the fact that low-end sensors can only provide severely quantized RSS readings due to their limited sensor readings, storage and bandwidth. Severely quantized RSS provides only coarse information on the relative range and sensor positions, thereby making the uncertainty information of each position estimate of high significance. The goal is to estimate the marginal posterior of each location, from which a position estimate, as well as its uncertainty information, can be obtained. This is achieved by adjusting message passing methods for this specific localization problem. More specifically, two versions of message passing methods: a particle-based algorithm and a parametric-based one, are developed in this chapter. The latter is computationally more efficient than the former one, but with the trade-off being slightly degraded localization accuracy.

This chapter is mainly based on [JYF+16]. The remainder of this chapter is organized as follows: Section 3.1 formulates the problem of cooperative localization using quantized RSS measurements mathematically. In Section 3.2, the particle-based message passing algorithm and the parametric one are detailed. The proposed algorithms are evaluated in Section 3.3 using real data. Finally, this chapter is summarized in Section 3.4.

3.1 Problem Formulation

We consider a wireless sensor network in a 2-dimensional space. There are two sets of nodes: target nodes (*agents*), whose positions $\mathbf{x}_i = [x_i, y_i]^T$ are unknown, and reference nodes (*anchors*), whose positions are given. Let $S_u = \{1, \ldots, N_u\}$ be the index set of all agents and $S_a = \{N_u + 1, \ldots, N\}$ be the index set of all anchors. The position of node i is modeled stochastically with a prior probability $f_i(\mathbf{x}_i)$ and the positions are assumed to be *a priori* independent, i.e., $f(\mathbf{x}_1, \ldots, \mathbf{x}_N) = \prod_{i=1}^{N} f_i(\mathbf{x}_i)$. Node i can communicate with a subset of sensors, which are called its neighbors and whose index set is denoted by Γ_i. If we have $k \in \Gamma_j$, $k \notin \Gamma_i$ and $j \in \Gamma_i$, then node k is called a 2-hop neighbor of node i.

Using the well-known log-distance path loss propagation model, the continuous-valued RSS, denoted by r_{ij}, coming from node i and received by node j, is given by

$$r_{ij} = \underbrace{A_0 - 10\,\alpha \log_{10}(d_{ij}/d_0)}_{g_{ij}(\mathbf{x}_i,\mathbf{x}_j)} + v_{ij}. \tag{3.1}$$

Here, A_0 denotes the reference power received at a predefined reference distance d_0, α denotes the path loss exponent, $d_{ij} \triangleq \|\mathbf{x}_i - \mathbf{x}_j\|$ is the Euclidean distance between node i and node j, and the measurement error $v_{ij} \sim \mathcal{N}(0, \sigma_{ij}^2)$ accounts for the propagation shadowing effect with $\mathcal{N}(0, \sigma_{ij}^2)$ denoting a Gaussian distribution with zero-mean and variance σ_{ij}^2 It is assumed that the measurement errors are mutually independent and parameters $A_0, d_0, \alpha, \sigma_{ij}$ are known.

The quantized RSS measurement, denoted by z_{ij}, is the output of a quantization operator $Q(\cdot)$ with S levels,

$$z_{ij} = Q(r_{ij}) = \begin{cases} 0 & \text{if } P_0 \le r_{ij} < P_1, \\ 1 & \text{if } P_1 \le r_{ij} < P_2, \\ \vdots & \vdots \\ S-1 & \text{if } P_{S-1} \le r_{ij} < P_S, \end{cases} \tag{3.2}$$

where $P_0, P_1, \ldots, P_S$ are the quantization levels with $P_0 = -\infty$, $P_S = +\infty$. The vector collection of all quantized RSS measurements is denoted by $\mathbf{z}$. In light of (3.1) and (3.2), we have

$$\Pr(z_{ij} = s \,\big|\, \mathbf{x}_i, \mathbf{x}_j) = \Phi\left(\frac{P_{s+1} - g_{ij}(\mathbf{x}_i, \mathbf{x}_j)}{\sigma_{ij}}\right) - \Phi\left(\frac{P_s - g_{ij}(\mathbf{x}_i, \mathbf{x}_j)}{\sigma_{ij}}\right), \tag{3.3}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the normal distribution. Our objective is to estimate the marginal posterior probability density function (pdf) of each agent position, $f(\mathbf{x}_i|\mathbf{z})$, from which the corresponding position $\mathbf{x}_i$ as well as its associated uncertainty can be obtained.

3.2 Message Passing Algorithms for Quantized RSS

To infer the marginal posterior of each position $\mathbf{x}_i$, $i \in S_u$, we start with the joint posterior $f(\mathbf{x}_1, \ldots, \mathbf{x}_N|\mathbf{z})$. Under the assumptions made in the preceding section, it has the form of

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_N|\mathbf{z}) \propto \prod_{i=1}^{N} \left(f(\mathbf{x}_i) \prod_{j \in \Gamma_i} f(z_{ij}|\mathbf{x}_i, \mathbf{x}_j) \right). \tag{3.4}$$

Intuitively, the marginal posterior, e.g., $f(\mathbf{x}_i|\mathbf{z})$, can be calculated as follows:

$$f(\mathbf{x}_i|\mathbf{z}) = \int \cdots \int f(\mathbf{x}_1, \ldots, \mathbf{x}_N|\mathbf{z}) \, \mathrm{d}\mathbf{x}_{1:N\setminus i}, \tag{3.5}$$

where $\mathrm{d}\mathbf{x}_{1:N\setminus i}$ denotes $\mathrm{d}\mathbf{x}_1, \ldots, \mathrm{d}\mathbf{x}_N$ excluding $\mathrm{d}\mathbf{x}_i$. In practice, such a straightforward operation is intractable since, first, there is no analytical solution to this integral in general, and second, this integral is computationally prohibitive for large scale problem. A well-known local message passing method, called *belief propagation* or *sum-product algorithm*, enables approximate marginalization for loopy graphs in an efficient fashion [Pea88]. In this local message passing method, a set of messages is calculated in an iterative manner, and each marginal posterior can be calculated or approximated based on a certain set of messages. More details on belief propagation can be found in Chapter 2.

3.2.1 Related Works

Belief propagation was used for cooperative localization by Ihler *et al.* in [IFMW05] for the first time. An alternative representative work on cooperative localization via message passing is the sum-product algorithm over a wireless network (SPAWN) by Wymeersch *et al.* in [WLW09]. Since then, there are a variety of variants developed for different purposes. To reduce the computational load, parametric representations of local beliefs have been proposed in [LFS$^+$12, CPWG10, OFW11, WL07]. Alternatively, a reduction in computational complexity is achieved in [JYF$^+$15] by, first, representing local beliefs using Gaussian mixture models, and second, adopting the sigma point methods. A further method is given in [SJMZ16], in which fast Fourier transform is utilized to reduce the computational complexity. To increase the localization accuracy in loopy networks, additional variants of the non-parametric belief propagation have been proposed in [SPZG10, SZ13].

The majority of the existing message passing algorithms only consider the case where measurements are relative distances. When other measurements, e.g., RSS, are available, a two-step procedure is commonly utilized for positioning via message passing, e.g., in [SPZG10,SZ13]. These measurements are first converted to estimates of relative distance. Then, positions are estimated using message passing algorithms with these distance estimates. Such a two-step procedure is however not a good solution for our problem. This is because, first, the distance estimates inferred from the severely quantized RSS measurements are very inaccurate, and second, the statistical model of the distance estimates deviates far from a Gaussian distribution, which is an assumption

in the existing message passing algorithms. In this chapter, we focus on developing message passing algorithms that are directly based on the severely quantized RSS measurements. More precisely, we start with the statistical model of the severely quantized RSS measurements and derive the SPAWN-type algorithms specifically for this statistical model.

3.2.2 Message Passing Algorithms

The key idea of belief propagation is to update a set of messages which contribute to calculating the marginal posteriors in an iterative manner. At each iteration, each node performs the same actions, i.e., message update and belief update, in parallel. Taking node i at the $(n+1)$-th iteration as an example, message update and belief update are given by

$$m_{ij}^{n}(\mathbf{x}_i) = \int \Pr(z_{ij}|\mathbf{x}_i, \mathbf{x}_j) B_j^{n}(\mathbf{x}_j)\, \mathrm{d}\mathbf{x}_j, \tag{3.6a}$$

$$B_i^{n+1}(\mathbf{x}_i) \propto f_i(\mathbf{x}_i) \prod_{j \in \Gamma_i} m_{ij}^{n}(\mathbf{x}_i). \tag{3.6b}$$

Here, $m_{ij}^{n}(\mathbf{x}_i)$ denotes the message about position $\mathbf{x}_i$ coming from its neighbor j, and $B_i^{n+1}(\mathbf{x}_i)$ denotes the local belief at node i about its own position $\mathbf{x}_i$. The message, $m_{ij}^{n}(\mathbf{x}_i)$, can be perceived as the belief about node i's position, coming from its neighbor, node j. The local belief, $B_i(\mathbf{x}_i)$, is then updated by collecting the messages from all its neighbors. After a sufficient number of iterations, the belief message, $B_i(\mathbf{x}_i)$, can be regarded as an approximation of the marginal posterior, $f(\mathbf{x}_i|\mathbf{z})$. Notably, the message-updating step in (3.6a) differs slightly from the original belief propagation algorithm in that $m_{ji}^{n-1}(\mathbf{x}_j)$ is included in the calculation of $m_{ij}^{n}(\mathbf{x}_i)$. Such a modification is proposed by Wymeersch *et al.* in [WLW09] and the resulting method is called SPAWN.

3.2.3 Particle-Based Algorithm

The message passing method can be interpreted as finding two sets of functions, $m_{ij}(\mathbf{x}_i)$ and $B_i(\mathbf{x}_i)$ for $j \in \Gamma_i$ and $i = 1, 2, \ldots, N$, that fulfill the message-/belief-updating rule in (3.6). In general, there is no analytical solution for this problem and we must resort to numerical approximation mechanisms. One feasible solution is to represent both the messages and the beliefs using a set of weighted particles. The method corresponding to this strategy is referred to as *particle-based SPAWN* algorithm.

First, we consider how to generate a set of weighted particles for $m_{ij}^n(\mathbf{x}_i)$, denoted by $\left\{\mathbf{x}_{j\to i}^{l,n}, w_{j\to i}^{l,n}\right\}_{l=1}^{L}$, based on the particle-based representation of $B_j^n(\mathbf{x}_j)$, denoted by $\left\{\mathbf{x}_j^{l,n}, w_j^{l,n}\right\}_{l=1}^{L}$. Following the sampling strategy in [IFMW05], $\left\{\mathbf{x}_{j\to i}^{l,n}, w_{j\to i}^{l,n}\right\}_{l=1}^{L}$ can be generated according to

$$\mathbf{x}_{j\to i}^{l,n} = \mathbf{x}_j^{l,n} + d_{ij}^{l,n}\left[\cos\left(\theta^{l,n}\right),\ \sin\left(\theta^{l,n}\right)\right]^T, \tag{3.7}$$

where $\left\{d_{ij}^{l,n}\right\}_{l=1}^{L}$ is a set of distance samples drawn from the likelihood function $\Pr(z_{ij} = s|d_{ij})$, and $\left\{\theta^{l,n}\right\}_{l=1}^{L}$ is a set of angle samples drawn from the uniform distribution $\mathcal{U}[0, 2\pi)$ with boundaries 0 and 2π. For notational brevity, the iteration index n and the subscript ij will be omitted.

The question is how to generate distance samples from $\Pr(z = s|d)$. Sampling distance directly from $\Pr(z = s|d)$ is intractable due to its special shapes, which correspond to the solid lines in Fig. 3.1. To solve this problem, we resort to an importance sampler in which the samples are drawn from a proposal distribution, which is close to the target distribution, $\Pr(z = s|d)$, but is simple to generate samples. Then these samples are assigned certain weights that compensate the difference between the target distribution and the proposal distribution. To find an appropriate proposal distribution, we observe the shapes of $\Pr(z = s|d)$, which is depicted in Fig. 3.1. Depending on the value of s, the shapes of $\Pr(z = s|d)$ can be classified into three cases. Based on this observation, we design the proposal distributions for each case of $\Pr(z = s|d)$ independently. The dashed lines and the solid lines in Fig. 3.1 correspond to the proposal distributions and the target distributions, respectively.

In the first and second cases, where $s = 0$ or $s = S-1$, we construct a proposal distribution using a mixture of a uniform distribution and a triangle distribution. Specifically, the proposal distribution $q(d)$ for the case $s = 0$ is given by

$$q(d) = \frac{1}{1+C_a}\, f_{\mathcal{U}[d_h, d_{\text{thres}})}(d) + \frac{C_a}{1+C_a} f_{\Delta(d_l, d_h, d_l)}(d), \tag{3.8a}$$

$$d_l = d_0 10^{\frac{A_0 - P_1 - 3\sigma}{10\alpha}},\ d_h = d_0 10^{\frac{A_0 - P_1 + 3\sigma}{10\alpha}}, \tag{3.8b}$$

where C_a is a scaling factor that makes $q(d)$ continuous, d_{thres} can be set to the communication range, $f_{\mathcal{U}[d_1,d_2)}(d)$ denotes the pdf of the uniform distribution in the interval $[d_1, d_2)$, and $f_{\Delta(a,b,c)}(d)$ denotes the pdf of a triangle distribution that is defined as

$$f_{\Delta(a,b,c)}(d) = \begin{cases} \frac{2(d-a)}{(c-a)(b-a)} & \text{if } a \le d < c, \\ \frac{2(b-d)}{(b-c)(b-a)} & \text{if } c \le d \le b, \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}$$

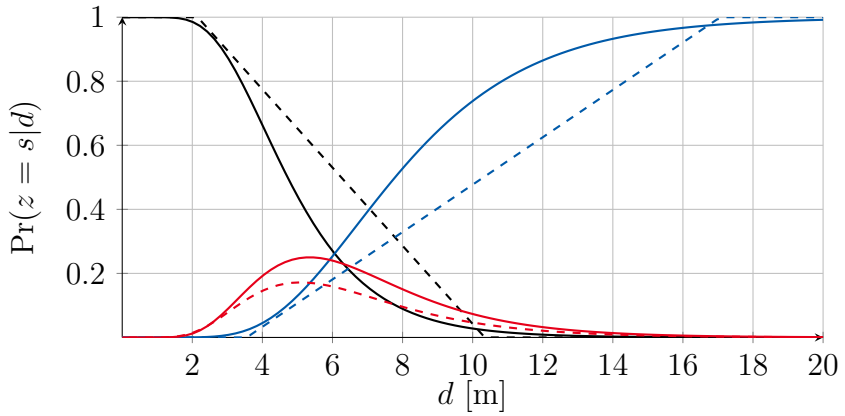


Figure 3.1: The likelihood function $\Pr(z = s|d)$ and the proposal distribution $q(d)$ are indicated by the solid line and dashed lines, respectively. From the left to the right, the black lines correspond to the case $s = S - 1$, the red lines for $s = 1, 2, \ldots, S - 2$, and the blue lines for $s = 0$.

For the case $s = S - 1$, we have

$$q(d) = \frac{1}{1 + C_a} f_{\mathcal{U}[0,d_l)}(d) + \frac{C_a}{1 + C_a} f_{\Delta(d_l,d_h,d_l)}(d), \tag{3.10a}$$

$$d_l = d_0 10^{\frac{A_0 - P_{S-1} - 3\sigma}{10\alpha}}, \; d_h = d_0 10^{\frac{A_0 - P_{S-1} + 3\sigma}{10\alpha}}. \tag{3.10b}$$

Both d_l and d_h are chosen such that $q(d)$ approaches 1 at d_l and 0 at d_h.

For the third case, where $s = 1, 2, \ldots, S - 2$, we use a log-normal distribution as the proposal distribution,

$$q(d) = \frac{C}{d} \exp\left(- \frac{\left(10\,\alpha \log_{10}(\frac{d}{d_0}) - P_0 + P_s\right)^2}{2\sigma^2_{\text{new}}} \right), \tag{3.11a}$$

$$\sigma^2_{\text{new}} = \sigma^2 + \frac{(P_{s+1} - P_s)^2}{3}, \; C = \frac{10\alpha}{\sqrt{2\pi}\sigma_{\text{new}} \log(10)}, \tag{3.11b}$$

where σ_{new} and C are chosen for a good match between the proposal distribution $q(d)$ and the target distribution $\Pr(z = s|d)$.

Up to this point, a distance sample, d^l, can be generated according to (3.8), or (3.10) or (3.11), depending on the value of the quantized RSS measurement. The weight is then simply the ratio of the target distribution and the proposal distribution

$$w_d^l \propto \frac{\Pr(z = s|d^l)}{q(d^l)} \tag{3.12}$$

with $\left\{w_d^l\right\}_{l=1}^L$ summing up to 1.

Eventually, a particle-based representation of the internal message can be generated using (3.7) and the corresponding weights are given by

$$w_{j\to i}^{l} \propto w_j^l w_d^l \tag{3.13}$$

with $\left\{w_{j\to i}^{l}\right\}_{l=1}^{L}$ summing up to one and $\left\{w_j^l\right\}_{l=1}^{L}$ being the weights associated with $\left\{\mathbf{x}_j^l\right\}_{l=1}^{L}$. Based on the weighted particles, the analytical approximation of the internal message is given by

$$m_{ij}(\mathbf{x}_i) \approx \sum_{l=1}^{L} w_{j\to i}^{l} f_{\mathcal{N}}(\mathbf{x}_i; \mathbf{x}_{j\to i}^{l}, \mathbf{\Sigma}_{ij}), \tag{3.14}$$

where $f_{\mathcal{N}}\big(\mathbf{x}_i; \mathbf{x}_{j\to i}^{l}, \mathbf{\Sigma}_{ij}\big)$ stands for the pdf of a Gaussian distribution with mean $\mathbf{x}_{j\to i}^{l}$ and covariance matrix $\mathbf{\Sigma}_{ij}$ which must be chosen appropriately.

Next, it comes to particle-based belief update according to (3.6b). Based on the particle-based representation of messages, a set of weighted particles, $\left\{\mathbf{x}_i^{l,n+1}, w_i^{l,n+1}\right\}_{l=1}^{L}$, should be generated from the belief, $B_i^{n+1}(\mathbf{x}_i)$, according to (3.6b). This can be performed by using the mixture importance sampling strategy given in [IFMW05] and will not be detailed here.

3.2.4 Parametric Algorithm

One major shortcoming of the particle-based algorithm is its high computational load. This is because the manipulation of a large number of weighted particles, including sampling and calculating the associated weights, can be computationally very extensive. To reduce the computational load, we aim for designing an appropriate parametric model to characterize the messages. To reduce the communication load, an appropriate parametric model is designed to represent the local beliefs. Similar to our previous work in [JYF+15], the beliefs are approximated by using a finite mode Gaussian mixture model, i.e.,

$$B_j(\mathbf{x}_j) \approx \sum_{k=1}^{K_j} \phi_j^k f_{\mathcal{N}}\big(\mathbf{x}_j; \boldsymbol{\mu}_j^k, \mathbf{\Sigma}_j^k\big), \tag{3.15}$$

where ϕ_j^k is the component coefficient fulfilling $\sum_{k=1}^{K} \phi_j^k = 1$, and K_j is the number of Gaussian components. Here, K_j can be chosen either empirically or using, for instance, the greedy expectation maximization algorithm which determines an appropriate number of Gaussian components [VVK03]. Inserting (3.15) into (3.6a) gives

$$m_{ij}(\mathbf{x}_i) = \sum_{k=1}^{K_j} \phi_j^k \underbrace{\int \Pr(z_{ij}|\mathbf{x}_i, \mathbf{x}_j) f_{\mathcal{N}}\big(\mathbf{x}_j; \boldsymbol{\mu}_j^k, \mathbf{\Sigma}_j^k\big)\, \mathrm{d}\mathbf{x}_j}_{G_{ij}^k(\mathbf{x}_i)}. \tag{3.16}$$

It is apparent from (3.16) that each integral $G_{ij}^k(\mathbf{x}_i)$ is the convolution of a nonlinear function with a Gaussian density function. We propose to approximate the convolution result, first, by replacing the Gaussian with its mean parameter, and second, by expanding the resulting function appropriately. Consequently, $G_{ij}^k(\mathbf{x}_i)$ is approximated using $\mathcal{G}(\mathbf{x}_i; \boldsymbol{\mu}, \hat{\sigma})$ that is the nonlinear function itself with tunable parameters,

$$\mathcal{G}(\mathbf{x}_i; \boldsymbol{\mu}, \hat{\sigma}) = \Phi\left(\frac{P_{s+1} - g_{ij}(\mathbf{x}_i, \boldsymbol{\mu})}{\hat{\sigma}}\right) - \Phi\left(\frac{P_s - g_{ij}(\mathbf{x}_i, \boldsymbol{\mu})}{\hat{\sigma}}\right), \tag{3.17}$$

where $P_s, P_{s+1}, A_0, \alpha, d_0$ are the parameters of this model, but ignored here for notational brevity. With this model, the convolution result is estimated using

$$G_{ij}^k(\mathbf{x}_i) \approx \mathcal{G}(\mathbf{x}_i; \boldsymbol{\mu}_j^k, \hat{\sigma}_{ij}^k) \tag{3.18}$$

with $\hat{\sigma}_{ij}^k = \sqrt{\sigma_{ij}^2 + \mathrm{Tr}(\boldsymbol{\Sigma}_j^k)}$. Finally, the parametric model of the message, $m_{ij}(\mathbf{x}_i)$, is given by

$$m_{ij}(\mathbf{x}_i) \approx \sum_{k=1}^{K_j} \phi_j^k \, \mathcal{G}(\mathbf{x}_i; \boldsymbol{\mu}_j^k, \hat{\sigma}_{ij}^k). \tag{3.19}$$

For completeness, the SPAWN-type algorithms, both the particle-based algorithm and the parametric one, are summarized in Algorithm 2 for node i at the n-th iteration.

3.3 Numerical Results

The proposed algorithms are evaluated using the real sensor network and the continuous-valued RSS measurements published in [PHP+03a]. The RSS measurements are uniformly quantized with different quantization levels, $S \in \{2, 4, 8\}$, giving rise to quantized RSS measurements. The environmental parameters, $A_0, \alpha, d_0, \sigma^2$, are chosen as reported in [PHP+03a]. The distributed maximum likelihood estimator (MLE), i.e., Algorithm 1 in [WLW09], is used as a competitor. For the distributed MLE, the initial position of each agent is randomly chosen over the space. For the SPAWN-type algorithms, uniform distribution over the space is used for the prior distribution $f_i(\mathbf{x}_i)$ for $i \in S_u$. In the SPAWN-type algorithms, $L = 1000$ particles are used and the maximum number of Gaussian components is 5. The SPAWN-type algorithms find the final position estimates using the minimum mean squared error (MMSE) estimator based on the estimated marginal posteriors.

Algorithm 2 SPAWN-type Algorithms for Quantized RSS

1: Broadcast $B_i^n(\mathbf{x}_i)$
   - $\left\{\mathbf{x}_i^{l,n}, w_i^{l,n}\right\}_{l=1}^{L}$ ← particle-based SPAWN
   - $\left\{\phi_i^{k,n}, \boldsymbol{\mu}_i^{k,n}, \boldsymbol{\Sigma}_i^{k,n}\right\}_{k=1}^{K_i^n}$ ← parametric SPAWN
2: Receive $B_j^n(\mathbf{x}_j)$ from all neighbors, $j \in \Gamma_i$
3: Generate weighted particles $\left\{\mathbf{x}_{j\to i}^{l,n}, w_{j\to i}^{l,n}\right\}_{l=1}^{L}$ from each $m_{ij}^n(\mathbf{x}_i)$, $j \in \Gamma_i$
   - Draw $d_{ji}^{l,n}$ from (3.8) or (3.10) or (3.11)
   - Draw $\theta^{l,n} \sim \mathcal{U}[0, 2\pi)$
   - $\mathbf{x}_{j\to i}^{l,n+1} = \mathbf{x}_j^{l,n} + d_{ji}^{l,n}\,[\cos(\theta^{l,n}), \sin(\theta^{l,n})]^T$
   - Calculate weights $w_{j\to i}^{l,n}$ according to (3.13)
4: Draw $\frac{L}{|\Gamma_i|}$ samples, $\mathbf{x}_i^{l,n+1}$, from each $m_{ij}^n(\mathbf{x}_i)$
5: Calculate weights $w_i^{l,n+1} \propto \frac{\prod_{j\in\Gamma_i} m_{ij}^n(\mathbf{x}_i^{l,n+1})}{\sum_{j\in\Gamma_i} m_{ij}^n(\mathbf{x}_i^{l,n+1})}$
   - $m_{ij}^n(\mathbf{x}_i)$ is defined as (3.14) ← particle-based SPAWN
   - $m_{ij}^n(\mathbf{x}_i)$ is defined as (3.19) ← parametric SPAWN
6: Calculate $\left\{\phi_i^{k,n+1}, \boldsymbol{\mu}_i^{k,n+1}, \boldsymbol{\Sigma}_i^{k,n+1}\right\}_{k=1}^{K_i^{n+1}}$ based on $\left\{\mathbf{x}_i^{l,n+1}, w_i^{l,n+1}\right\}_{l=1}^{L}$ ← only necessary for parametric SPAWN

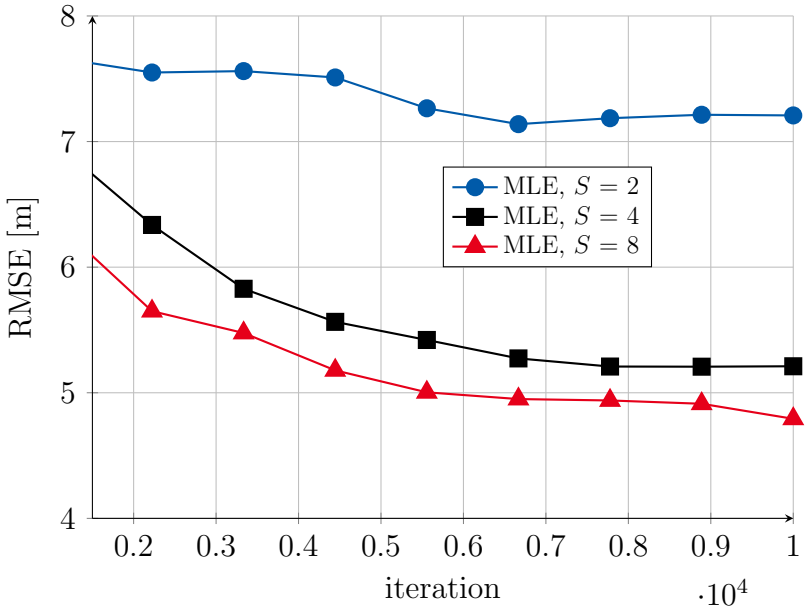


Figure 3.2: RMSE of the distributed MLE.

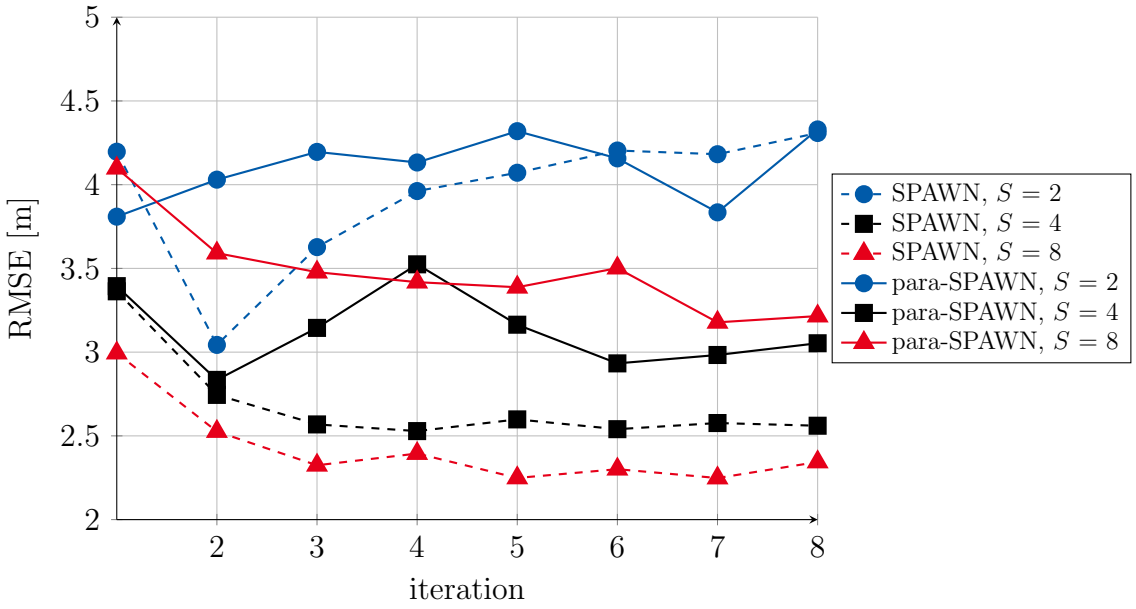


Figure 3.3: The localization accuracy of the SPAWN-type algorithms at different iterations.

The overall root mean squared errors (RMSEs) of the distributed MLE and the SPAWN-type algorithms for different quantization levels are shown in Fig. 3.2 and Fig. 3.3. To keep the range of y axis in both figures as close as possible, the result of the distributed MLE in the first 2000 iterations are omitted. There are several observations from both figures. First, the proposed SPAWN-type algorithms provide higher localization accuracy than the distributed MLE. This result is expected and can be explained as follows. The optimization problem in the distributed MLE contains many local optima and the MLE may easily get stuck in a local optimum. Additionally, only point estimates of positions are communicated between neighbors in the distributed MLE; while estimated marginal posteriors are communicated between neighbors in the proposed SPAWN-type algorithms. The estimated marginal posteriors surely contain more information than the point estimates of positions, leading to the good performance of the SPAWN-type algorithms. Second, the SPAWN-type algorithms converge after only several iterations; while the distributed MLE requires several thousands iterations. In each iteration, the amount of parameters broadcast by each node in the parametric SPAWN is comparable to that in the distributed MLE. Consequently, the communication load of the parametric SPAWN is much lighter than that of the distributed MLE. Third, a comparison between both SPAWN-type algorithms shows that the particle-based algorithm achieves a better localization performance than the parametric one. This is because the parametric models probably cannot approximate the messages and the beliefs as accurate as the particle-based approximations.

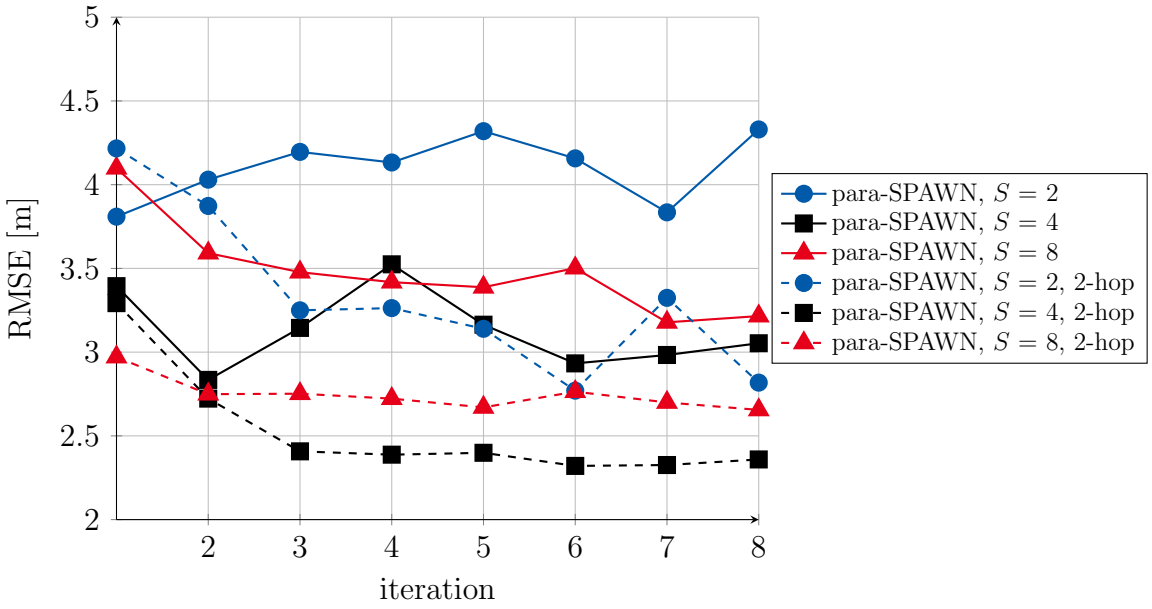


Figure 3.4: The performance of the parametric SPAWN with and without the knowledge about the 2-hop neighbors.

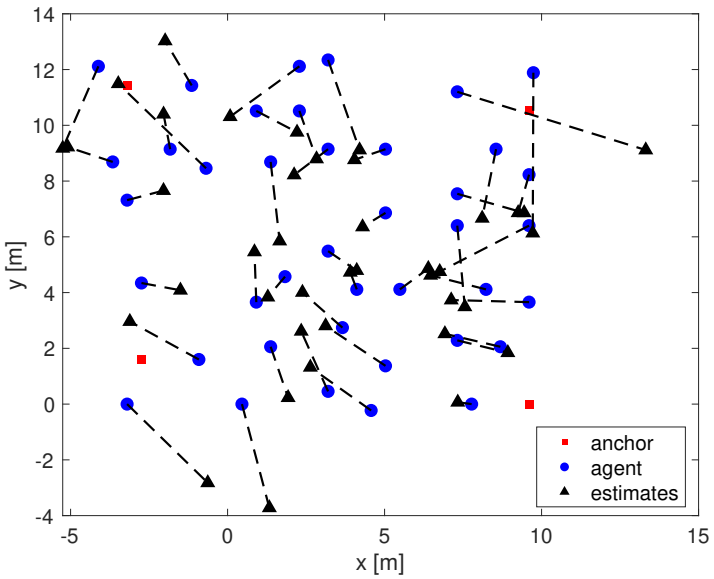


Figure 3.5: Estimated positions using the parametric SPAWN based on proximity measurements and knowledge about 2-hop neighbors.

Knowledge about the 2-hop neighbors can be obtained, but at the cost of a slightly increased communication load. In this simulation, the goal is to investigate the influence of the information from the 2-hop neighbors on the localization performance. Based on the fact that the true distance between a node and its 2-hop neighbor should be

larger than the communication range, the particles of a node's position that are close to its 2-hop neighbors should be punished.

Due to the large simulation time of the particle-based SPAWN, only the parametric SPAWN is considered in this simulation. The RMSEs of the parametric SPAWN with and without the knowledge about the 2-hop neighbors are depicted in Fig. 3.4. It is shown that the additional information from the 2-hop neighbors improves the localization performance of the parametric SPAWN. This result can be explained as follows. Thanks to the additional information from the 2-hop neighbors, more constraints are added to the localization problem and this improves the localization accuracy. The representative position estimates obtained from the parametric SPAWN based on the proximity measurements ($S = 2$) with the knowledge about the 2-hop neighbors are shown in Fig. 3.5.

3.4 Summary

In this chapter, severely quantized RSS measurements were utilized for cooperative localization. Message passing algorithms, including a particle-based algorithm and a parametric one, were developed. Numerical results demonstrated that the message passing algorithms are effective for localization with satisfactory localization accuracy. In particular, it was shown that the parametric SPAWN is computationally more efficient than the particle-based one, but at the cost of slightly degraded localization accuracy. It was further shown that knowledge about 2-hop neighbors has the potential of enhancing the localization accuracy of the parametric SPAWN.

Chapter 4

RSS-Based Cooperative Localization via Message Passing

In Chapter 3, quantized RSS measurements, which are provided by low-end sensors with limited sensor readings, was considered for cooperative localization. Continuous-valued RSS measurements are an additional type of measurements which are commonly available in different types of wireless networks. In this chapter, the problem is generalized to the scenario where RSS measurements are available but one key model parameter (path loss exponent) is unknown. Both positions and the path loss exponent are treated as random variables and a Bayesian framework is developed so that jointly estimating positions and the path loss exponent becomes a probabilistic inference problem. To solve this probabilistic inference problem, message passing methods with appropriate numerical strategies are developed in which the marginal posterior of each unknown variable is approximately calculated.

This chapter is mainly based on [JYF+20b]. The structure of this chapter is as follows: In Section 4.1, the state-of-the-art methods and the challenge are introduced for cooperative localization using RSS measurements. In Section 4.2, the problem of RSS-based cooperative localization with an unknown path loss exponent is formulated mathematically. Section 4.3 is dedicated to developing the message passing algorithms with appropriate numerical strategies. Some important issues are discussed in Section 4.4. The proposed algorithms are evaluated using extensive simulations in Section 4.5. Finally, Section 4.6 summarizes this chapter.

4.1 RSS-Based Cooperative Localization

Many existing works on RSS-based cooperative localization, such as [OWL10, PHP+03a], are based on the assumption that the classical log-distance path loss propagation model is perfectly known. This assumption is oversimplified and impractical for two reasons. First, the knowledge on these model parameters usually relies on a laborious calibration phase, where a large amount of datasets needs to be collected and processed. Such a calibration step is very time consuming and even impossible in many scenarios, such as monitoring and surveillance applications in hostile or inaccessible environments [Li06]. Second, these model parameters, particularly the path loss

exponent, are time varying, due to the changing environment, e.g., weather conditions or human behaviors [MAF07, Rap01]. Without a frequent re-calibration, the resulting mismatch may significantly deteriorate the localization performance. Among these model parameters, the path loss exponent plays the most crucial role for accurate localization. It has been theoretically and algorithmically demonstrated in [SGK12,SKG12] that a slight deviation of path loss exponent may severely deteriorate the localization performance. In this chapter, we focus on RSS-based cooperative localization with an unknown path loss exponent.

The problem of localization with an unknown path loss exponent has been considered by many researchers, but mostly in the context of non-parametric localization. Several representative examples are given in the following. In [Li06], the target position and the path loss exponent are jointly estimated by solving the optimization problem associated with the maximum likelihood estimator using the Levenberg-Marquardt algorithm. In [SKG12, CLIC11], the same optimization problem is first relaxed by linearizing the original problem and then simplified by replacing the position variable with a function of the path loss exponent. By doing so, the cost function depends only on the one-dimensional path loss exponent, and the resulting optimization problem can be readily solved by using grid search. In [Li06, GGL12], the location is estimated by eliminating the nuisance parameter: the path loss exponent (or several other model parameters). In [YFGZ13a], along with several model parameters, the location is estimated using the expectation and maximization criterion. In [SL11,WCLJ12], the location and the path loss exponent are estimated in an alternating manner. More precisely, the position is estimated based on an initialized (or estimated) path loss exponent, and afterwards the path loss exponent is estimated based on the updated position estimate. This procedure iterates until a certain termination condition is met. In the context of cooperative localization, RSS-based localization with an unknown path loss exponent is even more challenging and only very limited works exist. In [WCLJ12, VGBS13, TBD15], the alternating estimation strategy from the non-cooperative case is generalized to the cooperative case. Though such an alternating strategy is straightforward and simple, it is very heuristic and lack of theoretical support.

In this chapter, we treat the path loss exponent as a random variable and formulate the problem in a Bayesian framework. The reasons are as follows. First, when the path loss exponents between different propagation links differ, a random variable characterizing the averaging behavior of the collection of all path loss exponents is more suitable than just one deterministic path loss exponent. Second, characterizing the path loss exponent as a random variable enables us to integrate any prior information, if available, into the parameter estimation. Under the Bayesian umbrella, the cooperative localization problem with an unknown path loss exponent becomes a probabilistic inference

problem. To solve this problem, we derive the message passing algorithms in which the marginalized posterior distribution of each unknown parameter is calculated. For mathematical tractability, we combine the variable discretization and Monte-Carlo-based numerical approximation mechanisms. In addition, to reduce the computational complexity, we propose an auxiliary importance sampler for belief update that has a complexity order which scales linearly with the number of samples. Moreover, we develop a novel strategy for sampling from a normalized likelihood function, which plays an important role in the auxiliary importance sampler and mathematically interprets and corrects an existing heuristic sampling strategy. The proposed sampling strategy will benefit many existing works, such as [IFMW05, WLW09], since this task is an embedded step in many message-passing-based cooperative localization algorithms.

4.2 Problem Formulation

Consider a wireless network in 2-dimensional space with two types of nodes: target nodes (*agents*) with unknown locations and reference nodes (*anchors*) with known locations. Let $\mathbf{x}_i = [x_i, y_i]^T$ denote the location of each node, where $i \in S_u = \{1, \ldots, N_u\}$ for an agent and $i \in S_a = \{N_u + 1, \ldots, N\}$ for an anchor. The index set of all nodes is denoted by S, and we have $S = S_u \bigcup S_a$. Two nodes i and j are neighbors if they can communicate with each other. We denote the index set of node i's neighbors by Γ_i.

Using the well-known log-distance path loss propagation model, the RSS measurement r_{ij}, coming from node i and received by node j, is given by

$$r_{ij} = A_i - 10\,\alpha \log_{10}(d_{ij}/d_0) + v_{ij}, \tag{4.1}$$

where d_0 is a predefined reference distance, A_i denotes the reference power in dBm at d_0 that is assumed to be known, α denotes the *path loss exponent* that is assumed to be an unknown random variable, $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$ is the Euclidean distance, and v_{ij} stands for the log-normal shadowing error that is modeled by $v_{ij} \sim \mathcal{N}(0, \sigma_{ij}^2)$, where $\mathcal{N}(0, \sigma_{ij}^2)$ denotes a zero-mean Gaussian distribution with variance σ_{ij}^2. We assume measurements to be symmetric, i.e., r_{ij} and r_{ji}. The collection of all RSS measurements is denoted by $\mathbf{r} \triangleq \{r_{ij} : (i,j) \in \Gamma\}$, where (i,j) represents that nodes i and j are neighbors, and $\Gamma \triangleq \{(i,j) : j \in \Gamma_i \text{ and } j > i; i \in S_u\}$ denotes the set of all pairs of neighboring nodes. In line with the majority of the existing works, we assume that these shadowing measurement errors v_{ij} for all $(i,j) \in \Gamma$ are independent. The distribution of v_{ij}, denoted by $f_{v_{ij}}(v_{ij})$, is assumed to be known.

It is important to stress that the propagation environment implied by the measurement model in (4.1) is homogeneous with all links having a common path loss exponent, α. Practical localization environments are most likely inhomogeneous with each propagation link, e.g., connection link between node i and node j, having an individual path loss exponent, α_{ij}. However, incorporating such inhomogeneity results in the measurement model being under-determined, since besides the unknown locations, there will be an unknown parameter α_{ij} associated with each measurement r_{ij}. As a compromise between model accuracy and feasibility, a homogeneous propagation environment is considered throughout this chapter. It is noticeable that even for a homogeneous propagation environment, the problem of RSS-based cooperative localization with an unknown path loss exponent is readily challenging and has been rarely studied.

From a Bayesian perspective, we treat the path loss exponent, α, and each position, $\mathbf{x}_i, i \in S$, as random variables, whose prior distributions are denoted by $f(\alpha)$ and $f(\mathbf{x}_i), i \in S$, respectively. All positions and the path loss exponent are assumed to be mutually independent, i.e., $f(\alpha, \mathbf{x}_1, \ldots, \mathbf{x}_N) = f(\alpha) \cdot f(\mathbf{x}_1) \cdots f(\mathbf{x}_N)$. Our purpose is to infer the marginalized posterior distribution (marginal posterior) of each unknown parameter, including $f(\alpha|\mathbf{r})$ or $f(\mathbf{x}_i|\mathbf{r})$, $i \in S_u$, from the measurements $\mathbf{r}$ and the prior information about all parameters.

4.3 Message Passing Algorithms for RSS With Unknown Path Loss Exponent

To infer the marginal posterior of the path loss exponent, α, and that of each position, $\mathbf{x}_i, i \in S_u$, we start with the joint posterior $f(\mathbf{x}_1, \ldots, \mathbf{x}_N, \alpha|\mathbf{r})$. Under the assumptions made in the preceding section, it has the form of

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_N, \alpha|\mathbf{r}) \propto f(\alpha) \prod_{i=1}^{N} \Bigg(f(\mathbf{x}_i) \prod_{j \in \Gamma_i, j>i} f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j, \alpha) \Bigg). \tag{4.2}$$

The marginal posterior, e.g., $f(\mathbf{x}_i|\mathbf{r})$, can be intuitively calculated as follows:

$$f(\mathbf{x}_i|\mathbf{r}) = \int \cdots \int f(\mathbf{x}_1, \ldots, \mathbf{x}_N, \alpha|\mathbf{r}) \, \mathrm{d}\mathbf{x}_{1:N\setminus i} \, \mathrm{d}\alpha. \tag{4.3}$$

Such a straightforward approach is intractable, since no analytical solution exists and a numerical approximation of this integral is computationally prohibitive in practice. Following the same procedure as in Chapter 3, we adopt the message passing method, belief propagation, which solves marginalization problems in a computationally efficient manner.

4.3.1 Message-Updating Rule

In Chapter 4, and in many existing works on cooperative localization via belief propagation, e.g., [SIFW03, IM09], only pairwise potential functions are concerned. For the marginalization problem associated with (4.2), potential functions of order three are concerned. This makes the message-updating rule in Chapter 4 not directly applicable to this problem and, hence, we will derive it explicitly in what follows.

For problem with cliques of high orders, it is convenient to derive the message-updating rule based on a factor graph. The factor graph associated with the joint posterior $f(\mathbf{x}_1, \ldots, \mathbf{x}_N, \alpha|\mathbf{r})$ is depicted in Fig. 4.1. With a slight notation abuse, we use f_{ij} as a short-hand notation for the likelihood function $f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j, \alpha)$ from now on. Here, f_{ij} for all $(i,j) \in \Gamma$ are the potential functions corresponding to the cliques of order three and the path loss exponent variable is connected to all likelihood functions.

The key idea of the belief propagation is to update a set of messages iteratively, which contribute to calculating the marginal posteriors. We denote the message from the factor f_{ij} to the variable α by $m_{f_{ij}\to\alpha}(\alpha)$ and that from f_{ij} to $\mathbf{x}_i$ by $m_{f_{ij}\to\mathbf{x}_i}(\mathbf{x}_i)$. The messages $m_{f_{ij}\to\alpha}(\alpha)$ and $m_{f_{ij}\to\mathbf{x}_i}(\mathbf{x}_i)$ are updated according to the following rules:

$$m^n_{f_{ij}\to\alpha}(\alpha) \propto \iint f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j, \alpha)\, f(\mathbf{x}_i) f(\mathbf{x}_j) \prod_{s\in\Gamma_i\setminus j} m^{n-1}_{f_{si}\to\mathbf{x}_i}(\mathbf{x}_i) \prod_{t\in\Gamma_j\setminus i} m^{n-1}_{f_{tj}\to\mathbf{x}_j}(\mathbf{x}_j)\, \mathrm{d}\mathbf{x}_i\, \mathrm{d}\mathbf{x}_j, \tag{4.4a}$$

$$m^n_{f_{ij}\to\mathbf{x}_i}(\mathbf{x}_i) \propto \iint f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j, \alpha)\, f(\alpha) f(\mathbf{x}_j) \prod_{t\in\Gamma_j\setminus i} m^{n-1}_{f_{tj}\to\mathbf{x}_j}(\mathbf{x}_j) \prod_{(u,z)\in\Gamma\setminus(i,j)} m^{n-1}_{f_{uz}\to\alpha}(\alpha)\, \mathrm{d}\mathbf{x}_j\, \mathrm{d}\alpha. \tag{4.4b}$$

Here, $\Gamma\setminus(i,j)$ denotes the set of all pairs of connections excluding the connection (i,j). To facilitate compact notation, we will simplify $m_{f_{ij}\to\alpha}(\alpha)$ and $m_{f_{ij}\to\mathbf{x}_i}(\mathbf{x}_i)$ to $m_{ij}(\alpha)$ and $m_{ij}(\mathbf{x}_i)$, respectively. An illustrative explanation of (4.4b) is depicted in Fig. 4.1, where the messages enclosed in the blue dashed circle contribute to calculating the message $m_{ij}(\mathbf{x}_i)$. At the first glance, the message-updating rule in (4.4) seems tedious. In the subsequent context, a reformulation of (4.4) will be given in (4.6), thereby facilitating the interpretation of the messages $m_{ij}(\alpha)$ and $m_{ij}(\mathbf{x}_i)$. Based on these messages, the beliefs are updated by performing

$$B^n(\alpha) \propto f(\alpha) \prod_{(i,j)\in\Gamma} m^n_{ij}(\alpha), \tag{4.5a}$$

$$B^n(\mathbf{x}_i) \propto f(\mathbf{x}_i) \prod_{j\in\Gamma_i} m^n_{ij}(\mathbf{x}_i). \tag{4.5b}$$

Herein, $B^n(\alpha)$ and $B^n(\mathbf{x}_i)$ denote the belief of the path loss exponent, α, and the belief of the position variable, $\mathbf{x}_i$, in the n-th iteration, respectively. The belief-updating rule in (4.5), such as $B^n(\mathbf{x}_i)$, can be interpreted as multiplying the messages coming from all factors connected to $\mathbf{x}_i$. As an illustrative example, the belief-updating rule for $B(\mathbf{x}_i)$ is depicted in Fig. 4.1, where the messages enclosed in the red dotted circle contribute to updating $B(\mathbf{x}_i)$.

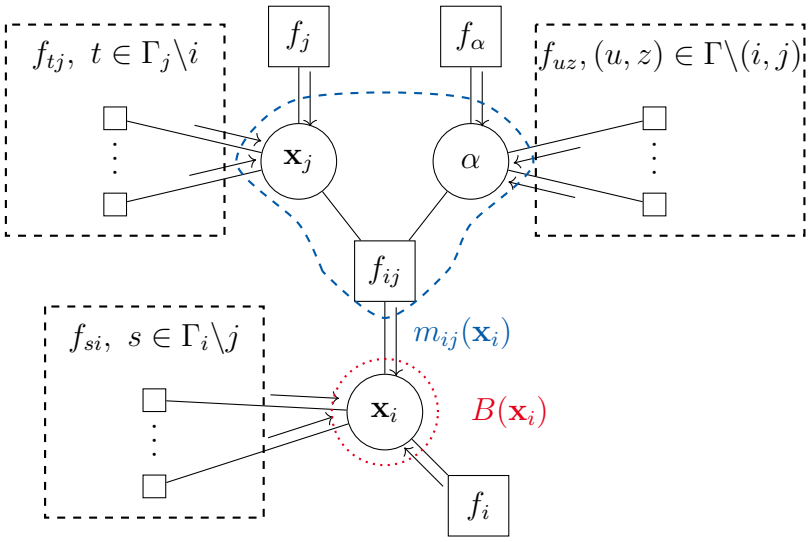


Figure 4.1: An illustration of factor graph and belief propagation associated with the joint posterior in (4.2). For clarity, the overlap between three rectangular blocks are omitted. Here, f_i and f_α are short notations for $f(\mathbf{x}_i)$ and $f(\alpha)$, respectively, and f_{ij} for the likelihood function $f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j, \alpha)$.

Comparing (4.4) with (4.5), it is obvious that certain terms in (4.4) can be replaced by (4.5). By doing so, the message-updating rule in (4.4) can be equivalently rewritten into a simpler form, namely,

$$m_{ij}^n(\alpha) \propto \int\int f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j, \alpha) \frac{B^{n-1}(\mathbf{x}_i)}{m_{ij}^{n-1}(\mathbf{x}_i)} \frac{B^{n-1}(\mathbf{x}_j)}{m_{ij}^{n-1}(\mathbf{x}_j)} \, \mathrm{d}\mathbf{x}_i \, \mathrm{d}\mathbf{x}_j, \tag{4.6a}$$

$$m_{ij}^n(\mathbf{x}_i) \propto \int\int f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j, \alpha) \frac{B^{n-1}(\mathbf{x}_j)}{m_{ij}^{n-1}(\mathbf{x}_j)} \frac{B^{n-1}(\alpha)}{m_{ij}^{n-1}(\alpha)} \, \mathrm{d}\mathbf{x}_j \, \mathrm{d}\alpha. \tag{4.6b}$$

The resulting message-updating rules in (4.6) admit a succinct representation, and the underlying meaning of the messages becomes more clear. Taking $m_{ij}^n(\mathbf{x}_i)$ as an example, it implies that certain information on $\mathbf{x}_i$ can be inferred from the likelihood function $f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j, \alpha)$, given the beliefs of $\mathbf{x}_j$ and α. In other words, the message $m_{ij}^n(\mathbf{x}_i)$ can be perceived as the belief on the position $\mathbf{x}_i$ coming from the node j and the node α. Alternatively, following the principle in [WLW09], the messages can be approximated by ignoring the denominator terms in (4.6), giving rise to the following

message-updating rules:

$$m_{ij}^{n}(\alpha) \propto \int\int f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j, \alpha) B^{n-1}(\mathbf{x}_i) B^{n-1}(\mathbf{x}_j)\, \mathrm{d}\mathbf{x}_i\, \mathrm{d}\mathbf{x}_j, \tag{4.7a}$$

$$m_{ij}^{n}(\mathbf{x}_i) \propto \int\int f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j, \alpha) B^{n-1}(\mathbf{x}_j) B^{n-1}(\alpha)\, \mathrm{d}\mathbf{x}_j\, \mathrm{d}\alpha. \tag{4.7b}$$

In this chapter, the message passing algorithm in light of (4.5) and (4.6) is referred to as the belief propagation; while that in light of (4.5) and (4.7) is referred to as the SPAWN. Note that the difference between the belief propagation and the SPAWN lies in the message-updating rule. As will be shown later in Section 4.4.1, the SPAWN message-updating rule according to (4.7) achieves a significant reduction in computational complexity, compared with the belief propagation.

The gist of the message passing algorithms is to perform two steps iteratively: updating the messages according to (4.6) (or (4.7)) and updating the beliefs according to (4.5). To give an overview, we summarize the resulting framework for inferring the marginal posteriors, $f(\alpha|\mathbf{r})$ and $f(\mathbf{x}_i|\mathbf{r}),\ i \in S_u$, in Algorithm 3. First, we initialize the beliefs, $B^0(\alpha)$ and $B^0(\mathbf{x}_i), i \in S$. Here, one sensible choice for the initial beliefs is their prior distributions. In the n-th iteration, the messages $m_{ij}^{n}(\alpha)$ and $m_{ij}^{n}(\mathbf{x}_i)$ are updated using Algorithm 4 and Algorithm 5, respectively, that will be detailed in Section 4.3.2. Then, the belief of each position, i.e., $B^n(\mathbf{x}_i),\ i \in S_u$, is updated, either using an importance sampler or using Algorithm 6, to be detailed in Section 4.3.3. Finally, the belief $B^n(\alpha)$ is updated, which will be discussed again in Section 4.3.3. These operations iterate until a certain termination condition is met, for instance, when the maximal number of iterations $N_{\max}$ is reached. Different from the existing works, this Bayesian framework treats both α and $\mathbf{x}_i,\ i \in S$, as random variables. It has the advantage that any prior knowledge on α and $\mathbf{x}_i,\ i \in S$, can be integrated. By doing so, $f(\alpha)$ reflecting the prior knowledge for any particular environment and $f(\mathbf{x}_i)$ representing the prior knowledge of any degree can be exploited. For instance, an anchor with imperfect position information can be easily handled in this framework. Moreover, this framework provides a marginal posterior estimate for each unknown parameter, which contains much more information than just one point estimate.

Except two special cases: case with discrete-valued variables and case with jointly Gaussian-distributed continuous-valued variables [SII+10], there is no analytical solution to the message-updating rules. Consequently, in general, the main challenge lies in finding two sets of functions, messages and beliefs, that fulfill the message-updating rules. In our problem, where the variables are continuous-valued but not jointly Gaussian distributed, we must resort to numerical approximation mechanisms. One naive and simple numerical approximation scheme is to define a grid, on which the beliefs and

Algorithm 3 Cooperative Localization Algorithms

Initialization: $B^0(\alpha)$ and $B^0(\mathbf{x}_i)$ for all $i \in S$
for $n = 1 : N_{\max}$
  for each $i \in S_u$
    for each $j \in \Gamma_i$
      if $j > i$, **then** calculate $m_{ij}^n(\alpha)$, see Algorithm 4
      compute $m_{ij}^n(\mathbf{x}_i)$, see Algorithm 5
    end for
    update and broadcast $B^n(\mathbf{x}_i)$, see the importance sampler in Section 4.3.4 or Algorithm 6
  end for
  calculate $B^n(\alpha)$ using (4.14)
end for

the messages are evaluated. This approach has two shortcomings. First, the number of the grid points grows exponentially with the dimensionality of the variable. Second, for a certain fixed granularity, the number of the grid points along one dimension grows linearly with its supported interval. Therefore, this approach is appropriate only when the variable is of low dimensionality and defined on a bounded interval, for instance, α varying in the range of $[1.5, 6]$ [Rap01]. An alternative numerical method is the Monte-Carlo-based numerical approximation methods proposed in [SIFW03,SII$^+$10], in which both the beliefs and the messages are approximated using a set of weighted samples. These samples are generated using certain stochastic methods, for instance, Markov Chain Monte Carlo methods in [SIFW03,SII$^+$10]. Compared with the naive discretization method, this sample-based method is suitable to deal with high-dimensional variables or variables with infinite or relatively large support, such as the position variable $\mathbf{x}_i$, $i \in S_u$.

Based on the analysis above, we will discretize α and sample $\mathbf{x}_i$, $i \in S_u$, using stochastic sampling methods. More specifically, the messages and the belief of α, e.g., $m_{ij}(\alpha)$ and $B(\alpha)$, are evaluated on a set of predefined grid points $\{\alpha_d^r\}_{r=1}^R$; while the messages and the beliefs of positions, e.g., $m_{ij}(\mathbf{x}_i)$ and $B(\mathbf{x}_i)$, are approximated based on weighted samples. In the next two sections, we will detail the approximation mechanisms for message updating and belief updating.

4.3.2 Particle-Based Message Updating

In this section, we consider how to update the messages $m_{ij}(\alpha)$ and $m_{ij}(\mathbf{x}_i)$ approximately. We proceed with the belief propagation message-updating rule, and message

updating using the SPAWN can be derived in analogy with the belief propagation. For the moment, we assume that $\{\mathbf{x}_i^{l,n-1}\}_{l=1}^L$, $\{\mathbf{x}_j^{l,n-1}\}_{l=1}^L$ and $\{B(\alpha_d^r)^{n-1}\}_{r=1}^R$ are available, which are the equally weighted samples of $B^{n-1}(\mathbf{x}_i)$, those of $B^{n-1}(\mathbf{x}_j)$ and the values of $B^{n-1}(\alpha)$ evaluated at $\{\alpha_d^r\}_{r=1}^R$, respectively.

The message $m_{ij}^n(\alpha)$ can be updated by approximating the double integral in (4.6a) using importance sampling [Bis06], giving rise to,

$$m_{ij}^n(\alpha) \propto \sum_{l=1}^{L} w_{ij\to\alpha}^{l,n} f(r_{ij}|\mathbf{x}_i^l, \mathbf{x}_j^l, \alpha), \tag{4.8a}$$

$$w_{ij\to\alpha}^{l,n} \propto \frac{B^{n-1}(\mathbf{x}_i^l) \cdot B^{n-1}(\mathbf{x}_j^l)}{m_{ij}^{n-1}(\mathbf{x}_i^l) \cdot m_{ij}^{n-1}(\mathbf{x}_j^l) \cdot q(\mathbf{x}_i^l, \mathbf{x}_j^l)}. \tag{4.8b}$$

Herein, $\{\mathbf{x}_i^l, \mathbf{x}_j^l\}_{l=1}^L$ are the samples generated from the proposal distribution $q(\mathbf{x}_i, \mathbf{x}_j)$, and $\{w_{ij\to\alpha}^{l,n}\}_{l=1}^L$ are the importance weights that satisfy $\sum_{l=1}^L w_{ij\to\alpha}^{l,n} = 1$. Based on the fact that $\mathbf{x}_i$ and $\mathbf{x}_j$ are decoupled in the non-normalized target distribution $\frac{B^{n-1}(\mathbf{x}_i)}{m_{ij}^{n-1}(\mathbf{x}_i)} \cdot \frac{B^{n-1}(\mathbf{x}_j)}{m_{ij}^{n-1}(\mathbf{x}_j)}$, it is natural to split $\mathbf{x}_i$ and $\mathbf{x}_j$ in the proposal distribution, i.e., $q(\mathbf{x}_i, \mathbf{x}_j) = q(\mathbf{x}_i) \cdot q(\mathbf{x}_j)$. The question that remains to answer is how to choose $q(\mathbf{x}_i)$ and $q(\mathbf{x}_j)$. A sensible choice for $q(\mathbf{x}_i)$ is the belief $B^{n-1}(\mathbf{x}_i)$. The reasons are twofold. First, the belief $B^{n-1}(\mathbf{x}_i)$ is part of $\mathbf{x}_i$'s non-normalized target distribution $B^{n-1}(\mathbf{x}_i)/m_{ij}^{n-1}(\mathbf{x}_i)$, and it is an approximate of $f(\mathbf{x}_i|\mathbf{r})$. Second, the samples from $B^{n-1}(\mathbf{x}_i)$ are available, and no extra effort is needed for sampling. For the same reasons, $B^{n-1}(\mathbf{x}_j)$ is used as the proposal distribution, i.e., $q(\mathbf{x}_j) \propto B^{n-1}(\mathbf{x}_j)$. Consequently, $m_{ij}^n(\alpha)$ can be approximated as follows:

$$m_{ij}^n(\alpha) \propto \sum_{l=1}^{L} w_{ij\to\alpha}^{l,n} f(r_{ij}|\mathbf{x}_i^{l,n-1}, \mathbf{x}_j^{l,n-1}, \alpha), \tag{4.9a}$$

$$w_{ij\to\alpha}^{l,n} \propto \frac{1}{m_{ij}^{n-1}(\mathbf{x}_i^{l,n-1}) \cdot m_{ij}^{n-1}(\mathbf{x}_j^{l,n-1})}. \tag{4.9b}$$

Herein, $\{\mathbf{x}_i^{l,n-1}\}_{l=1}^L$ and $\{\mathbf{x}_j^{l,n-1}\}_{l=1}^L$ are samples from $B^{n-1}(\mathbf{x}_i)$ and $B^{n-1}(\mathbf{x}_j)$, respectively, and the importance weights fulfill $\sum_{l=1}^L w_{ij\to\alpha}^{l,n} = 1$. Since we have defined the grid points $\{\alpha_d^r\}_{r=1}^R$, as the last step, $m_{ij}^n(\alpha)$ should be evaluated at $\{\alpha_d^r\}_{r=1}^R$. As will be seen later in Section 4.3.3, evaluating the messages of α at $\{\alpha_d^r\}_{r=1}^R$ can facilitate updating the belief $B(\alpha)$ significantly.

For $m_{ij}^n(\mathbf{x}_i)$ in (4.6b), we can directly combine the discretization approximation and the importance sampling technique, leading to

$$m_{ij}^n(\mathbf{x}_i) \propto \sum_{r=1}^{R} \sum_{l=1}^{L} w_{ij\to\mathbf{x}_i}^{l,n} \frac{B^{n-1}(\alpha_d^r)}{m_{ij}^{n-1}(\alpha_d^r)} f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^{l,n-1}, \alpha_d^r), \tag{4.10a}$$

$$w_{ij\to\mathbf{x}_i}^{l,n} \propto 1/m_{ij}^{n-1}(\mathbf{x}_j^{l,n-1}), \tag{4.10b}$$

where $\{\mathbf{x}_j^{l,n-1}\}_{l=1}^L$ denote the samples from $B^{n-1}(\mathbf{x}_j)$, and the importance weights fulfill $\sum_{l=1}^L w_{ij\to\mathbf{x}_i}^{l,n} = 1$. In contrast to $m_{ij}^n(\alpha)$ in (4.9a), the double-integral problem for $m_{ij}^n(\mathbf{x}_i)$ becomes a double summation in (4.10a). Compared with $m_{ij}^n(\alpha)$, $m_{ij}^n(\mathbf{x}_i)$ is computationally intensive, incurring that updating the belief $B(\mathbf{x}_i)$ will be computationally intensive as well. In order to approximate $m_{ij}^n(\mathbf{x}_i)$ in a computationally more efficient manner, we treat α similarly as $\mathbf{x}_j$ and perform importance sampling for both α and $\mathbf{x}_j$. In contrast to $\mathbf{x}_j$, α can be drawn only from the set of the grid points $\{\alpha_d^r\}_{r=1}^R$. Subsequently, $m_{ij}^n(\mathbf{x}_i)$ can be approximated as follows:

$$m_{ij}^n(\mathbf{x}_i) \propto \sum_{l=1}^{L} w_{ij\to\mathbf{x}_i}^{l,n} f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^{l,n-1}, \alpha^{l,n-1}), \tag{4.11a}$$

$$w_{ij\to\mathbf{x}_i}^{l,n} \propto \frac{1}{m_{ij}^{n-1}(\mathbf{x}_j^{l,n-1}) \cdot m_{ij}^{n-1}(\alpha^{l,n-1})}. \tag{4.11b}$$

Herein, $\{\mathbf{x}_j^{l,n-1}\}_{l=1}^L$ and $\{\alpha^{l,n-1}\}_{l=1}^L$ denote the samples from $B^{n-1}(\mathbf{x}_j)$ and $B^{n-1}(\alpha)$, respectively, and the importance weights fulfill $\sum_{l=1}^L w_{ij\to\mathbf{x}_i}^{l,n} = 1$. Thank to the additional sampling process, the double summation in (4.10a) is simplified to a single summation in (4.11a).

Next, the message $m_{ij}^n(\mathbf{x}_i)$ is further transformed as follows:

$$m_{ij}^n(\mathbf{x}_i) \propto \sum_{l=1}^{L} \tilde{w}_{ij\to\mathbf{x}_i}^{l,n} \tilde{f}(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^{l,n-1}, \alpha^{l,n-1}), \tag{4.12a}$$

where $\tilde{f}(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^{l,n-1}, \alpha^{l,n-1})$ is the normalized likelihood function, as given by

$$\tilde{f}(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^{l,n-1}, \alpha^{l,n-1}) = Z_{ij}^{-1} f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^{l,n-1}, \alpha^{l,n-1}), \tag{4.12b}$$

$$Z_{ij} = \int f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^{l,n-1}, \alpha^{l,n-1})\, \mathrm{d}\mathbf{x}_i, \tag{4.12c}$$

and the mixture weight $\tilde{w}_{ij\to\mathbf{x}_i}^{l,n}$ is given by

$$\tilde{w}_{ij\to\mathbf{x}_i}^{l,n} \propto Z_{ij} \cdot w_{ij\to\mathbf{x}_i}^{l,n} \tag{4.12d}$$

that satisfies $0 \leq \tilde{w}_{ij\to\mathbf{x}_i}^{l,n} \leq 1$ and $\sum_{l=1}^L \tilde{w}_{ij\to\mathbf{x}_i}^{l,n} = 1$. The integral in (4.12c) can be evaluated analytically with the details given in Appendix A.1. It is noteworthy that (4.12a) differs from (4.11a) in that the mixture component $\tilde{f}(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^l, \alpha^l)$ is a normalized likelihood function of $\mathbf{x}_i$, satisfying the properties of a probability density function (pdf), while $f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^l, \alpha^l)$ in (4.11a) does not. As will be seen later in Section 4.3.3, $m_{ij}^n(\mathbf{x}_i)$ in the form of (4.12a) is more advantageous than that in (4.11a), since it allow us to develop an efficient sampling procedure for updating $B(\mathbf{x}_i)$. Finally, Algorithm 4 and Algorithm 5 summarize the steps for updating the messages, $m_{ij}(\alpha)$ and $m_{ij}(\mathbf{x}_i)$, respectively.

Remark: In the SPAWN, the messages are updated in the same manner as the procedures above. The only difference is that the importance weights in (4.9b)-(4.11b) are replaced by

$$w_{ij\to\alpha}^{l,n} = 1/L, \tag{4.13a}$$
$$w_{ij\to\mathbf{x}_i}^{l,n} = 1/L. \tag{4.13b}$$

Algorithm 4 Message Update of $m_{ij}^n(\alpha)$

1: **Input**: $B^{n-1}(\mathbf{x}_i) \triangleq \{\mathbf{x}_i^{l,n-1}\}_{l=1}^L$
$B^{n-1}(\mathbf{x}_j) \triangleq \{\mathbf{x}_j^{l,n-1}\}_{l=1}^L$
2: **Output**: $\{m_{ij}^n(\alpha_d^r)\}_{r=1}^R$
3: calculate $w_{ij\to\alpha}^{l,n}$ using (4.9b) $\leftarrow$ belief propagation
or (4.13a) $\leftarrow$ SPAWN
4: evaluate $m_{ij}^n(\alpha)$ at $\{\alpha_d^r\}_{r=1}^R$ using (4.9a)

Algorithm 5 Message Update of $m_{ij}^n(\mathbf{x}_i)$

1: **Input**: $B^{n-1}(\alpha) \triangleq \{B^{n-1}(\alpha_d^r)\}_{r=1}^R$
$B^{n-1}(\mathbf{x}_j) \triangleq \{\mathbf{x}_j^{l,n-1}\}_{l=1}^L$
2: **Output**: $\{\tilde{w}_{ij\to\mathbf{x}_i}^{l,n}, \tilde{f}(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^{l,n-1}, \alpha^{l,n-1})\}_{l=1}^L$
3: draw $\alpha^{l,n-1} \sim B^{n-1}(\alpha)$
4: calculate $w_{ij\to\mathbf{x}_i}^{l,n}$ using (4.11b) $\leftarrow$ belief propagation
or (4.13b) $\leftarrow$ SPAWN
5: compute $\tilde{w}_{ij\to\mathbf{x}_i}^{l,n}$ and $\tilde{f}(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^{l,n-1}, \alpha^{l,n-1})$ using (4.12)

4.3.3 Particle-Based Belief Updating

In this section, we will discuss the numerical approximation mechanisms for updating the beliefs: $B(\alpha)$ and $B(\mathbf{x}_i)$, $i \in S_u$. First, we consider how to update the belief $B^n(\alpha)$ according to the update rule in (4.5a). For the reason that $\left\{m_{ij}^n(\alpha_d^r)\right\}_{r=1}^R$ are available for each pair of connection $(i,j) \in \Gamma$, $B^n(\alpha)$ can be readily evaluated at $\{\alpha_d^r\}_{r=1}^R$ according to

$$B^n(\alpha_d^r) \propto f_\alpha(\alpha_d^r) \prod_{(i,j)\in\Gamma} m_{ij}^n(\alpha_d^r). \tag{4.14}$$

Thanks to the discretization, updating $B^n(\alpha)$ can be conducted by simply multiplying $|\Gamma|$ real-valued numbers at the R predefined grid points.

Next, we consider how to update the beliefs of position variables, for instance $B^n(\mathbf{x}_i)$. By combining (4.5b) and (4.12a), we obtain $B^n(\mathbf{x}_i)$ in the form of

$$\begin{aligned} B^n(\mathbf{x}_i) &\propto f(\mathbf{x}_i) \prod_{j \in \Gamma_i} m^n_{f_{ij} \to \mathbf{x}_i}(\mathbf{x}_i) \\ &\propto f(\mathbf{x}_i) \prod_{j \in \Gamma_i} \left(\sum_{l=1}^{L} \tilde{w}^{l,n}_{ij \to \mathbf{x}_i} \tilde{f}(r_{ij} | \mathbf{x}_i, \mathbf{x}_j^{l,n-1}, \alpha^{l,n-1}) \right). \end{aligned} \tag{4.15}$$

Our purpose is to conduct efficient sampling, i.e., $\mathbf{x}_i \sim B^n(\mathbf{x}_i)$. Here, the target distribution, $B^n(\mathbf{x}_i)$, is a product of $|\Gamma_i|$ mixtures, each being a sum of L weighted normalized likelihood functions. Note that the component $\tilde{f}(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^{l,n-1}, \alpha^{l,n-1})$ is in general non-Gaussian. Therefore, updating $B(\mathbf{x}_i)$ boils down to sampling from a product of non-Gaussian mixtures. For notational convenience, we simplify (4.15) to

$$B(\mathbf{x}) \propto f(\mathbf{x}) \prod_{j=1}^{J} M_j(\mathbf{x}) \tag{4.16a}$$

$$M_j(\mathbf{x}) = \sum_{l=1}^{L} \tau_j^l f_j^l(\mathbf{x}). \tag{4.16b}$$

One straightforward sampling strategy is to construct all components explicitly and to sample from them. This is, however, computationally prohibitive, since the product of J mixtures, each containing L components, is itself a mixture of L^J components. Besides, there exist several samplers in the existing works, including the Gibbs sampler [SIFW03] and its related multi-scale sampling strategies in [ISFW03, RW07]. These approaches, however, require a prerequisite that is each $M_j(\mathbf{x})$ must be a Gaussian mixture, and, hence, they are not applicable to our problem. In the following, we will first revisit an existing sampling approach and then propose an alternative sampler, which has a significantly reduced computational complexity.

4.3.4 Importance Sampling as Baseline

First, we consider the technique of importance sampling. The samples and the associated weights are obtained as follows:

$$\mathbf{x}^l \sim q(\mathbf{x}), \tag{4.17a}$$

$$w^l \propto B(\mathbf{x}^l)/q(\mathbf{x}^l), \tag{4.17b}$$

where $q(\mathbf{x})$ is an appropriate proposal distribution, and the importance weight w^l satisfies $\sum_{l=1}^{L} w^l = 1$. Possible choices for $q(\mathbf{x})$ include the prior distribution $f(\mathbf{x})$, an

evenly weighted sum of J mixtures $\sum_{j=1}^{J} J^{-1} M_j(\mathbf{x})$ [IFMW05] and the message with the smallest entropy, e.g., $M_j(\mathbf{x})$, [LFS+12]. The resulting Algorithm 3 with the beliefs updated using the importance sampler in (4.17) is referred to as belief-propagation-IS or SPAWN-IS, for which the messages are updated according to the belief propagation or the SPAWN, respectively. One shortcoming of the importance sampler lies in the high computational load, since computing these L weights $\left\{w^l\right\}_{l=1}^{L}$ according to (4.17) requires operations of order $\mathcal{O}(JL^2)$ [LFS+12]. In order to reduce the computational load, we propose an alternative sampler in what follows.

4.3.5 Auxiliary Importance Sampler

Motivated by [BDS05], we develop an efficient sampler, named as *auxiliary importance sampler* (AIS), for the sampling problem $\mathbf{x} \sim B(\mathbf{x})$. The key idea is to introduce an auxiliary variable ψ_j to each mixture $M_j(\mathbf{x})$. The auxiliary variable ψ_j plays the role of a component label indicator, indicating which component is drawn from the mixture $M_j(\mathbf{x}) = \sum_{l=1}^{L} \tau_j^l f_j^l(\mathbf{x})$, and it can take value $\psi_j = \kappa$, where $\kappa \in \{1, \ldots, L\}$. For instance, if we have $\psi_j = \kappa$, it denotes that the κ-th component $\tau_j^\kappa f_j^\kappa(\mathbf{x})$ is drawn from the mixture $M_j(\mathbf{x}) = \sum_{l=1}^{L} \tau_j^l f_j^l(\mathbf{x})$. Stacking all J auxiliary variables into a vector, we have the compact auxiliary variable $\boldsymbol{\psi} = [\psi_1, \ldots, \psi_J]^T$.

With the help of the auxiliary variable $\boldsymbol{\psi}$, the sampling task $\mathbf{x} \sim B(\mathbf{x})$ can be achieved in two steps:

1. Draw $\boldsymbol{\psi}^l \sim p(\boldsymbol{\psi})$,
$$p(\boldsymbol{\psi}) = \int f(\mathbf{x}, \boldsymbol{\psi}) \, \mathrm{d}\mathbf{x} = Z_1^{-1} \int \prod_{j=1}^{J} \tau_j^{\psi_j} f_j^{\psi_j}(\mathbf{x}) \, \mathrm{d}\mathbf{x}; \tag{4.18}$$

2. Draw $\mathbf{x}^l \sim f(\mathbf{x}|\boldsymbol{\psi}^l)$, conditional on $\boldsymbol{\psi}^l$,
$$f(\mathbf{x}|\boldsymbol{\psi}^l) = Z_2^{-1} \prod_{j=1}^{J} f_j^{\psi_j^l}(\mathbf{x}). \tag{4.19}$$

Here, Z_1 and Z_2 are two normalization constants. The samples $\{\mathbf{x}^l\}_{l=1}^{L}$ generated in such a two-step procedure follow the distribution in (4.16). However, when directly sampling from $p(\boldsymbol{\psi})$ and $f(\mathbf{x}|\boldsymbol{\psi}^l)$ is impossible, as in our case, we can generate samples from two proposal distributions $q(\boldsymbol{\psi})$ and $q(\mathbf{x}|\boldsymbol{\psi}^l)$ and assign certain importance weights to them. This gives rise to the following three-step procedure:

1. Draw $\boldsymbol{\psi}^l \sim q(\boldsymbol{\psi})$;
2. Draw $\mathbf{x}^l \sim q(\mathbf{x}|\boldsymbol{\psi}^l)$, conditional on $\boldsymbol{\psi}^l$;
3. Calculate the importance weight w^l

$$w^l \propto \frac{f(\mathbf{x}^l,\ \boldsymbol{\psi}^l)}{q(\mathbf{x}^l, \boldsymbol{\psi}^l)} = \frac{f(\mathbf{x}^l,\ \boldsymbol{\psi}^l)}{q(\boldsymbol{\psi}^l) \cdot q(\mathbf{x}^l|\boldsymbol{\psi}^l)} \tag{4.20}$$

with the non-normalized joint distribution $f(\mathbf{x}^l, \boldsymbol{\psi}^l)$ given by

$$f(\mathbf{x}^l, \boldsymbol{\psi}^l) = f(\mathbf{x}^l) \prod_{j=1}^{J} \tau_j^{\psi_j^l} f_j^{\psi_j^l}(\mathbf{x}^l). \tag{4.21}$$

Up to this point, the problem remained is how to design $q(\boldsymbol{\psi})$ and $q(\mathbf{x}|\boldsymbol{\psi}^l)$, which will be addressed in the following.

Remark: Note that the underlying condition in the AIS is that the target distribution $B(\mathbf{x})$ must be a product of several mixtures, each being a sum of multiple weighted pdfs. Thanks to the additional message transformation in (4.12), the belief in (4.15) satisfies the properties of this condition, meaning that the message transformation in (4.12) is a prerequisite for the development of the AIS.

4.3.5.1 Auxiliary Variable

First, we focus on designing an appropriate proposal distribution $q(\boldsymbol{\psi})$. Ideally, $q(\boldsymbol{\psi})$ should resemble the corresponding target distribution $p(\boldsymbol{\psi})$ as closely as possible, and, at the same time, it should be feasible to draw samples from it. To this end, we first recover the original form of the target distribution $p(\boldsymbol{\psi})$. Replacing $\mathbf{x}$, $\tau_j^{\psi_j}$ and $f_j^{\psi_j}(\mathbf{x})$ in (4.18) with $\mathbf{x}_i$, $\tilde{w}_{ij\to\mathbf{x}_i}^{\psi_j}$ and $\tilde{f}(r_{ij}\big|\mathbf{x}_i, \mathbf{x}_j^{\psi_j}, \alpha^{\psi_j})$, respectively, gives rise to

$$p(\boldsymbol{\psi}) \propto \int \prod_{j\in\Gamma_i} \tilde{w}_{ij\to\mathbf{x}_i}^{\psi_j} \tilde{f}(r_{ij}\big|\mathbf{x}_i, \mathbf{x}_j^{\psi_j}, \alpha^{\psi_j})\ \mathrm{d}\mathbf{x}_i. \tag{4.22}$$

To ensure mathematical tractability, we assume that all auxiliary variables in $\{\psi_j : j \in \Gamma_i\}$ are independent, giving rise to $q(\boldsymbol{\psi}) = \prod_{j=\Gamma_i} q(\psi_j)$ with

$$q(\psi_j = \kappa) = \tilde{w}_{ij\to\mathbf{x}_i}^{\kappa} \int \tilde{f}(r_{ij}\big|\mathbf{x}_i, \mathbf{x}_j^{\kappa}, \alpha^{\kappa})\ \mathrm{d}\mathbf{x}_i = \tilde{w}_{ij\to\mathbf{x}_i}^{\kappa}, \tag{4.23}$$

where the second equality is due to (4.12b) and (4.12c).

4.3.5.2 Position Variable x

In order to design $q(\mathbf{x}|\boldsymbol{\psi}^l)$, again, we recover the original form of $f(\mathbf{x}|\boldsymbol{\psi}^l)$. This is done by replacing $\mathbf{x}$ and $f_j^{\psi_j^l}(\mathbf{x})$ in (4.19) with $\mathbf{x}_i$ and $\tilde{f}(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^{\psi_j^l}, \alpha^{\psi_j^l})$, respectively, giving rise to

$$f(\mathbf{x}_i|\boldsymbol{\psi}^l) \propto \prod_{j\in\Gamma_i} \tilde{f}(r_{ij}\big|\mathbf{x}_i, \mathbf{x}_j^{\psi_j^l}, \alpha^{\psi_j^l}). \tag{4.24}$$

To capture each mixture component in (4.24), we design $q(\mathbf{x}_i|\boldsymbol{\psi}^l)$ in the form of

$$q(\mathbf{x}_i|\boldsymbol{\psi}^l) = \sum_{j\in\Gamma_i} |\Gamma_i|^{-1} q(\mathbf{x}_i\big|\mathbf{x}_j^{\psi_j^l}, \alpha^{\psi_j^l}, r_{ij}), \tag{4.25}$$

where $q(\mathbf{x}_i|\mathbf{x}_j^{\psi_j^l}, \alpha^{\psi_j^l}, r_{ij})$ should resemble $\tilde{f}(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^{\psi_j^l}, \alpha^{\psi_j^l})$ as closely as possible, and, at the same time, drawing samples from it remains feasible. For notational clarity, we will replace ψ_j^l with l', thereby simplifying $q(\mathbf{x}_i|\mathbf{x}_j^{\psi_j^l}, \alpha^{\psi_j^l}, r_{ij})$ to $q(\mathbf{x}_i|\mathbf{x}_j^{l'}, \alpha^{l'}, r_{ij})$.

Next, we proceed with designing the proposal distribution $q(\mathbf{x}_i|\mathbf{x}_j^{l'}, \alpha^{l'}, r_{ij})$ for the target distribution $\tilde{f}(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^{l'}, \alpha^{l'})$, which is the normalized likelihood function $f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^{l'}, \alpha^{l'})/\int f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^{l'}, \alpha^{l'})\mathrm{d}\mathbf{x}_i$. This task is actually an embedded step in many other works, for instance, under different measurement models in [IFMW05, WLW09, LFS+12, JYF+15, JYF+16]. Therefore, instead of being specific, we generalize this sampling problem to a generic measurement model, as given by

$$r_{ij} = h(d_{ij}) + v, \quad v \sim f_v(v). \tag{4.26}$$

Here r_{ij} denotes any distance related measurement, $h(d_{ij})$ is a function of the distance $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$, and v is an additive measurement error. Our purpose is to sample from the normalized likelihood function, namely,

$$\mathbf{x}_i^l \sim Z^{-1} f(r_{ij}\big|\mathbf{x}_i, \mathbf{x}_j^{l'}), \tag{4.27}$$

where $\mathbf{x}_j^{l'}$ is a reference position, and Z is a normalization constant, to be precise, $Z = \int f(r_{ij}\big|\mathbf{x}_i, \mathbf{x}_j^{l'})\ \mathrm{d}\mathbf{x}_i$. The proposal distribution $q(\mathbf{x}_i\big|\mathbf{x}_j^{l'}, r_{ij})$ for the sampling problem in (4.27) can be designed in a bottom-up manner, meaning that we first develop a sampling strategy and then derive the associated distribution $q(\mathbf{x}_i\big|\mathbf{x}_j^{l'}, r_{ij})$. Given r_{ij}, $\mathbf{x}_j^{l'}$ and the measurement model in (4.26), an intuitive and reasonable approach to generate $\mathbf{x}_i^l$ is as follows:

$$\theta_{ij}^l \sim \mathcal{U}[0, 2\pi), \tag{4.28a}$$

$$v^l \sim f_v(v), \tag{4.28b}$$

$$d_{ij}^l = h^{-1}\big(r_{ij} - v^l\big), \tag{4.28c}$$

$$\mathbf{x}_i^l = \mathbf{x}_j^{l'} + \big[d_{ij}^l \cdot \cos\theta_{ij}^l,\ d_{ij}^l \cdot \sin\theta_{ij}^l\big]^T. \tag{4.28d}$$

In words, the sample $\mathbf{x}_i^l$ is obtained by moving $\mathbf{x}_j^{l'}$ in a random direction θ_{ij}^l by a random distance d_{ij}^l, which is generated based on the measurement model and the measurement r_{ij}. We denote the distributions of θ_{ij}, d_{ij} and $\mathbf{x}_i$ associated with the sampling procedures in (4.28a), (4.28c) and (4.28d) by $q_\theta(\theta_{ij})$, $q_d(d_{ij}|r_{ij})$ and $q(\mathbf{x}_i|\mathbf{x}_j^{l'}, r_{ij})$, respectively. Note that the subscripts θ and d in $q_\theta(\theta_{ij})$ and $q_d(d_{ij}|r_{ij})$ indicate that they are the distributions of θ_{ij} and d_{ij}, respectively. However, it is not straightforward to obtain the proposal distribution $q(\mathbf{x}_i\big|\mathbf{x}_j^{l'}, r_{ij})$.

As one of our contributions, we provide a mathematical interpretation and justification for the sampling procedure in (4.28), upon which, we further derive the proposal distribution $q(\mathbf{x}_i\big|\mathbf{x}_j^{l'}, r_{ij})$. The underlying idea of the sampling procedure in (4.28) is the transformation between a pair of random variables, from polar coordinate $[d_{ij}, \theta_{ij}]^T$ to Cartesian coordinate $\mathbf{x}_i$. Equivalently speaking, drawing the position sample $\mathbf{x}_i^l$ is transformed to a problem of drawing the sample pair of distance and angle, i.e., $\left[d_{ij}^l, \theta_{ij}^l\right]^T$. Consequently, the distributions $q(\mathbf{x}_i\big|\mathbf{x}_j^{l'}, r_{ij})$ and $q_{d,\theta}(d_{ij}, \theta_{ij}|r_{ij}) = q_d(d_{ij}|r_{ij}) \cdot q_\theta(\theta_{ij})$ are related according to

$$q(\mathbf{x}_i|\mathbf{x}_j^{l'}, r_{ij}) = \frac{q_d\big(d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j^{l'}\|\big|r_{ij}\big)}{2\pi \cdot \|\mathbf{x}_i - \mathbf{x}_j^{l'}\|}. \tag{4.29}$$

Thanks to (4.29), deriving $q(\mathbf{x}_i\big|\mathbf{x}_j^{l'}, r_{ij})$ reverts to the problem of deriving $q_d(d_{ij}|r_{ij})$, which should not be difficult for most measurement models. In our problem, where the measurement model is defined in (4.1), $q_d(d_{ij}|r_{ij})$ has the form of

$$q_d(d_{ij}|r_{ij}) = \frac{1}{\sqrt{2\pi}\frac{d_{ij}}{d_0}\tilde{\sigma}} \exp\left(-\frac{1}{2\tilde{\sigma}^2}\left(\log\frac{d_{ij}}{d_0} - \tilde{\mu}\right)^2\right), \tag{4.30a}$$

$$\tilde{\mu} = \frac{\log 10}{10\alpha^{l'}} \cdot (A - r_{ij}), \tag{4.30b}$$

$$\tilde{\sigma}^2 = \left(\frac{\log 10}{10\alpha^{l'}}\right)^2 \cdot \sigma^2. \tag{4.30c}$$

Replacing $q_d(d_{ij}|r_{ij})$ in (4.29) with (4.30) gives rise to the proposal distribution $q(\mathbf{x}_i|\mathbf{x}_j^{l'}, r_{ij})$, which is equivalent to $q(\mathbf{x}_i\big|\mathbf{x}_j^{l'}, \alpha^{l'}, r_{ij})$ in our original problem. The derivation of (4.30) is given in Appendix A.2.

4.3.5.3 Importance Weight w_i^l

For the auxiliary variable sample $\boldsymbol{\psi}^l$ and the position sample $\mathbf{x}_i^l$, which are generated from $q(\boldsymbol{\psi})$ and $q(\mathbf{x}_i|\boldsymbol{\psi}^l)$, respectively, the associated importance weight w_i^l is given by

$$w_i^l \propto \frac{\prod_{j\in\Gamma_i} f(r_{ij}\big|\mathbf{x}_i^l, \mathbf{x}_j^{\psi_j^l}, \alpha^{\psi_j^l})}{\sum_{j\in\Gamma_i} |\Gamma_i|^{-1} q(\mathbf{x}_i^l\big|\mathbf{x}_j^{\psi_j^l}, \alpha^{\psi_j^l}, r_{ij})}. \tag{4.31}$$

Finally, Algorithm 6 lists the steps for updating $B(\mathbf{x}_i)$ using the proposed AIS. The resulting Algorithm 3 with the beliefs updated using Algorithm 6 are named as belief-propagation-AIS or SPAWN-AIS, for which the messages are updated according to the belief propagation or the SPAWN, respectively.

Algorithm 6 Belief Update Using AIS

1: **Input**: $m_{ij}^n(\mathbf{x}_i)$ for all $j \in \Gamma_i$
2: **Output**: $B^n(\mathbf{x}_i) \triangleq \{\mathbf{x}_i^{l,n}\}_{l=1}^L$
3: draw $\boldsymbol{\psi}^l \sim q(\boldsymbol{\psi})$ as follows:
4: **for** each $j \in \Gamma_i$
5: draw $\psi_j^l \sim q(\psi_j)$
6: **end for**
7: draw $\mathbf{x}_i^l \sim q(\mathbf{x}_i \big| \boldsymbol{\psi}^l)$ as follows:
8: **for** each $j \in \Gamma_i$
9: draw $\mathbf{x}_i^l \sim q(\mathbf{x}_i | \mathbf{x}_j^{\psi_j^l}, \alpha^{\psi_j^l}, r_{ij})$ using (4.28)
10: **end for**
11: calculate w_i^l using (4.29)-(4.31).
12: resampling

4.4 Some Important Issues

In this section, several important issues, including a theoretical analysis on the computational complexity, sampling from the normalized likelihood function and RSS-based localization in inhomogeneous propagation environments, will be discussed.

4.4.1 Computational Complexity

The four main parts of Algorithm 3, including updating $m_{ij}(\alpha)$, $m_{ij}(\mathbf{x}_i)$, $B(\mathbf{x}_i)$ and $B(\alpha)$, will be analyzed in terms of computational complexity. To be general, we write $\mathbb{C}_{\mathrm{c}}(M, N)$ to denote the complexity of drawing M samples from an N-categorical distribution. First, we consider updating $m_{ij}(\alpha)$ using Algorithm 4. Importance weights $\{w_{ij\to\alpha}^{l,n}\}_{l=1}^L$ are calculated with a complexity order of $\mathcal{O}(L^2)$ according to (4.9b) in the belief propagation, but $\mathcal{O}(1)$ according to (4.13a) in the SPAWN. Evaluating $m_{ij}(\alpha)$ at $\{\alpha_d^r\}_{r=1}^R$ requires operations of order $\mathcal{O}(L \cdot R)$. Second, for updating $m_{ij}(\mathbf{x}_i)$ using Algorithm 5, drawing samples $\{\alpha^l\}_{l=1}^L$ from $B(\alpha)$ needs operations of order $\mathcal{O}(\mathbb{C}_{\mathrm{c}}(L, R))$,

$m_{ij}(\alpha)$	importance weight	$\mathcal{O}(L^2)$	(4.9b) in belief propagation
		$\mathcal{O}(1)$	(4.13a) in SPAWN
	evaluating $m_{ij}(\alpha_d^r)$	$\mathcal{O}(L \cdot R)$	
$m_{ij}(\mathbf{x}_i)$	sample α	$\mathcal{O}(\mathbb{C}_c(L, R))$	
	importance weight	$\mathcal{O}(L^2)$	(4.11b) in belief propagation
		$\mathcal{O}(1)$	(4.13b) in SPAWN
	normalization	$\mathcal{O}(L)$	

Table 4.1: Complexity for updating $m_{ij}(\alpha)$ and $m_{ij}(\mathbf{x}_i)$.

$B(\mathbf{x}_i)$	importance sampler	sampling	$\mathcal{O}(L)$
		importance weight	$\boldsymbol{\mathcal{O}(\lvert\Gamma_i\rvert \cdot L^2)}$
		resampling	$\mathcal{O}(L)$
	auxiliary importance sampling	label indicator	$\mathcal{O}(\lvert\Gamma_i\rvert \cdot \mathbb{C}_{\mathrm{c}}(\frac{L}{\lvert\Gamma_i\rvert}, L))$
		position sample	$\mathcal{O}(L)$
		importance weight	$\boldsymbol{\mathcal{O}(\lvert\Gamma_i\rvert \cdot L)}$
		resampling	$\mathcal{O}(L)$
$B(\alpha)$		evaluate $B(\alpha_d^r)$	$\mathcal{O}(\lvert\Gamma\rvert \cdot R)$

Table 4.2: Complexity for updating $B(\mathbf{x}_i)$ and $B(\alpha)$. Here, the bold fonts are used to highlight the reduction in computational complexity, where the quadratic order in the importance sampler is reduced to the linear order in the proposed AIS.

calculating $\{w_{ij\to\mathbf{x}_i}^{l,n}\}_{l=1}^{L}$ has the same complexity as for $\{w_{ij\to\alpha}^{l,n}\}_{l=1}^{L}$, and converting $m_{ij}(\mathbf{x}_i)$ from (4.11a) to (4.12a) is done with a complexity order of $\mathcal{O}(L)$. Third, $B(\mathbf{x}_i)$ can be updated either using the importance sampler or using the proposed AIS. For the importance sampler in (4.17), L position samples and the corresponding importance weights are obtained with complexity orders of $\mathcal{O}(L)$ and $\mathcal{O}(|\Gamma_i| \cdot L^2)$, respectively. The subsequent resampling is conducted with a complexity order of $\mathcal{O}(L)$ [HSG06]. For the proposed AIS in Algorithm 6, generating L label indicators has a complexity order of $\mathcal{O}(|\Gamma_i| \cdot \mathbb{C}_{\mathrm{c}}(\frac{L}{|\Gamma_i|}, L))$ approximately. Generating L position samples and calculating L importance weights according to (4.31) have complexity orders of $\mathcal{O}(L)$ and $\mathcal{O}(|\Gamma_i| \cdot L)$,

respectively. The resampling step requires additional operations of order $\mathcal{O}(L)$. Lastly, $B(\alpha)$ is updated by simply multiplying $|\Gamma|$ real-valued numbers for R times, according to (4.14).

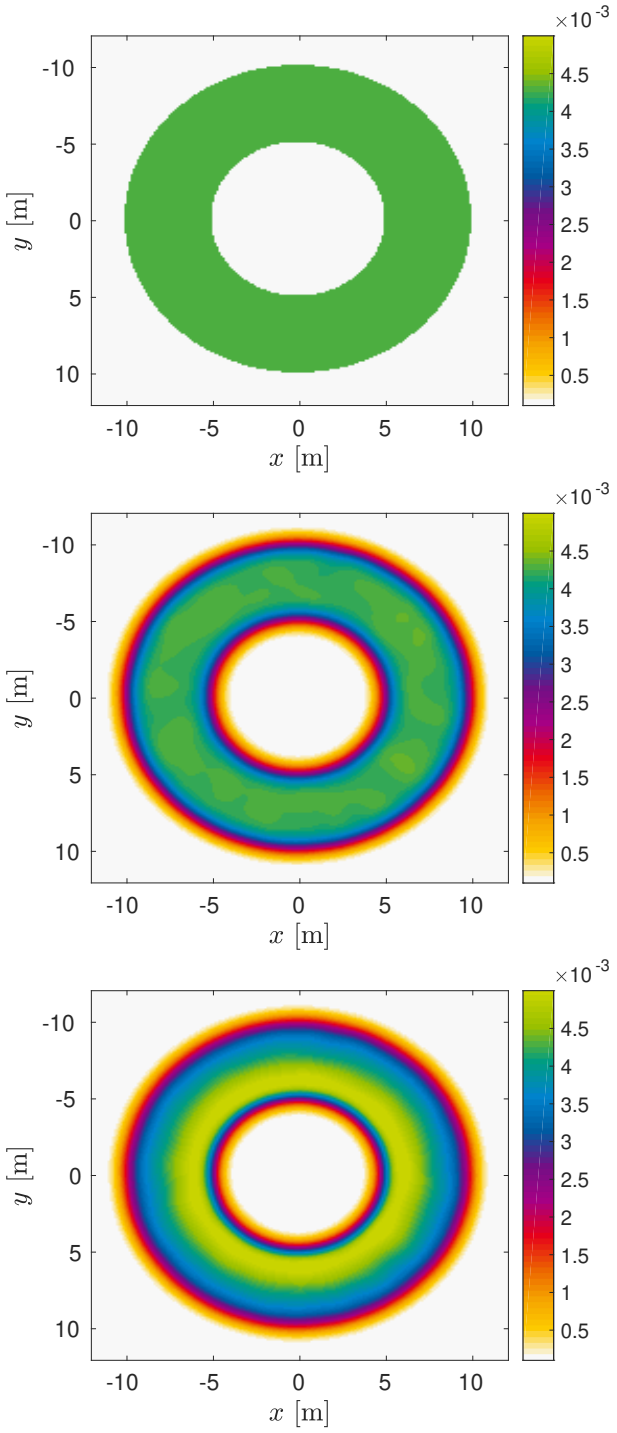


Figure 4.2: A comparison between the proposed sampler (middle) and the heuristic sampler (below) versus the groundtruth (top) for sampling $\mathbf{x}_i$ from $Z^{-1}f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^{l'})$ for the measurement model $r_{ij} = d_{ij} + v$ with $d_{ij} = 7.5$ and $v \sim \mathcal{U}[-2.5, 2.5]$.

The computational complexities for updating messages and beliefs are summarized in Table 4.1 and Table 4.2, respectively. For updating $m_{ij}(\alpha)$ and $m_{ij}(\mathbf{x}_i)$, the belief propagation is computationally substantially more intensive than the SPAWN, see Table 4.1. Regarding updating $B(\mathbf{x}_i)$ using the importance sampler, calculating the importance weights is computationally the most intensive step, requiring operations of order $\mathcal{O}(|\Gamma_i| \cdot L^2)$. Thanks to the introduction of the auxiliary variable $\boldsymbol{\psi}$ in the proposed AIS, the quadratic complexity order is reduced to the linear order $\mathcal{O}(|\Gamma_i| \cdot L)$, see Table 4.2.

4.4.2 Sampling From a Normalized Likelihood Function

In this section, we return to the problem that we have addressed in developing the AIS in Section 4.3.5. That is sampling from the normalized likelihood function $Z^{-1} f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^{l'})$. The sampling strategy proposed by us is essentially an importance sampler combined with random variable transformation. With the help of random variable transformation, the position samples are generated according to (4.28), and the associated proposal distribution $q(\mathbf{x}_i|\mathbf{x}_j^{l'}, r_{ij})$ is derived. In the context of this sampling problem, an importance weight, denoted by $w(\mathbf{x}_i^l)$, is assigned to the sample $\mathbf{x}_i^l$, as given by

$$w(\mathbf{x}_i^l) \propto \frac{f(r_{ij}|\mathbf{x}_i^l, \mathbf{x}_j^{l'})}{q(\mathbf{x}_i^l|\mathbf{x}_j^{l'}, r_{ij})} = \frac{f(r_{ij}|d_{ij}^l)}{q_d(d_{ij}^l|r_{ij})} \cdot d_{ij}^l, \tag{4.32}$$

where $f(r_{ij}|d_{ij}^l)$ is $f(r_{ij}|\mathbf{x}_i^l, \mathbf{x}_j^{l'})$ with $\|\mathbf{x}_i^l - \mathbf{x}_j^{l'}\|$ replaced by d_{ij}^l. This sampling strategy is related to a heuristic sampling strategy in [IFMW05]. A straightforward extension of this heuristic sampler leads to the same sample-generating mechanism, i.e., (4.28). The heuristic sampling strategy in [IFMW05] differs from our sampler in that these samples are deemed as following the normalized likelihood function, $Z^{-1} f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^{l'})$, though they actually follow $q(\mathbf{x}_i|\mathbf{x}_j^{l'}, r_{ij})$. A question that naturally arises is that under which condition are $Z^{-1} f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j^{l'})$ and $q(\mathbf{x}_i|\mathbf{x}_j^{l'}, r_{ij})$ proportional. Referring to the relation in (4.32), it holds only under the condition

$$f(r_{ij}|d_{ij}) \propto q_d(d_{ij}|r_{ij})/d_{ij}. \tag{4.33}$$

Unfortunately, the condition in (4.33) is not fulfilled in general, and, hence, the heuristic sampler may suffer from performance loss.

Next, we will compare these two samplers in a concrete example. Consider the measurement model $r_{ij} = d_{ij} + v$ with the true distance $d_{ij} = 7.5$. Suppose that the measurement error $v \sim \mathcal{U}[-2.5, 2.5]$ is uniformly distributed in the interval $[-2.5, 2.5]$.

For a given measurement r_{ij} and a predefined sample $\mathbf{x}_j^{l'}$, the purpose is to generate a set of (weighted) samples $\left\{\mathbf{x}_i^l\right\}_{l=1}^L$ that can approximately represent the normalized likelihood function $Z^{-1}f(r_{ij}\big|\mathbf{x}_i, \mathbf{x}_j^{l'})$. Without loss of generality, we consider that the measurement is $r_{ij} = 7.5$ and $\mathbf{x}_j^{l'}$ is fixed at $(0, 0)$. First, a large amount of samples are generated using (4.28). For our sampler, the importance weight associated with each sample is calculated using (4.32). Second, in order to visualize these two samplers for comparison purposes, we reconstruct the plots based on the (weighted) samples, which can be done, for instance, using a (weighted) kernel density estimator [Sil86]. The reconstruction of our sampler and that of the heuristic sampler are illustrated in the middle and the below plots in Fig. 4.2, respectively. As groundtruth, the true normalized likelihood function is depicted in the top plot in Fig. 4.2.

Several observations from Fig. 4.2 are listed as follows: First, the true normalized likelihood function (top), i.e., $Z^{-1}f(r_{ij}\big|\mathbf{x}_i, \mathbf{x}_j^{l'})$, shows a ring-shape, implying that the measurement r_{ij} is equally probable for any position $\mathbf{x}_i$ in the green area. Second, both samplers are able to roughly represent the ring-shape of $Z^{-1}f(r_{ij}\big|\mathbf{x}_i, \mathbf{x}_j^{l'})$, but with different accuracy. The ring-shape of the heuristic sampler (below) is not equally valued. The inner-ring has higher values than the outer-ring. This implies that, using the heuristic sampling strategy, the measurement r_{ij} becomes more probable for $\mathbf{x}_i$ that is located closer to $\mathbf{x}_j$. Our sampler (middle) can approximate $Z^{-1}f(r_{ij}\big|\mathbf{x}_i, \mathbf{x}_j^{l'})$ more accurately than the heuristic sampler, since the ring-shape of our sampler is more equally valued.

Note that the difference between both samplers lies essentially in the importance weight. Our sampler computes the importance weight according to (4.32), for which the underlying pdf $q(\mathbf{x}_i|\mathbf{x}_j^{l'}, r_{ij})$ associated with (4.28) is derived rigorously. In contrast, the heuristic sampling strategy ignores the mismatch between $q(\mathbf{x}_i|\mathbf{x}_j^{l'}, r_{ij})$ and $Z^{-1}f(r_{ij}\big|\mathbf{x}_i, \mathbf{x}_j^{l'})$, and simply assign equal weights to all samples. Consequently, the equally weighted samples may not represent the normalized likelihood function, $Z^{-1}f(r_{ij}\big|\mathbf{x}_i, \mathbf{x}_j^{l'})$, well. Note that this deviation becomes severe, when the likelihood function covers a broad area. On the other hand, when the likelihood function is sharp, both samplers can provide very satisfactory approximation results. As shown in the example above, the heuristic sampler deviates from the groundtruth. Such a deviation can be justified theoretically. It has been shown that the equally weighted samples of the heuristic sampler follow the desired normalized likelihood function $Z^{-1}f(r_{ij}\big|\mathbf{x}_i, \mathbf{x}_j^{l'})$ only under the condition in (4.33). In this example, this condition is not fulfilled here, since we have $f(r_{ij}|d_{ij}) = q_d(d_{ij}|r_{ij}) = f_v(r_{ij} - d_{ij})$.

4.4.3 RSS-Based Localization in Inhomogeneous Propagation Environments

Although developing RSS-based cooperative localization algorithm for inhomogeneous environments is beyond the scope of this chapter, a discussion on this issue is important. If no prior experimental campaign can be conducted, RSS-based cooperative localization for inhomogeneous environments is a challenging problem and has been rarely studied. In the context of non-cooperative localization, there are several works on RSS-based localization in inhomogeneous environments. In [MBL+09], the unknown and unequal path loss exponents are estimated using measures of compatibility of the distance estimates. In [Man16], a generalized two-region channel model is used and the path loss exponent in the NLOS region is adapted online while positioning. Similarly, a mode-dependent propagation model is proposed in [YFGZ13a], and all unknown parameters are jointly estimated using the expectation-maximization criterion. Moreover, the recent work [LZJ+19] has proposed a piecewise convex approximation aided localization dealing with the nonlinear and non-convex problem, resulting from the unknown and unequal path loss exponents.

If it is affordable to conduct an experimental campaign to build a dataset, fingerprinting methods, e.g., [BP00], can generally provide quite satisfying localization performance. However, the major drawbacks of the fingerprinting approaches lie in the requirement of a dataset and the inability to adapt to a new environment. Furthermore, besides the dataset, if more advanced measurements can be collected, such as the waveform of ultra-wideband signals, there are various potential solutions for cooperative localization in inhomogeneous environments. One potential solution is the paradigm based on the concept of soft-range-information [MCAW18,CMB+19]. In contrast to the conventional techniques that rely on the most likely distance estimate, soft-range information enables soft-decision localization by capturing the odds of all possible distances. Based on the training data collected during the experimental campaign, a generative model is estimated via machine learning and then stored, and the soft-range information is estimated using the stored generative model.

4.5 Numerical Results

In this section, we comprehensively evaluate the proposed algorithms. For comparison purpose, Tomic's SDP estimator from [TBD15], which outperforms other existing works [VGBS13, WCLJ12], is chosen as a competitor. Here, the SDP estimator is slightly

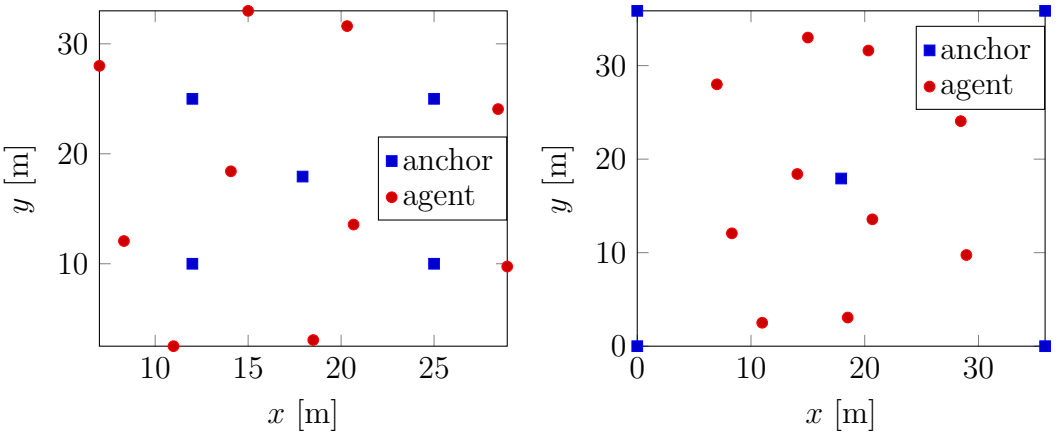


Figure 4.3: Network layout: Network I (left) and Network II (right)

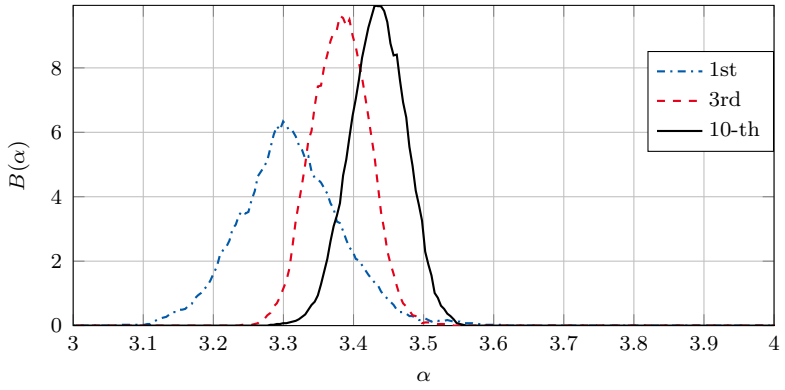
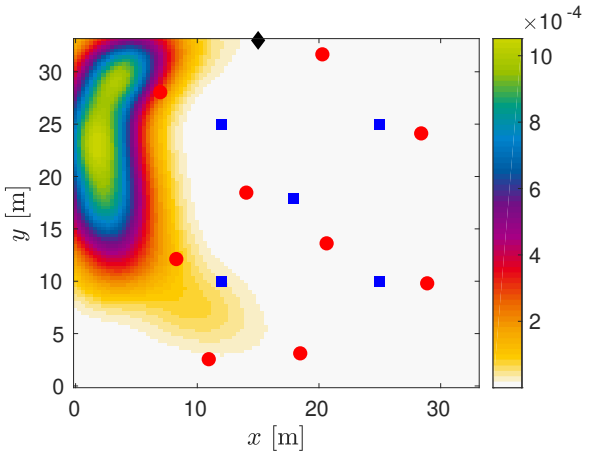
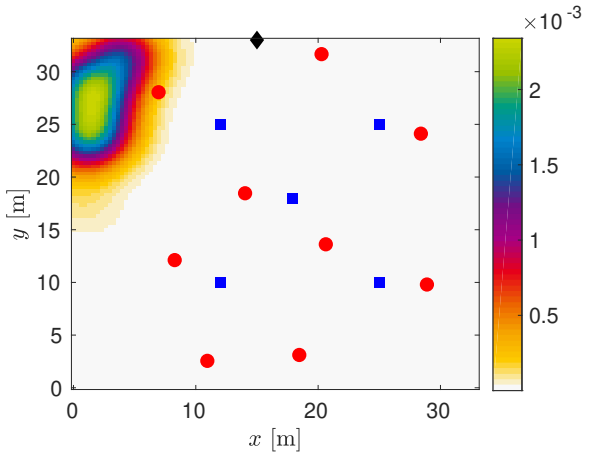


Figure 4.4: Example of $B(\alpha)$ over iterations with the true $\alpha = 3.5$.

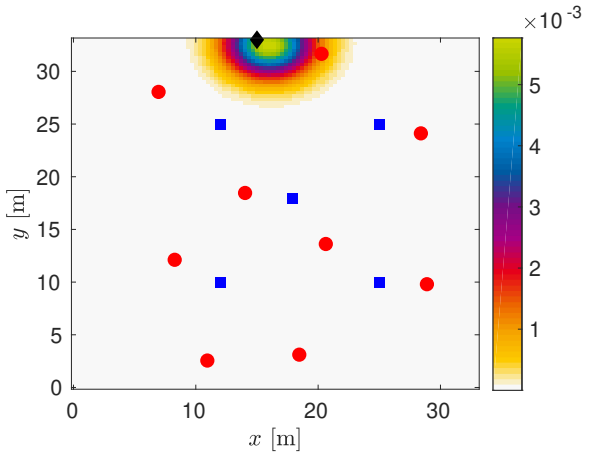
adjusted so that the estimate of the path loss exponent is constrained in the predefined region, in accordance with $f(\alpha)$. Note that such an adjustment can improve the original SDP estimator, since unreasonable estimates of the path loss exponent are avoided. The SDP estimator terminates, either when $N_{\max} = 100$ iterations are achieved or when $|C(n) - C(n-1)|/|C(n-1)|$ is smaller than 10^{-5}, where $C(n)$ is the logarithm of the cost function in the n-th iteration. The convex optimization problem in the SDP estimator is solved using the CVX Toolbox [GB17] with the SeDuMi solver. For the proposed algorithms, the maximal number of iterations is set to $N_{\max} = 10$, $L = 1000$ particles are used, and $R = 100$ grid points $\{\alpha_d^r\}_{r=1}^R$ are chosen. For a fair comparison with the SDP estimator, in the proposed algorithms, a point estimate is further inferred from the marginal posterior estimate for each unknown parameter. This is done by finding the highest mode of the analytical form of $B(\mathbf{x}_i)$, which is recovered using kernel density estimation, based on the samples of $B(\mathbf{x}_i)$. Due to the fact that both the belief-propagation-IS and the belief-propagation-AIS are computationally very intensive, we



(a)



(b)



(c)

Figure 4.5: Example of $B(\mathbf{x}_i)$ in the 1st (a), 3rd (b) and 10-th (c) iteration. The agent of interest locates at ◆, the other agents locate at ●, and anchors locate at ■.

will only demonstrate the performance of the SPAWN-IS and that of the SPAWN-AIS.

As shown in many experimental works, such as in [PHP+03a, CDG+14, Rap01], the value of the path loss exponent is determined by the propagation environment. In general, for an LOS propagation link, we have $\alpha \in [1.5, 2]$, and for an NLOS link, we have $\alpha \in (2, 6]$. The propagation environment considered in this chapter is homogeneous, meaning that a common path loss exponent, α, is shared by all propagation links. Accordingly, the proposed algorithms will be evaluated mainly for homogeneous environments. In Section 4.5.1, environments with the following six values of α, i.e., $\{1.5, 2, 3, 4, 5, 6\}$, are considered, corresponding to a tunnel-like scenario ($\alpha = 1.5$), a free-space scenario ($\alpha = 2$) and different NLOS scenarios ($\alpha = 3$, 4, 5 or 6). To make the simulation complete, inhomogeneous propagation environments are considered in Section 4.5.3 as well.

We choose two representative networks with 10 agents and 5 anchors: Network I, in which some of the agents are outside the convex hull of the anchors, and Network II, in which all agents are within the convex hull, see Fig. 4.3. The reference power is set to $A_i = -30$ dBm for all $i \in S$ at a reference distance of $d_0 = 1$ meter. For a fair comparison, we set the prior distribution of the path loss exponent α as a uniform distribution, $\alpha \sim \mathcal{U}[1.5, 6]$, and that of each position as a uniform distribution in a square area that is determined by the maximum and the minimum of all nodes' positions. All simulation results are based on 100 Monte Carlo runs. The mean squared error (MSE) of the estimator $\hat{\alpha}$, the bias of $\hat{\alpha}$ and the RMSE, defined in [YFJ+15], are chosen as performance metrics.

4.5.1 Varying Path Loss Exponent

In this section, the goal is to investigate the performance of the proposed algorithms at different values of the path loss exponent in different network layouts. We set the standard deviation of the measurement error to $\sigma = 3$ and the communication range to 20 meter. As an illustrative example, we first demonstrate how the beliefs evolve with iterations and depict the kernel density estimates of $B(\alpha)$ and $B(\mathbf{x}_i)$ in Figs. 4.4-4.5, respectively. It is observed in Fig. 4.4 that over iterations, the uncertainty on α reduces, and $B(\alpha)$ moves towards the true path loss exponent $\alpha = 3.5$, as shown in Fig. 4.5. Similarly, over iterations $B(\mathbf{x}_i)$ becomes more and more concentrated and shifts towards the true position. It is noteworthy that the prior distributions adopted are quite coarse. More precisely, a uniform distribution $\mathcal{U}[1.5, 6]$ is used for the path loss exponent. Even so, the proposed algorithms can provide marginal posterior estimates that are relatively sharp and close to the true parameters.

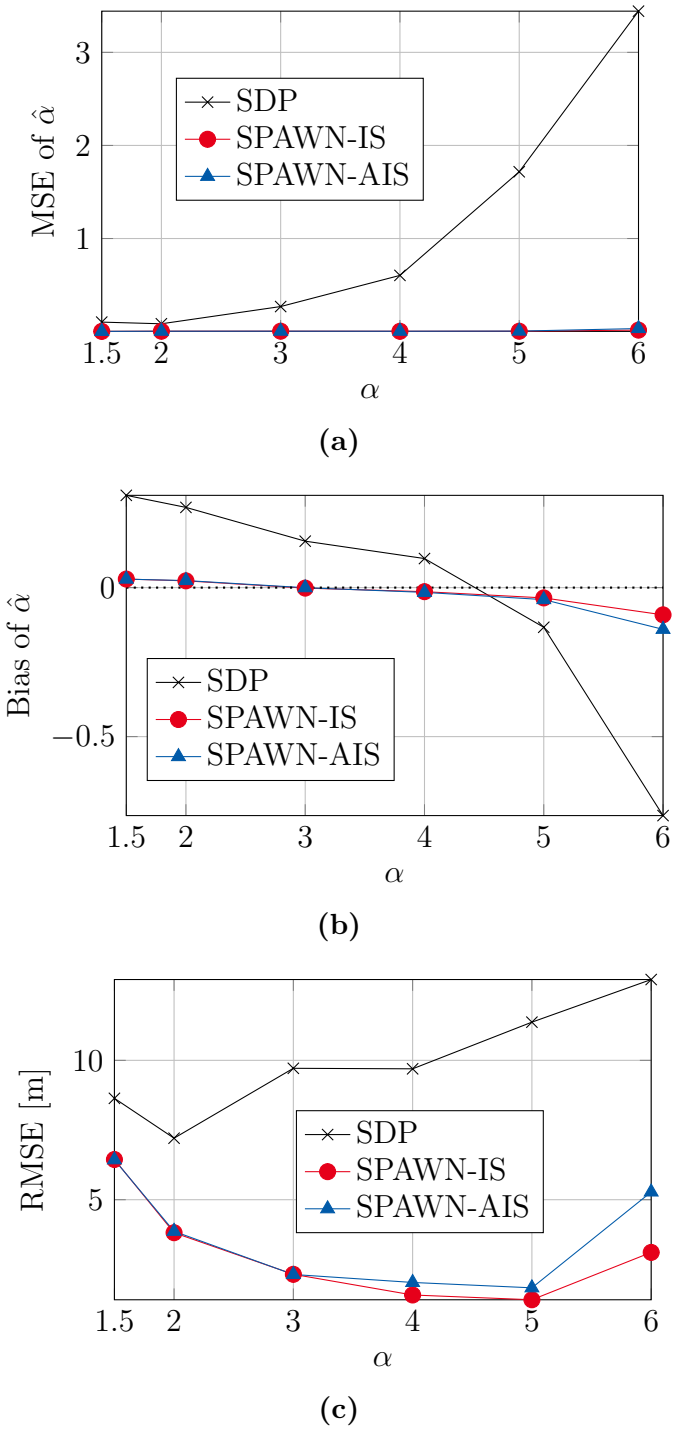


Figure 4.6: Network I: (a) The MSE of $\hat{\alpha}$, (b) the bias of $\hat{\alpha}$ and (c) the RMSE versus the true path loss exponent α. Here, the standard deviation of the measurement error is $\sigma = 3$, and the communication range is 20 meter.

The overall performance of different algorithms is evaluated in terms of the MSE of $\hat{\alpha}$, the bias of $\hat{\alpha}$ and the RMSE. The results are depicted in Fig. 4.6 and Fig. 4.7 for

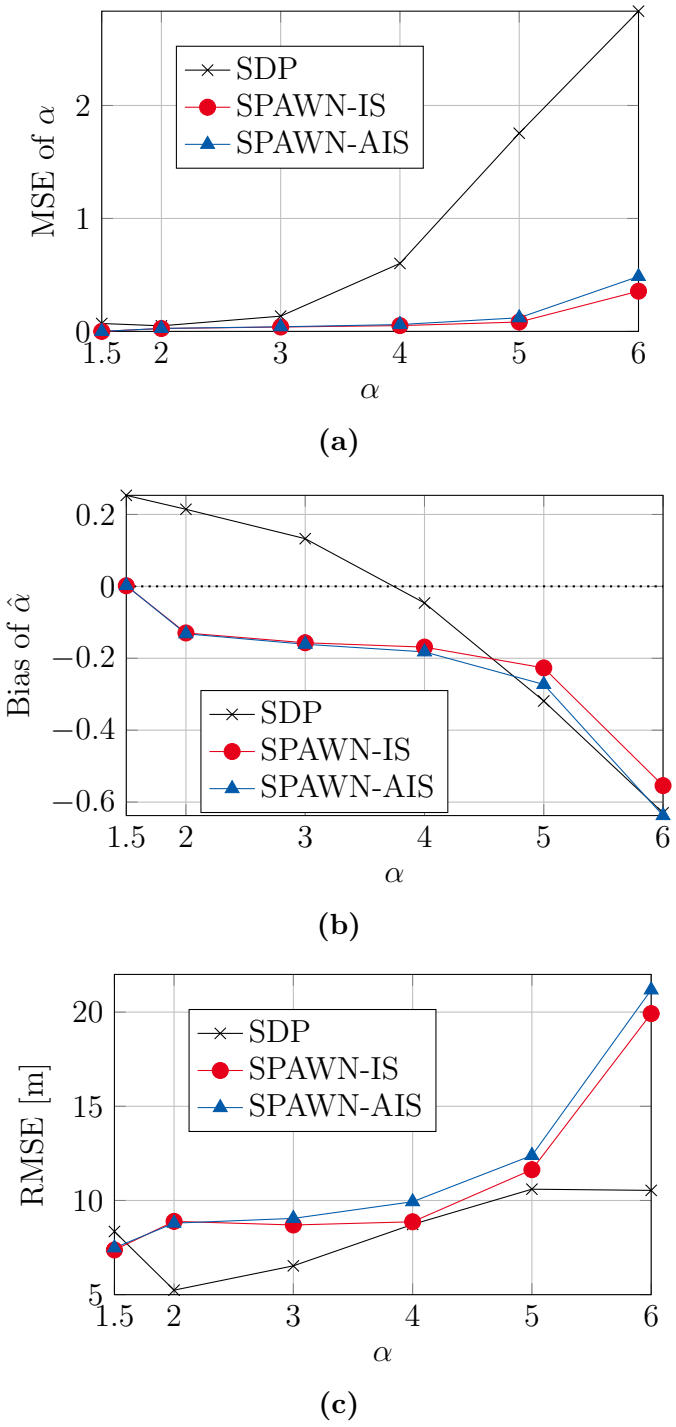
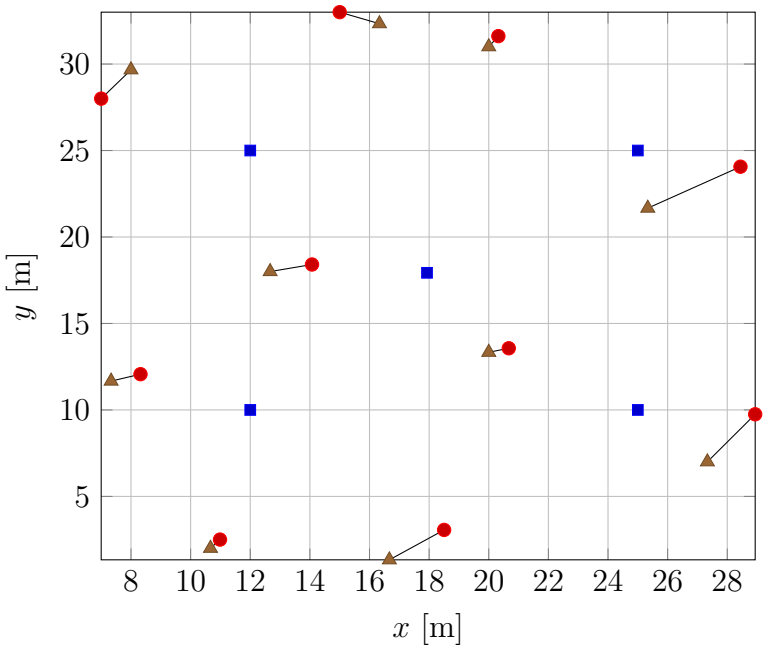
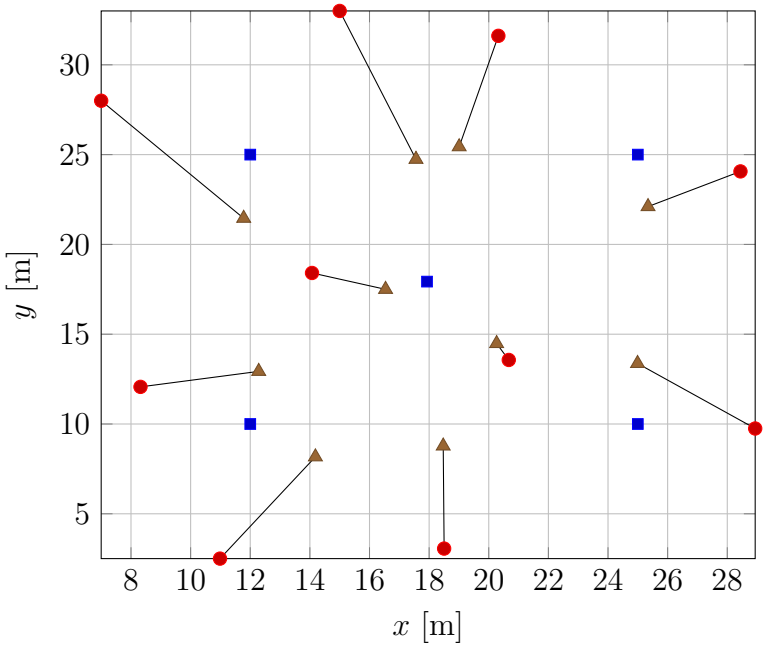


Figure 4.7: Network II: (a) The MSE of $\hat{\alpha}$, (b) the bias of $\hat{\alpha}$ and (c) the RMSE versus the true path loss exponent α. Here, the standard deviation of the measurement error is $\sigma = 3$, and the communication range is 20 meter.

Networks I and II, respectively. For Network I, it is remarkable that, both proposed algorithms (SPAWN-IS and SPAWN-AIS) have comparable performance for both the



(a)



(b)

Figure 4.8: Position estimates obtained by the SPAWN-AIS (a) and by the SDP (b) with anchors ■, agents ●, estimated agents ▲ .

path loss exponent α and the position $\mathbf{x}_i, i \in S_u$, though the SPAWN-AIS has a significantly lower computational complexity than the SPAWN-IS. Compared with the SDP estimator, the proposed algorithms have improved performance for both α and

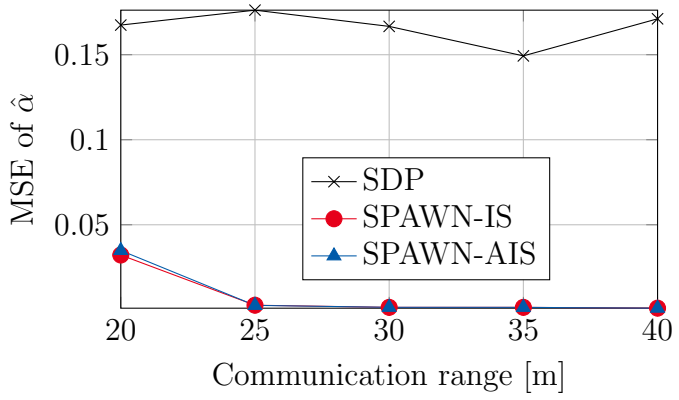
$\mathbf{x}_i$, $i \in S_u$. For a better visualization, we depict the representative position estimates obtained from both the SPAWN-AIS and the SDP estimator in Fig. 4.8. It is clear to see that the localization accuracy of the SDP estimator is quite low if part of the agents are outside the convex hull of the anchors, while the SPAWN-AIS does not suffer from this problem. We notice that this type of network topology is rarely examined in the existing literature, although its existence is very probable in practical networks.

Furthermore, it is observed in Fig. 4.6 that the RMSE curves of the proposed algorithms first drop and then rise as the true path loss exponent α varies. It results from two factors: the true path loss exponent α and the estimation performance of $\hat{\alpha}$. For $\alpha \in \{1.5, 2, 3, 4, 5\}$, the proposed algorithms provide relatively accurate estimation of α. It can be approximately treated as if the path loss exponent is known. As theoretically demonstrated in [PHP+03a], for RSS-based localization with known path loss exponent, the Cramér Rao bound is proportional to σ^2/α^2. It means that the localization performance of RSS-based algorithms generally improves when the ratio σ/α decreases (when α increases for a fixed σ). This explains why the RMSE curves first drop. On the other hand, when the true underlying α is 6, $\hat{\alpha}$ tends to underestimate α. The influence of an incorrect path loss exponent (bias of $\hat{\alpha}$) on RSS-based localization has been studied in [SKG12]. As shown in [SKG12], the localization error becomes considerably large, when $\hat{\alpha}$ underestimates α. This explains why the curves rise at $\alpha = 6$.

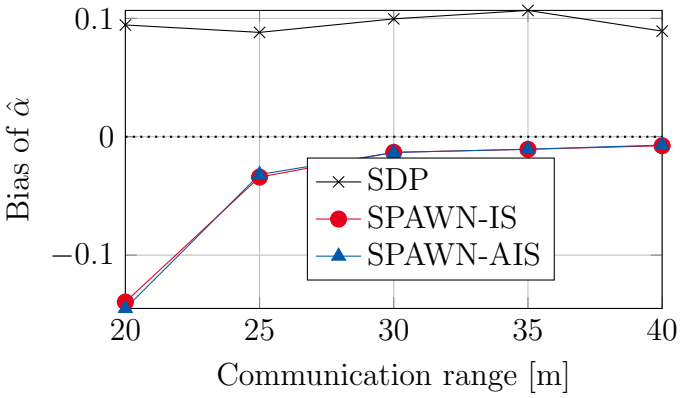
For Network II (Fig. 4.7), again, the MSE curve of $\hat{\alpha}$ of the proposed algorithms is under that of the SDP estimator, meaning that the proposed algorithms have quite stable estimation performance for α. However, for this network, the localization accuracy of the proposed algorithms is comparable to or slightly lower than that of the SDP estimator. This localization performance degradation in the proposed algorithms results from a biased estimation of α, which can be seen in the middle plot in Fig. 4.7. The possible reason lies in that there could be certain performance loss when we infer the unknown parameter from its marginal posterior, instead of jointly inferring all unknown parameters from the joint posterior. Nevertheless, this problem will be alleviated either when the communication range increases or when the measurement noise decreases, as will be demonstrated in the following simulations.

4.5.2 Varying Communication Range and Standard Deviation

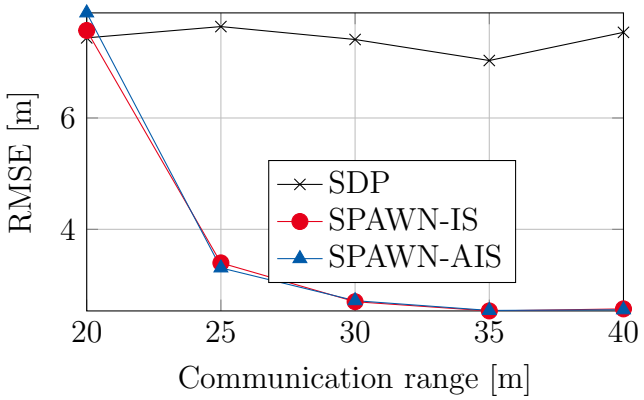
The goal in this section is to assess the performance of the proposed algorithms at varying communication range and varying standard deviation of the measurement error.



(a)



(b)



(c)

Figure 4.9: Network II: (a) The MSE of $\hat{\alpha}$, (b) the bias of $\hat{\alpha}$ and (c) the RMSE versus the communication range. Here, the standard deviation of the measurement error is $\sigma = 3$, and the true underlying path loss exponent is $\alpha = 3$.

It has been shown that for Network I the proposed algorithms have quite satisfying performance for both the positions and the path loss exponent. Hence, in the following

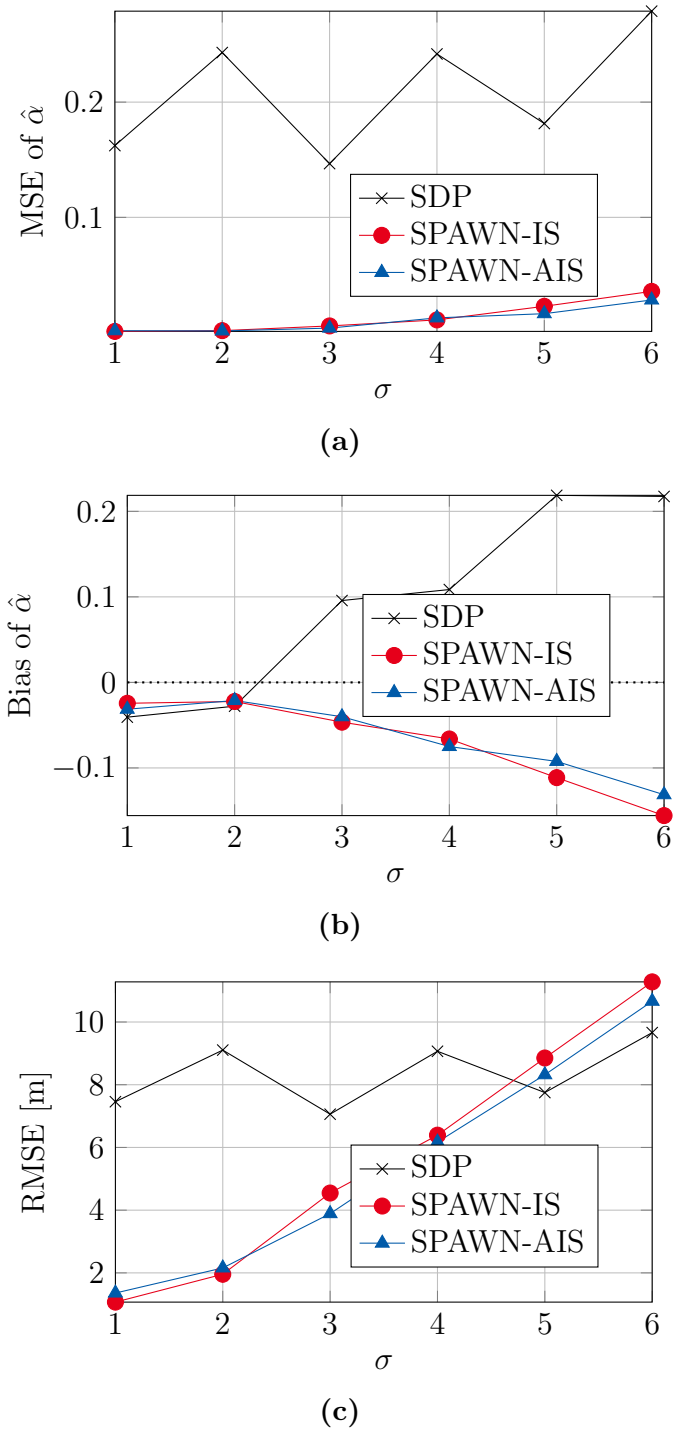
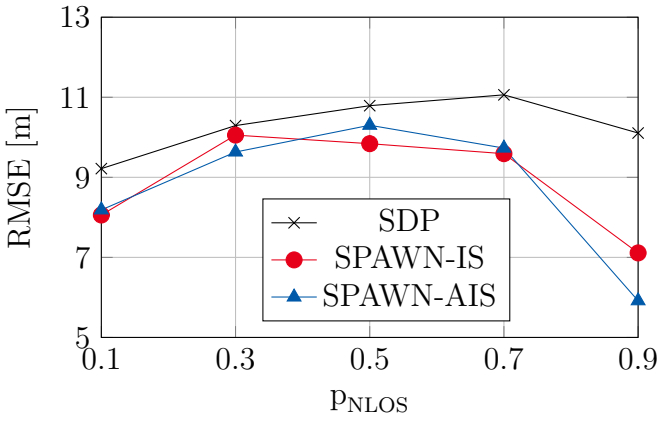
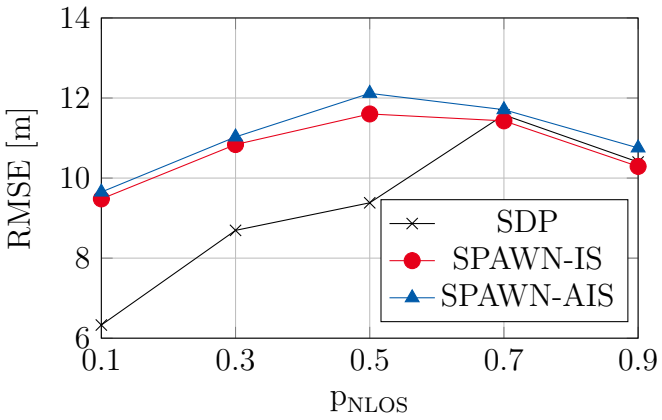


Figure 4.10: Network II: (a) The MSE of $\hat{\alpha}$, (b) the bias of $\hat{\alpha}$ and (c) the RMSE versus the standard deviation of the measurement error. Here, the true underlying path loss exponent is $\alpha = 3$, and the communication range is 25 meter.

simulations, we will only focus on Network II. For the simulation with varying communication range, the true path loss exponent and the standard deviation are set to $\alpha = 3$



(a)



(b)

Figure 4.11: Performance obtained in the inhomogeneous environments for (a) Network I and (b) Network II. Part of the propagation links (LOS) have $\alpha = 2$ and the others (NLOS) have $\alpha = 4$. The probability that one link is in the NLOS condition, denoted as $\mathrm{p_{NLOS}}$, varies from 0.1 to 0.9 with a stepsize of 0.2. Here, the standard deviation of the measurement is $\sigma = 3$, and the communication range is 25 meter.

and $\sigma = 3$, respectively, and for the other simulation, the true path loss exponent is set to $\alpha = 3$, and the communication range is set to 25 meter. The results are depicted in Fig. 4.9 and Fig. 4.10 for the cases of varying communication range and varying standard deviation, respectively.

From the figures we can see that for the proposed algorithms, all three error-curves drop substantially and eventually attain relatively small values when the communication range increases or when the standard deviation decreases. In contrast, no obvious improvement is seen for the SDP estimator. This result is expected and can be explained as follows. In the proposed Bayesian algorithms, the marginal posterior of

each unknown parameter is inferred. When more information is collected, for instance through increasing communication range (network connectivity) or through decreasing measurement error, the marginal posterior can reflect the unknown parameter more accurately. On the other hand, the SDP estimator suffers from certain performance loss due to the relaxation procedure, and this performance loss may be so dominating that the increase in the information cannot improve the estimation accuracy any more. This result highlights that the proposed algorithms can benefit from the increase in the information to a large extent. Lastly, we stress that although the SPAWN-AIS has a significant reduction on computational cost, it achieves similar estimation performance as the SPAWN-IS.

4.5.3 Inhomogeneous Propagation Environments

Here, we consider the case when part of the propagation links are LOS links and the others are NLOS links. The probability that one measurement is in the NLOS condition is denoted by P_{NLOS}. To represent different scenarios, it varies from 0.1 to 0.9 with a stepsize of 0.2. A LOS measurement is generated using (4.1) with $\alpha = 2$; while an NLOS one is generated with $\alpha = 4$. The other parameters stay the same as in Section 4.5.1. The localization performance of all algorithms is depicted in Fig. 4.11 for both networks. A certain performance degradation is observed for the proposed algorithms (SPAWN-IS and SPAWN-AIS), and for the competing algorithm as well. Although the proposed algorithms have lower localization accuracy for network II, they consistently outperform the SDP estimator for network I. It can be seen from Fig. 4.11 that almost all error-curves follow the ascending-descending trend, for instance, the RMSE curve of the SPAWN-IS algorithm in the top plot starts from 8 [m] at $P_{\text{NLOS}} = 0.1$, approaching 10 [m] at $P_{\text{NLOS}} = 0.5$, ends at 7 [m] at $P_{\text{NLOS}} = 0.9$. This observation implies that as the propagation environment deviates more from a homogeneous one, the performance degradation becomes more prominent. However, such performance degradation is expected, because none of these algorithms are developed for the inhomogeneous environments.

4.6 Summary

In this chapter, cooperative localization using RSS measurements in homogeneous mixed LOS/NLOS environments was considered for the case where the key model parameter (path loss exponent) is unknown. The cooperative localization problem

was formulated as a statistical inference problem in a Bayesian framework with all unknown parameters being treated as random variables. The goal was to infer the marginal posterior of each unknown parameter, from which a position estimate, as well as its uncertainty information, can be obtained. This was achieved by deriving message passing methods and designing appropriate numerical strategies. One notable advantage of the resulting message passing methods is that the localization performance does not require a good initialization. Comprehensive numerical studies were conducted to demonstrate the performance of the proposed message passing methods. It was shown that the proposed message passing methods have the potential of achieving improved localization accuracy and provide accurate estimates of the path loss exponent.

Chapter 5

Robust Cooperative Localization via Exploiting Sparsity of Ranging Biases

In this chapter, we consider a more general case where statistical knowledge of the LOS/NLOS measurement errors is completely unknown and range measurements, which are believed to be accurate but quite sensitive to NLOS propagation, are available. This is addressed by first modeling each ranging bias as an unknown parameter. It is shown that the ranging biases are sparse in LOS-dominating environments. This sparsity is then exploited by introducing a sparsity-promoting term in the conventional cost functions, giving rise to a generic sparsity-promoting regularized formulation. By bounding the sum of the residual errors, an alternative generic bound-constrained regularized formulation is developed. To avoid the problem of local optima solutions, a specific SDP-based solver is developed for each generic regularized formulation. It is theoretically shown that, under certain conditions, these two SDPs provide equivalent optimal solutions. Furthermore, in order to select an appropriate regularization parameter, an efficient data-driven strategy is developed which exploits the special structure of the bound-constrained regularized SDP.

This chapter is mainly based on [JYF+20a, JYZS21]. The structure of this chapter is as follows: Section 5.1 discusses the challenge caused by NLOS propagation and the state-of-the-art methods dealing with NLOS measurements. Section 5.2 introduces the measurement model and formulates the problem mathematically. Section 5.3 is dedicated to the sparsity property of the ranging bias parameters and two generic regularized optimization problems, which are then solved via SDP in Section 5.4. In Section 5.5, we discuss the regularization parameter selection strategy. Numerical results are given in Section 5.6. Finally, Section 5.7 summarizes this chapter.

Notation: Boldface lowercase letter $\mathbf{a}$ and boldface uppercase letter $\mathbf{A}$ are reserved for vectors and matrices, respectively. $\mathbf{A}_{(l)}$ and $\mathbf{A}_{(l)}^T$ denote the l-th column of matrix $\mathbf{A}$ and $\mathbf{A}_{(l)}$'s transpose, respectively. $\mathbf{A}_{(a:b)}$ and $\mathbf{A}_{(a:b)}^T$ denote the submatrix of $\mathbf{A}$ constructed by its columns a to b and $\mathbf{A}_{(a:b)}$'s transpose, respectively. $\mathbf{A}_{[c:d,a:b]}$ denotes the submatrix of $\mathbf{A}$ constructed by its rows c to d and columns a to b. $\mathbf{I}_n$ is the $n \times n$ identity matrix. $\mathbf{e}_i$ is a vector whose i-th entry is one and the rest are zero. $\mathbf{1}_n$ is the $n \times 1$ vector whose entries are all equal to one. $\mathbf{0}_n$ is the $n \times 1$ vector whose entries are all zero. $\mathrm{tr}(\mathbf{A})$ stands for the trace of matrix $\mathbf{A}$. $\|\cdot\|$ represents the Euclidean norm, and $|\cdot|$ represents the cardinality of a set. $\mathcal{N}(\mu, \sigma^2)$ denotes a Gaussian distribution with

mean μ and variance σ^2. $\mathcal{U}[a,b]$ denotes a uniform distribution in the interval $[a,b]$. The symbol $\mathbb{R}^n_+$ stands for the n-dimensional non-negative orthant, and the symbol $\mathcal{S}^n_+$ represents the n-dimensional positive semi-definite cone. $\mathbf{A} \succeq \mathbf{0}$ indicates that $\mathbf{A}$ is a symmetric positive semi-definite matrix and $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B} \succeq \mathbf{0}$.

5.1 Localization in Mixed LOS/NLOS Environments

NLOS propagation is one of the main challenges in practical localization because it usually introduces a large positive bias to the range measurements, thereby severely degrading the positioning performance. To mitigate this effect, different approaches have been developed. The most common idea is to identify the LOS/NLOS condition of each link. Then, the measurements detected as NLOS are either discarded or considered with a low weight. However, this approach has two major shortcomings. First, reliable LOS/NLOS identification, such as applying machine learning-based techniques in [MGWW10], usually relies on additional laborious experimental campaigns to build a database. Such a data-collection phase is, in general, very time consuming and requires intensive human resources, which is unrealistic for many wireless network applications. Second, an environmental change may lead to mismatches between experimental propagation conditions and practical conditions, thereby deteriorating the LOS/NLOS identification accuracy and resulting in considerable localization errors. These two drawbacks also exist in other localization methods that require measurement campaigns, e.g., [WML+18].

To avoid the aforementioned drawbacks, localization in mixed LOS/NLOS environments without requiring LOS/NLOS identification or any experimental campaigns is of great interest. This problem has been studied in the literature. The majority of existing works are carried out in the context of the conventional non-cooperative paradigm [WCLA14, CWA19, YZ12, YFGZ13b, YFGZ14, FHKZ09, HWZ09, HZ11, VB07]. For cooperative localization, broadly speaking, there are two categories: parametric approaches [YZFG13, YFJ+15, YAZ+13, CWW+12] where the measurement errors are assumed to follow a certain probability distribution versus non-parametric approaches [GSG19, WZCL18, VB15, WGW+19, WGW+20] where no specific distribution is assumed to model the measurement errors.

In [YZFG13, YFJ+15, YAZ+13], the measurement error is modeled as a Gaussian mixture with unknown model parameters. Either the expectation-conditional maximization (ECM) or the expectation maximization criterion is adopted to approximate the

maximum likelihood estimator of the unknown positions as well as the Gaussian mixture model parameters. In [CWW+12], similar estimators are developed for three representative scenarios where different amounts of prior knowledge about the link conditions and the NLOS measurement error statistics are considered. Such parametric methods, however, suffer from significant performance degradation when there is a mismatch between the assumed model and the data. From this perspective, the following non-parametric cooperative methods are more advantageous.

In [GSG19], a location estimator is developed based on the principle of maximum entropy, i.e., the entropy of the distance residuals is maximized. In [WZCL18], the measurement errors are limited in certain intervals but with unknown bounds. Following the principle of robust optimization, localization is then formulated as a minimax problem, in which the sum of the residual errors in the least favorable case is minimized. The authors claim that this method works well in NLOS-heavy scenarios. In [VB15], a maximum likelihood estimator for the positions is approximated using a semi-definite program where the ranging biases are treated as nuisance parameters and a set of range-related constraints are designed. In addition, to constrain the ranging bias to a feasible value, a regularization term is added to the objective function. However, this method suffers from increased computational complexity and infeasibility, both are caused by the extra set of range-related constraints. In a similar vein, the maximum likelihood estimator of the positions and NLOS biases is relaxed to a semi-definite program in [WGW+19, WGW+20]. The novelty lies in that the spatial geometric relationship of the ranging biases is exploited in terms of optimization constraints. However, its practical behavior needs further verification, since the simulation results shown therein are merely based on small networks and all agents are closely located.

In this chapter, we aim to develop a non-parametric cooperative localization approach that requires neither LOS/NLOS identification nor experimental campaign in advance while still being robust against NLOS measurements. By treating the ranging bias, not only for NLOS but also for LOS, as an unknown parameter, we first indicate that the ranging biases are sparse in LOS-dominating environments and then exploit this sparsity property for NLOS mitigation.

5.2 Problem Formulation

Consider a wireless network in a 2-dimensional space. There are two sets of nodes: target nodes (*agents*), whose positions $\mathbf{x}_i = [x_i, y_i]^T$ are unknown, and reference nodes

(*anchors*), whose positions $\mathbf{x}_i$ are given. We let $S_u = \{1, \ldots, N_u\}$ be the index set of all agents and $S_a = \{N_u + 1, \ldots, N\}$ be that of all anchors. In general, the network is not fully connected. If there is a range measurement between node i and node j, these two nodes are called neighbors. Γ_i denotes the index set of node i's neighbors, and Γ denotes the set of all pairs of neighbors.

Following many existing studies, e.g., [CWW+12], the range measurement r_{ij} between nodes i and j in a mixed LOS/NLOS propagation environment is modeled as

$$r_{ij} = \begin{cases} \|\mathbf{x}_i - \mathbf{x}_j\| + v_{ij} & \text{for LOS,} \\ \|\mathbf{x}_i - \mathbf{x}_j\| + \delta_{ij} + v_{ij} & \text{for NLOS.} \end{cases} \tag{5.1}$$

Here, $\|\mathbf{x}_i - \mathbf{x}_j\|$ denotes the Euclidean distance between $\mathbf{x}_i$ and $\mathbf{x}_j$, v_{ij} denotes the measurement noise and δ_{ij} denotes the NLOS bias which is positive and unknown. Note that we consider a symmetric scenario, meaning that we have $r_{ij} = r_{ji}$. The range can be estimated from the TOA as well as the RSS measurements, which can be obtained using many technologies, such as WLAN, cellular networks, ultra-wideband based on radio frequencies as well as ultra and audible sound. For TOA and RSS measurements, NLOS propagation leads to large propagation time and strong attenuation, respectively [ZB11]. Both effects result in range estimates with positive biases[1].

The objective is to estimate the agent positions, $\mathbf{x}_i, \forall i \in S_u$, based on the anchor positions and all range measurements. For localization in mixed LOS/NLOS environments, fundamental questions include: how much information about NLOS propagation is available? More concretely, is there any available prior knowledge regarding LOS/NLOS identification? Is there any available prior statistical knowledge regarding v_{ij} and δ_{ij}; in other words, is the distribution of v_{ij} or δ_{ij} available? In this chapter, we consider the least favorable case. More precisely, we assume that 1) there is no LOS/NLOS identification information, and 2) there is no statistical knowledge of v_{ij} and δ_{ij}, except an assumption that v_{ij} is zero-mean. Note that, however, this zero-mean assumption on v_{ij} is a common and reasonable assumption.

5.3 Exploiting Sparsity of Ranging Biases

When the link status (LOS or NLOS) is not known in advance, it is natural to re-write the measurement model (5.1) in a unified form

$$r_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\| + \delta_{ij} + v_{ij}, \qquad \forall (i,j) \in \Gamma. \tag{5.2}$$

[1]The path loss exponent used for the range estimate is assumed to be 2.

If link (i, j) is an LOS connection, δ_{ij} is zero. If it is an NLOS link, δ_{ij} is an unknown positive number. This leads to the following two remarkable properties of the bias vector $\boldsymbol{\delta} = \{\delta_{ij}\}_{(i,j)\in\Gamma}$:

- $\boldsymbol{\delta}$ possesses a sparsity property in LOS-dominating environments, since the majority of the elements in vector $\boldsymbol{\delta}$ are zero;
- $\boldsymbol{\delta}$ is inherently elementwise non-negative, meaning that $\delta_{ij} \geq 0$, $\forall (i, j) \in \Gamma$.

Surprisingly, and to the best of our knowledge, the sparsity property of the bias vector $\boldsymbol{\delta}$ has not been studied so far. Our goal in the following is to exploit sparsity for cooperative localization.

5.3.1 Sparsity-Promoting Regularized Formulation

The objective now becomes inferring the unknown parameters, $\mathbf{x}_i$, $i \in S_u$, from $\mathbf{r} = \{r_{ij}\}_{(i,j)\in\Gamma}$ according to (5.2), where $\boldsymbol{\delta}$ is unknown. This set of $|\Gamma|$ equations in (5.2) describes an underdetermined system because there are $|\Gamma|+2\cdot N_u$ unknown parameters but only $|\Gamma|$ equations, where $|\Gamma|$ denotes the cardinality of set Γ. To overcome this problem, we propose to exploit the sparsity of $\boldsymbol{\delta}$ and formulate the localization problem as a sparsity-promoting regularized optimization problem:

$$\min_{\mathbf{X},\, \boldsymbol{\delta}\geq 0} \rho\big(\mathbf{X}, \boldsymbol{\delta}\big) + \gamma\, g(\boldsymbol{\delta}). \tag{5.3}$$

Here, $\mathbf{X}$ stands for the matrix of unknown positions, i.e., $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_u}]$. The first term, $\rho(\cdot)$, is a cost function related to the residual error. The second term is a regularization term. The function $g(\cdot)$ promotes the sparsity of $\boldsymbol{\delta}$. The predefined free parameter, γ, controls the tradeoff between the residual error term and the sparsity-inducing term. Such a compound objective function enables the optimizer to find a solution that contains many zero elements in $\boldsymbol{\delta}$ and results in a small residual error. Hence, the purpose of jointly estimating the positions and the bias parameters can be achieved by solving this optimization problem. Based on the fact that there is a sparsity-promoting regularization term, we call the generic formulation of (5.3) the *sparsity-promoting regularized formulation*.

Name & Reference	Cost function	Advantage
Log-likelihood [PHP+03a, SL14]	$\rho(\mathbf{X}, \boldsymbol{\delta}) = -\sum_{(i,j)\in\Gamma} \ln\big(f_{ij}(\mathbf{x}_i, \mathbf{x}_j, \delta_{ij}\|r_{ij})\big)$	Asymptotic efficient
Sum of squared errors [Li07, WLW09]	$\rho(\mathbf{X}, \boldsymbol{\delta}) = \sum_{(i,j)\in\Gamma} w_{ij}\big(r_{ij} - \|\mathbf{x}_i - \mathbf{x}_j\| - \delta_{ij}\big)^2$	Statistical information not required
Huber loss [ZKCM12, Kv09]	$\rho(\mathbf{X}, \boldsymbol{\delta}) = \sum_{(i,j)\in\Gamma} h_\eta\big(r_{ij} - \|\mathbf{x}_i - \mathbf{x}_j\| - \delta_{ij}\big)$ with $h_\eta(x) = \begin{cases} \frac{1}{2}x^2, & \text{for } x \le \eta, \\ \eta\big(\|x\| - \frac{1}{2}\eta\big), & \text{otherwise.} \end{cases}$	Less sensitive to outliers

Table 5.1: A collection of well-known cost functions for localization purpose.

5.3.2 Bound-Constrained Regularized Formulation

In addition to the sparsity-promoting regularized formulation (5.3), the objective of estimating $\mathbf{X}$ and promoting the sparsity of $\boldsymbol{\delta}$ can be simultaneously achieved using an alternative formulation:

$$\begin{aligned} \min_{\mathbf{X},\, \boldsymbol{\delta}\ge 0} \quad & g(\boldsymbol{\delta}) \\ \text{subject to} \quad & \rho\big(\mathbf{X}, \boldsymbol{\delta}\big) \le \sigma_e. \end{aligned} \tag{5.4}$$

Here, the residual error term $\rho\big(\mathbf{X}, \boldsymbol{\delta}\big)$ is bounded by a predefined parameter σ_e. In this formulation, the constraint $\rho\big(\mathbf{X}, \boldsymbol{\delta}\big) \le \sigma_e$ plays the role of regularizing the optimization problem, and the objective function $g(\boldsymbol{\delta})$ promotes the sparsity of $\boldsymbol{\delta}$. Consequently, the regularization parameter γ is no longer needed, but the residual error bound σ_e becomes a new regularization parameter in the formulation of (5.4). We call this alternative generic form the *bound-constrained regularized formulation*.

5.3.3 Discussion

We remark that the formulations in (5.3) and (5.4) provide a general framework for NLOS effect mitigation by exploiting the sparsity of ranging biases. Different residual error functions and sparsity-promoting functions can be adopted, giving rise to different realizations of (5.3) and (5.4). Several examples of $\rho\big(\mathbf{X}, \boldsymbol{\delta}\big)$ are given in Table 5.1, being the well-known cost functions ever used for localization purposes. Depending on the objective, which could be, e.g., ending up with an optimization problem without local optima, a specific $\rho\big(\mathbf{X}, \boldsymbol{\delta}\big)$ is chosen or designed. Enumerating all possible realizations of (5.3) and (5.4) is out of the scope of this paper. Here, we only focus on developing one concrete realization of (5.3) and (5.4), to be elaborated in Section 5.4.

5.4 Semi-definite Programming Model

In this section, we will show one concrete example of realizing the generic formulations in (5.3) and (5.4). Location estimation using the formulations in (5.3) and (5.4) relies on solving an optimization problem, which may contain local optima. Finding the global optima of a high dimensional non-convex problem without an appropriate initial guess is quite challenging. Therefore, convex realizations of (5.3) and (5.4) are highly desired and serve as our target in this section. The difficulty, however, lies in that the optimization problems in (5.3) and (5.4) are non-convex for most reasonable choices of $\rho(\cdot)$. To overcome this difficulty, one common path is to design the cost function $\rho(\cdot)$ properly so that the resulting non-convex problem can be transformed into a convex one using relaxation methods. In this section, we follow this path and design the cost function $\rho(\cdot)$ as well as the sparsity-promoting function $g(\cdot)$ so that the resulting problem can be conveniently relaxed to an SDP. Aiming for an SDP relaxation method is driven by the following facts. Firstly, the SDP is a classical method for cooperative localization. Though there exist other relaxation methods, e.g., second-order-cone-programming relaxation in [Tse07], it is widely known that the semi-definite relaxation is tight and provides quite satisfying localization performance, hence, has been intensively pursued in the literature [CMS04, BLWY06, CWW$^+$12, SL14, VB15]. Secondly, the bound-constrained regularized formulation of the resulting SDP will become primal infeasible when σ_e is too small. This property can be exploited to develop a data-driven methodology for selecting an appropriate σ_e.

5.4.1 Sparsity-Promoting Regularized SDP

Based on the procedure in [JYF$^+$20a], we start with the following residual error function:

$$\begin{aligned}\rho(\mathbf{X},\boldsymbol{\delta}) &= \sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \left(\|\mathbf{x}_i-\mathbf{x}_j\|+\delta_{ij}\right)^2 \right| \\ &= \sum_{(i,j)\in\Gamma} \Big| r_{ij}^2 - \|\mathbf{x}_i-\mathbf{x}_j\|^2 - \underbrace{\left(2\|\mathbf{x}_i-\mathbf{x}_j\|\delta_{ij}+\delta_{ij}^2\right)}_{\text{transformed bias: } \epsilon_{ij}} \Big|.\end{aligned} \tag{5.5}$$

A careful examination of the transformed bias, $\epsilon_{ij} = 2\|\mathbf{x}_i-\mathbf{x}_j\|\delta_{ij}+\delta_{ij}^2$, shows that

- $\delta_{ij}=0$ gives rise to $\epsilon_{ij}=0$;
- $\delta_{ij}>0$ gives rise to $\epsilon_{ij}>0$.

This guarantees that the vector $\boldsymbol{\epsilon} = \{\epsilon_{ij}\}_{(i,j)\in\Gamma}$ possesses the same properties as $\boldsymbol{\delta}$, namely the sparsity property and the elementwise nonnegativity.

Following the same design principle of (5.3), the sparsity property of $\boldsymbol{\epsilon}$ is exploited and the following regularized optimization problem is formulated:

$$\min_{\mathbf{X},\,\boldsymbol{\epsilon}\geq 0} \sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \|\mathbf{x}_i - \mathbf{x}_j\|^2 - \epsilon_{ij} \right| + \gamma\, g(\boldsymbol{\epsilon}). \tag{5.6}$$

As for the sparsity-inducing function $g(\cdot)$, a common convex choice is the l_1 norm, i.e.,

$$g(\boldsymbol{\epsilon}) = \sum_{(i,j)\in\Gamma} |\epsilon_{ij}| = \sum_{(i,j)\in\Gamma} \epsilon_{ij}. \tag{5.7}$$

Here, the second equality is due to the nonnegativity of $\boldsymbol{\epsilon}$. Inserting (5.7) into (5.6) gives rise to

$$\min_{\mathbf{X},\,\boldsymbol{\epsilon}\geq 0} \sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \|\mathbf{x}_i - \mathbf{x}_j\|^2 - \epsilon_{ij} \right| + \gamma \sum_{(i,j)\in\Gamma} \epsilon_{ij}. \tag{5.8}$$

The resulting optimization problem in (5.8) is though non-convex, it can be conveniently related to an SDP by following the same procedure as in [BLWY06]. For completeness, we briefly repeat the relaxation procedure. First, Problem (5.8) is equivalently reformulated as:

$$\begin{aligned} \min_{\boldsymbol{\epsilon},\,\mathbf{Z}} \quad & \sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \mathrm{tr}\big(\mathbf{E}_{ij}\mathbf{Z}\big) - \epsilon_{ij} \right| + \gamma \sum_{(i,j)\in\Gamma} \epsilon_{ij} \\ \text{subject to} \quad & \epsilon_{ij} \geq 0, \quad \forall (i,j) \in \Gamma, \\ & \mathbf{Z} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{X} \\ \mathbf{X}^T & \mathbf{Y} \end{bmatrix}, \\ & \mathbf{Y} = \mathbf{X}^T\mathbf{X} \end{aligned} \tag{5.9}$$

with matrix $\mathbf{E}_{ij}$ defined as

$$\mathbf{E}_{ij} = \begin{cases} [0,0,\mathbf{e}_i^T - \mathbf{e}_j^T]^T\, [0,0,\mathbf{e}_i^T - \mathbf{e}_j^T],\ i,j \in S_u, \\ [\mathbf{x}_i^T, -\mathbf{e}_j^T]^T\, [\mathbf{x}_i^T, -\mathbf{e}_j^T],\ i \in S_a, j \in S_u. \end{cases} \tag{5.10}$$

Herein, $\mathrm{tr}(\cdot)$ denotes the trace of a matrix, $\mathbf{I}_2$ is the 2-dimensional identity matrix, and $\mathbf{e}_i$ is a vector with its i-th entry being one and the rest being zero. Next, the quadratic constraint $\mathbf{Y} = \mathbf{X}^T\mathbf{X}$ is relaxed to $\mathbf{Y} \succeq \mathbf{X}^T\mathbf{X}$, which means that $\mathbf{Y} - \mathbf{X}^T\mathbf{X}$ is positive semi-definite. The relaxed constraint is equivalent to $\mathbf{Z} \succeq \mathbf{0}$. Subsequently,

the non-convex problem in (5.9) can be relaxed to the following SDP:

$$\begin{aligned} \min_{\boldsymbol{\epsilon},\, \mathbf{Z}} \quad & \sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \operatorname{tr}\big(\mathbf{E}_{ij}\mathbf{Z}\big) - \epsilon_{ij} \right| + \gamma \sum_{(i,j)\in\Gamma} \epsilon_{ij} \\ \text{subject to} \quad & \epsilon_{ij} \geq 0, \quad \forall (i,j) \in \Gamma, \\ & \mathbf{Z} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{X} \\ \mathbf{X}^T & \mathbf{Y} \end{bmatrix} \succeq \mathbf{0}. \end{aligned} \tag{5.11}$$

The resulting problem has a similar form as the generic formulation of (5.3), and it is referred to as the *sparsity-promoting regularized SDP*. Compared with (5.3), the optimization problem in (5.11) contains two different optimization variables, $\boldsymbol{\epsilon}$ and $\mathbf{Z}$. The position variable $\mathbf{X}$ is contained in $\mathbf{Z}$. The variable $\boldsymbol{\delta}$ is replaced by $\boldsymbol{\epsilon}$, which inherits the same properties as $\boldsymbol{\delta}$.

5.4.2 Bound-Constrained Regularized SDP

Similar to the generic bound-constrained regularized formulation, the residual error term $\sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \|\mathbf{x}_i - \mathbf{x}_j\|^2 - \epsilon_{ij} \right|$ in (5.8) can be bounded by a predefined parameter. This gives rise to the following formulation:

$$\begin{aligned} \min_{\mathbf{X},\, \boldsymbol{\epsilon}\geq 0} \quad & \sum_{(i,j)\in\Gamma} \epsilon_{ij} \\ \text{subject to} \quad & \sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \|\mathbf{x}_i - \mathbf{x}_j\|^2 - \epsilon_{ij} \right| \leq \sigma_e. \end{aligned} \tag{5.12}$$

Adopting the same relaxation procedure, Problem (5.12) can be relaxed to the following SDP:

$$\begin{aligned} \min_{\boldsymbol{\epsilon},\, \mathbf{Z}} \quad & \sum_{(i,j)\in\Gamma} \epsilon_{ij} \\ \text{subject to} \quad & \epsilon_{ij} \geq 0, \quad \forall (i,j) \in \Gamma, \\ & \mathbf{Z} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{X} \\ \mathbf{X}^T & \mathbf{Y} \end{bmatrix} \succeq \mathbf{0}, \\ & \sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \operatorname{tr}\big(\mathbf{E}_{ij}\mathbf{Z}\big) - \epsilon_{ij} \right| \leq \sigma_e. \end{aligned} \tag{5.13}$$

The resulting problem has a similar form as the generic formulation of (5.4), and it is referred to as the *bound-constrained regularized SDP*. A detailed analysis of the relation between (5.11) and (5.13) as well as two special cases is given in the following section.

5.4.3 Analysis of the Convex Approximations

5.4.3.1 Equivalence of Problems (5.11) and (5.13)

A comparison of (5.11) and (5.13) shows that both formulations are the same except for one difference: The residual error, $\sum_{(i,j)\in\Gamma}\left|r_{ij}^2-\mathrm{tr}\big(\mathbf{E}_{ij}\mathbf{Z}\big)-\epsilon_{ij}\right|$, is one term of the objective function in (5.11), but it is bounded as one constraint in (5.13). One question arises naturally: what is the relation between these two formulations? To answer this question, we will introduce several conditions and three theorems which state the relation of these two formulations.

Condition 1): The minima of Problems (5.11) and (5.13) exist, and Slater's condition is fulfilled for both problems.

Condition 2): The inequality constraint $\sum_{(i,j)\in\Gamma}\left|r_{ij}^2-\mathrm{tr}\big(\mathbf{E}_{ij}\mathbf{Z}\big)-\epsilon_{ij}\right|\leq\sigma_e$ is *strictly* active in Problem (5.13), meaning that the optimal KKT multiplier associated with this inequality constraint is non-zero.

Condition 3): The KKT multiplier associated with the inequality constraint $\sum_{(i,j)\in\Gamma}\left|r_{ij}^2-\mathrm{tr}\big(\mathbf{E}_{ij}\mathbf{Z}\big)-\epsilon_{ij}\right|\leq\sigma_e$ is unique.

Condition 4): For any $A>0$, let $\boldsymbol{\epsilon}^*$ and $\mathbf{Z}^*$ be the optimal solution of Problem (5.11) with $\gamma=A$. The value of the summation $\sum_{(i,j)\in\Gamma}\left|r_{ij}^2-\mathrm{tr}\big(\mathbf{E}_{ij}\mathbf{Z}^*\big)-\epsilon_{ij}^*\right|$ is unique.

Theorem 2 *Let $\boldsymbol{\epsilon}^*$ and $\mathbf{Z}^*$ be the optimal solution of Problem* (5.13)*. Under Conditions 1) and 2), $\boldsymbol{\epsilon}^*$ and $\mathbf{Z}^*$ must be the optimal solution of Problem* (5.11) *with $\gamma=\frac{1}{\lambda_\sigma^*}$, where λ_σ^* is the KKT multiplier associated with the inequality constraint $\sum_{(i,j)\in\Gamma}\left|r_{ij}^2-tr\big(\mathbf{E}_{ij}\mathbf{Z}\big)-\epsilon_{ij}\right|\leq\sigma_e$.*

Proof: See Appendix A.3. □

Theorem 3 *Let $\boldsymbol{\epsilon}^*$ and $\mathbf{Z}^*$ be the optimal solution of Problem* (5.11)*. Under Conditions 1) and 2), $\boldsymbol{\epsilon}^*$ and $\mathbf{Z}^*$ must be the optimal solution of Problem* (5.13) *with $\sigma_e=\sum_{(i,j)\in\Gamma}\left|r_{ij}^2-tr\big(\mathbf{E}_{ij}\mathbf{Z}^*\big)-\epsilon_{ij}^*\right|$.*

Proof: See Appendix A.4. □

Theorem 4 *Under Conditions 1) - 4), for each σ_e, there exists one γ such that the optimal solutions of both problems are equivalent, and for each γ, there exists one σ_e such that the optimal solutions of both problems are equivalent.*

Proof: See Appendix A.5. □

According to Theorem 4, under Conditions 1) - 4), both regularized problems in (5.11) and (5.13) are equivalent in the sense that they have the same optimal solution.

Remark: We stress that this equivalence relation holds only for (5.11) and (5.13) but not for the two generic formulations (5.3) and (5.4) or for (5.8) and (5.12). This is because the equivalence requires another condition: both formulations must be convex problems, and neither (5.8) and (5.12) nor (5.3) and (5.4) are convex in general.

5.4.3.2 Special case: infinite γ and σ_e

As both parameters γ and σ_e approach infinity, Problems (5.11) and (5.13) simplify to the following optimization problem:

$$\begin{aligned} \min_{\boldsymbol{\epsilon},\, \mathbf{Z}} \quad & \sum_{(i,j)\in\Gamma} \epsilon_{ij} \\ \text{subject to} \quad & \epsilon_{ij} \geq 0, \quad \forall (i,j) \in \Gamma, \\ & \mathbf{Z} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{X} \\ \mathbf{X}^T & \mathbf{Y} \end{bmatrix} \succeq \mathbf{0}, \end{aligned} \tag{5.14}$$

which is equivalent to

$$\begin{aligned} \text{find} \quad & \mathbf{Z} \\ \text{subject to} \quad & \mathbf{Z} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{X} \\ \mathbf{X}^T & \mathbf{Y} \end{bmatrix} \succeq \mathbf{0}. \end{aligned} \tag{5.15}$$

Apparently, when a positive semi-definite matrix $\mathbf{Z}$ fulfills that $\mathbf{Z}_{[1:2,1:2]} = \mathbf{I}_2$ with $\mathbf{Z}_{[1:2,1:2]}$ denoting the submatrix of $\mathbf{Z}$ constructed by its first two rows and first columns, it must be an optimal solution of Problem (5.15). Hence, the optimization problem in (5.15) is meaningless for localization purposes.

5.4.3.3 Special case: inactive inequality constraint

Theorem 5 *Let $\boldsymbol{\epsilon}^*$ and $\mathbf{Z}^*$ be the optimal solution of Problem* (5.13)*. If the inequality constraint $\sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - tr\big(\mathbf{E}_{ij}\mathbf{Z}\big) - \epsilon_{ij} \right| \leq \sigma_e$ is inactive at $\boldsymbol{\epsilon} = \boldsymbol{\epsilon}^*$ and $\mathbf{Z} = \mathbf{Z}^*$, under Condition 1), it holds that $\boldsymbol{\epsilon}^* = \mathbf{0}_{|\Gamma|}$.*

Here, $|\Gamma|$ is the cardinality of Γ and $\mathbf{0}_{|\Gamma|}$ denotes the $|\Gamma| \times 1$ vector whose entries are all equal to zero.

Proof: See Appendix A.6. □

Corollary 1 *If the inequality constraint* $\sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - tr\big(\mathbf{E}_{ij}\mathbf{Z}\big) - \epsilon_{ij} \right| \leq \sigma_e$ *is inactive at the optimal solution,* $\boldsymbol{\epsilon} = \boldsymbol{\epsilon}^*$, *and* $\mathbf{Z} = \mathbf{Z}^*$, *Problem* (5.13) *is equivalent to the following optimization problem:*

$$\begin{aligned} \textit{find} \quad & \mathbf{Z} \\ \textit{subject to} \quad & \mathbf{Z} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{X} \\ \mathbf{X}^T & \mathbf{Y} \end{bmatrix} \succeq \mathbf{0}, \\ & \sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - tr\big(\mathbf{E}_{ij}\mathbf{Z}\big) \right| \leq \sigma_e. \end{aligned} \tag{5.16}$$

Proof: See Appendix A.7. □

This tells us that when the inequality constraint is inactive, Problem (5.13) simplifies to Problem (5.16). This happens when there exists a matrix, $\mathbf{Z}$, that fulfills the constraints in (5.16). As σ_e increases, it becomes increasingly likely that the feasible area of Problem (5.16) is nonempty. In a pure LOS scenario, this may occur for a moderate σ_e, while for a mixed LOS/NLOS scenario, this may occur only for a sufficiently large σ_e. It is remarkable that the inequality constraint $\sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \mathrm{tr}\big(\mathbf{E}_{ij}\mathbf{Z}\big) - \epsilon_{ij} \right| \leq \sigma_e$ cannot be ignored even if it is inactive. Without the inequality constraint, Problem (5.13) becomes Problem (5.15), which is a feasibility problem. It is apparent that the feasible area of Problem (5.15) is significantly larger than that of Problem (5.16). Hence, the inequality constraint $\sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \mathrm{tr}\big(\mathbf{E}_{ij}\mathbf{Z}\big) \right| \leq \sigma_e$ is significant in Problem (5.16) because it limits the feasible area.

5.4.4 Complexity Analysis

Here, we analyze the computational complexity of the proposed methods. Both (5.11) and (5.13) are SDPs and can be solved in polynomial time by a standard interior point method. It is shown in [BTN01] that the worst-case computational complexity for solving an SDP is on the order of

$$\left(N_{\mathrm{var}}^3 + N_{\mathrm{var}}^2 \sum_{i=1}^{N_{\mathrm{SD}}} (k_i)^2 + N_{\mathrm{var}} \sum_{i=1}^{N_{\mathrm{SD}}} (k_i)^3 \right) \cdot B(\mathcal{K}) \cdot \ln(1/\omega), \tag{5.17}$$

where N_{var} is the number of variables, N_{SD} is the number of semi-definite cone constraints, k_i is the dimension of the i-th semi-definite cone, $\omega > 0$ is the solution precision, and $B(\mathcal{K})$ is the so-called barrier parameter associated with cone $\mathcal{K}$, measuring the geometric complexity of cone $\mathcal{K}$. For an SDP, we have $B(\mathcal{K}) = \sum_{i=1}^{N_{\text{SD}}} k_i$.

To obtain the worst-case complexity of the SDP in (5.11), we first need to convert it into a standard form without taking the absolute value:

$$\begin{aligned}
\min_{\boldsymbol{\epsilon},\, \boldsymbol{\beta}^+,\, \boldsymbol{\beta}^-,\, \mathbf{Z}} \quad & \sum_{(i,j)\in\Gamma} \left(\beta_{ij}^+ + \beta_{ij}^-\right) + \gamma \sum_{(i,j)\in\Gamma} \epsilon_{ij} \\
\text{subject to} \quad & \epsilon_{ij} \geq 0,\ \beta_{ij}^+ \geq 0,\ \beta_{ij}^- \geq 0, \quad \forall (i,j) \in \Gamma, && \text{(5.18a)} \\
& \mathbf{Z} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{X} \\ \mathbf{X}^T & \mathbf{Y} \end{bmatrix} \succeq \mathbf{0}, && \text{(5.18b)} \\
& \epsilon_{ij} + \beta_{ij}^+ - \beta_{ij}^- + \operatorname{tr}\big(\mathbf{E}_{ij}\, \mathbf{Z}\big) = r_{ij}^2,\ \forall (i,j) \in \Gamma. && \text{(5.18c)}
\end{aligned}$$

In (5.18), there are $3|\Gamma| + \binom{2N_u+2}{2}$ variables, $3|\Gamma|$ one dimensional semi-definite cones associated with (5.18a)[2], one semi-definite cone of size $2N_u + 2$ associated with (5.18b) and $2|\Gamma|$ one dimensional semi-definite cones associated with (5.18c). Hence, the worst-case computational complexity of (5.11) is on the order of

$$N_{\text{var}} \cdot \left(N_{\text{var}}^2 + N_{\text{var}} \left(5|\Gamma| + (2N_u+2)^2 \right) + \left(5|\Gamma| + (2N_u+2)^3 \right) \right) \cdot B_1(\mathcal{K}) \cdot \ln(1/\omega) \quad \text{(5.19)}$$

with $N_{\text{var}} = 3|\Gamma| + \binom{2N_u+2}{2}$ and $B_1(\mathcal{K}) = 5|\Gamma| + 2N_u + 2$. Following the same procedure, the worst-case complexity of (5.13) is on the order of

$$N_{\text{var}} \cdot \left(N_{\text{var}}^2 + N_{\text{var}} \left(5|\Gamma| + (2N_u+2)^2 + 1 \right) + \left(5|\Gamma| + (2N_u+2)^3 + 1 \right) \right) \cdot B_2(\mathcal{K}) \cdot \ln(1/\omega) \quad \text{(5.20)}$$

with N_{var} being defined as above and $B_2(\mathcal{K}) = 5|\Gamma| + 2N_u + 3$. This analysis shows that both (5.11) and (5.13) have comparable worst-case complexity. Lastly, we note that the complexity bounds in (5.19) and (5.20) are in the worst-case sense and both (5.11) and (5.13) can be solved more efficiently by exploiting certain structure of the constraint matrices.

5.5 Regularization Parameter Selection Strategy

In this section, we consider the question: how to determine an appropriate regularization parameter γ or σ_e in (5.11) or (5.13), respectively. We first analyze the influence

[2]The linear constraint $\mathbf{a}^T\mathbf{x} - b \geq 0$ is equivalent to the one dimensional linear matrix inequality constraint $\mathbf{a}^T\mathbf{x} - b \in \mathcal{S}_+^1$, where $\mathbf{x}$ is the optimization variable, $\mathbf{a}$ and b are the given constraint coefficients.

of both parameters on the localization accuracy. The analysis below shows that the most appropriate regularization parameter, γ or σ_e, depends on the underlying localization environment and sticking to a fixed parameter may lead to certain performance loss. This motivates us to develop a reliable data-driven method that dynamically determines an appropriate parameter for the specific localization environment.

5.5.1 Effect of γ and σ_e

The regularization parameter γ controls how strong the sparsity-inducing term $\sum_{(i,j)\in\Gamma} \epsilon_{ij}$ is penalized. If γ is too large, ϵ_{ij} associated with an NLOS link may be reduced to zero. On the other hand, if γ is too small, ϵ_{ij} associated with an LOS link may become non-zero. In fact, the most appropriate γ depends on the specific propagation environment, e.g., the percentage of NLOS links (P_{NLOS}) and the statistics of v_{ij} and δ_{ij}. An example is given in Fig. 5.1. The localization accuracy is evaluated in terms of the root-mean-squared-error (RMSE). Here, we focus on interpreting Fig. 5.1 and postpone the definition of RMSE and the simulation setup to Section 5.6. In a pure LOS scenario, i.e., $P_{\text{NLOS}} = 0\%$, $\gamma \geq 0.1$ provides the highest localization accuracy, while for the scenario of $P_{\text{NLOS}} = 60\%$, $\gamma \leq 0.01$ is more appropriate. The value of $\gamma = 0.1$ leads to 11% performance loss for $P_{\text{NLOS}} = 60\%$. Although this quantity is not immense, it tells us that a fixed γ cannot be the most appropriate parameter for all localization scenarios. Next, we investigate the influence of σ_e on the localization

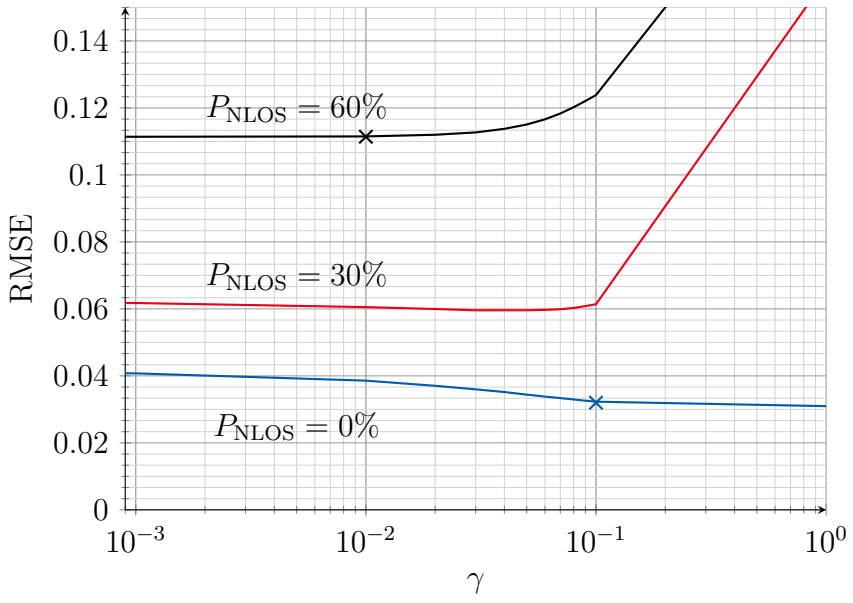
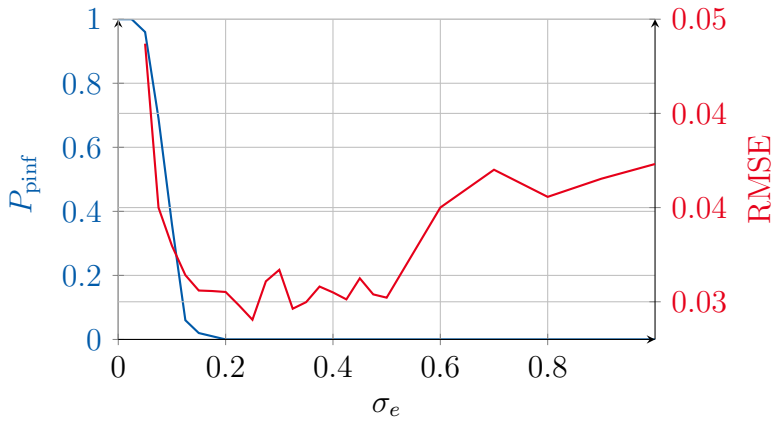
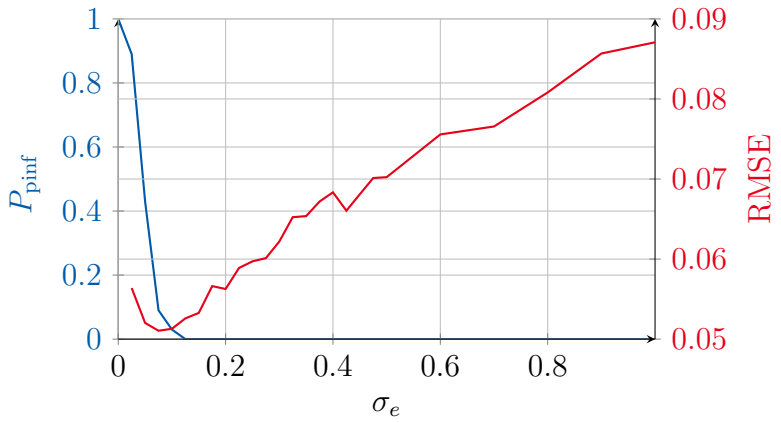


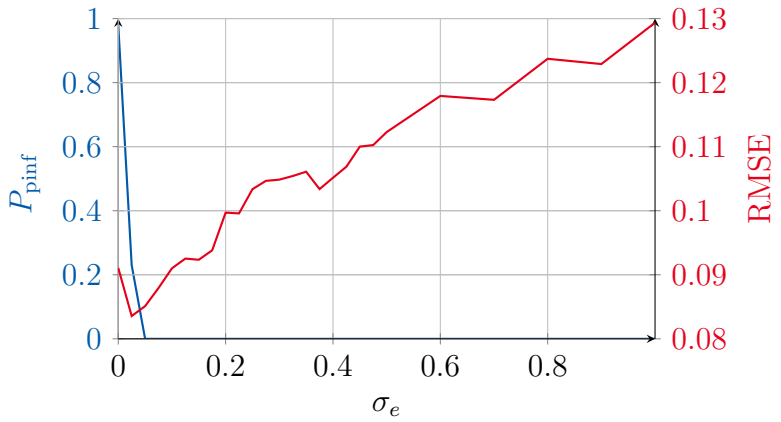
Figure 5.1: Influence of γ on localization accuracy when $v_{ij} \sim \mathcal{N}(0, 0.01^2)$ and $\delta_{ij} \sim \mathcal{U}[0, 0.5]$.



(a) $P_{\text{NLOS}} = 10\%$



(b) $P_{\text{NLOS}} = 20\%$



(c) $P_{\text{NLOS}} = 40\%$

Figure 5.2: Influence of σ_e on probability of primal infeasibility (P_{pinf}) and localization accuracy in terms of RMSE at $v_{ij} \sim \mathcal{N}(0, 0.01^2)$ and $\delta_{ij} \sim \mathcal{U}[0, 1]$.

accuracy as well as the probability of primal infeasibility (P_{pinf}). Figure 5.2 depicts one representative result, where the blue curve, associated with the left y-axis, represents the probability of primal infeasibility P_{pinf} and the red curve, associated with the right y-axis, represents the localization accuracy in terms of the RMSE. All primal infeasible cases do not contribute to the calculation of the RMSE. There are several observations related to Fig. 5.2. Firstly, the probability of primal infeasibility, P_{pinf}, decreases monotonically with an increasing σ_e. This result is expected because if σ_e is too small, there may not exist a $\mathbf{Z}$ that fulfills all the constraints and Problem (5.13) becomes primal infeasible. Secondly, roughly speaking, the RMSE curves rise with an increasing σ_e. A comparison of all three figures shows that the value of σ_e, at which the lowest RMSE is achieved, declines as P_{NLOS} increases. This tells us that, similar to γ, the most appropriate σ_e depends on the specific scenario, e.g., the underlying P_{NLOS}. In addition to these three scenarios, we have examined various other scenarios: P_{NLOS} varying from 0% to 100% with different distributions of v_{ij} and δ_{ij}. For all tested scenarios, the P_{pinf} curve and the RMSE curve demonstrate similar behaviors.

5.5.2 Data-Driven Parameter Selection Strategy

We have seen that the most appropriate γ or σ_e varies with the localization scenarios, e.g., P_{NLOS}, which is unknown in practice. This motivates us to develop a reliable data-driven parameter selection method, either for γ or for σ_e. By examining Fig. 5.2 more carefully, we have an interesting observation: the RMSE curve first shows a short decreasing tendency and then starts to increase with σ_e. This short declining trend implies that the boundary value $\sigma_e^{\min}$ is too tight and a slight loosening of σ_e achieves improved localization accuracy. Here, $\sigma_e^{\min}$ denotes the minimal value of σ_e, for which Problem (5.13) is primal feasible. Inspired by this observation, we propose to use the boundary value $\sigma_e^{\min}$ as a reference value and set $\sigma_e = c \cdot \sigma_e^{\min}$ with $c \geq 1$ being a user-defined constant.

This gives rise to the following data-driven parameter selection strategy. First, we determine the lower bound, $\sigma_e^{\min}$, by solving:

$$\begin{aligned} \sigma_e^{\min} = \min_{\sigma_e, \boldsymbol{\epsilon} \geq 0, \mathbf{Z}} \quad & \sigma_e \\ \text{subject to} \quad & \mathbf{Z} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{X} \\ \mathbf{X}^T & \mathbf{Y} \end{bmatrix} \succeq \mathbf{0}, \\ & \sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \operatorname{tr}\big(\mathbf{E}_{ij}\,\mathbf{Z}\big) - \epsilon_{ij} \right| \leq \sigma_e. \end{aligned} \tag{5.21}$$

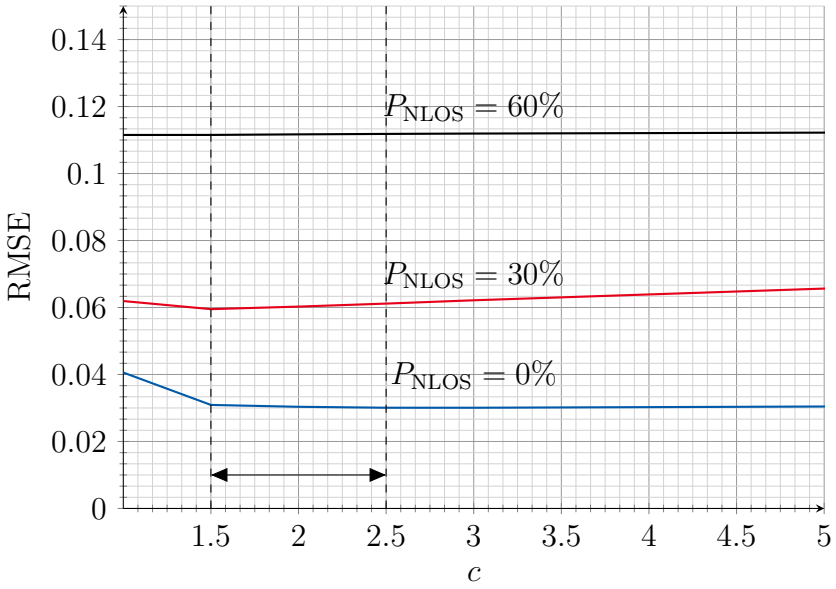


Figure 5.3: Influence of c on the localization accuracy obtained with $v_{ij} \sim \mathcal{N}(0, 0.01^2)$ and $\delta_{ij} \sim \mathcal{U}[0, 0.5]$.

Here, $\sigma_e^{\min}$ is the global optimum of (5.21). Different from (5.13), σ_e is an optimization variable in (5.21) while not a predefined constant. Then, based on $\sigma_e^{\min}$, the position estimates are obtained by solving (5.13) with $\sigma_e = c \cdot \sigma_e^{\min}$:

$$\begin{aligned} \min_{\boldsymbol{\epsilon} \geq 0,\, \mathbf{Z}} \quad & \sum_{(i,j)\in\Gamma} \epsilon_{ij} \\ \text{subject to} \quad & \mathbf{Z} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{X} \\ \mathbf{X}^T & \mathbf{Y} \end{bmatrix} \succeq \mathbf{0}, \\ & \sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \operatorname{tr}\left(\mathbf{E}_{ij}\, \mathbf{Z}\right) - \epsilon_{ij} \right| \leq c \cdot \sigma_e^{\min}. \end{aligned} \tag{5.22}$$

Here, $c \geq 1$ is a user-defined parameter. To illustrate the influence of c on the localization accuracy, we depict a typical case in Fig. 5.3. This figure shows that for LOS-dominating scenarios, enlarging c to certain extent can help to improve the localization accuracy. On the other hand, the performance for NLOS-heavy scenarios, e.g., $P_{\text{NLOS}} = 60\%$, remains almost unchanged. Besides, the localization accuracy is not sensitive to c. In this example, any c in the range $[1.5, 2.5]$ is a suitable choice.

We have two remarks on the proposed data-driven parameter selection strategy. First, this parameter selection strategy makes use of the fact that Problem (5.13) is not always primal feasible. This fact enables us to dynamically obtain a boundary value, $\sigma_e^{\min}$, for each localization problem. Then, as an estimate of the most appropriate parameter, we use $\sigma_e = c \cdot \sigma_e^{\min}$. Second, σ_e is determined at the cost of a slight increase in computational complexity, i.e., the computational cost of solving Problem (5.21). Compared

with most existing parameter selection methods, which require solving Problems (5.11) and (5.13) on a grid of γ and σ_e, respectively, this additional computational cost for solving Problem (5.21) once is almost negligible.

5.6 Numerical Results

In this section, the performance of the proposed methods is evaluated in a variety of scenarios. Following [BLWY06], we consider a 1×1 area in the normalized unit-free scale. Note that the noise level and the localization accuracy should be scaled accordingly to obtain the corresponding physical values. For instance, a standard deviation of 0.01 in the normalized unit-free sense corresponds to 1 m for an actual area of $100\,\mathrm{m} \times 100\,\mathrm{m}$. Each network consists of 100 agents and $N_a = 4$ or $N_a = 8$ anchors. For all Monte Carlo runs, the anchor positions are fixed. The agent positions are randomly and uniformly generated in each Monte Carlo run. We set the communication range to 0.2 unless mentioned otherwise. For the measurement error, v_{ij}, different distributions are considered, including a zero-mean Gaussian distribution $\mathcal{N}(0, \sigma_{ij}^2)$ with homogeneous variance $\sigma_{ij}^2 = \sigma_{\mathrm{LOS}}^2$ or inhomogeneous distance-related variance $\sigma_{ij}^2 = \kappa^2 \|\mathbf{x}_i - \mathbf{x}_j\|^2$ and the Laplace distribution. The NLOS bias δ_{ij} is randomly generated from a uniform distribution, $\mathcal{U}[a, b]$. The localization accuracy is evaluated in terms of the RMSE, which is defined as

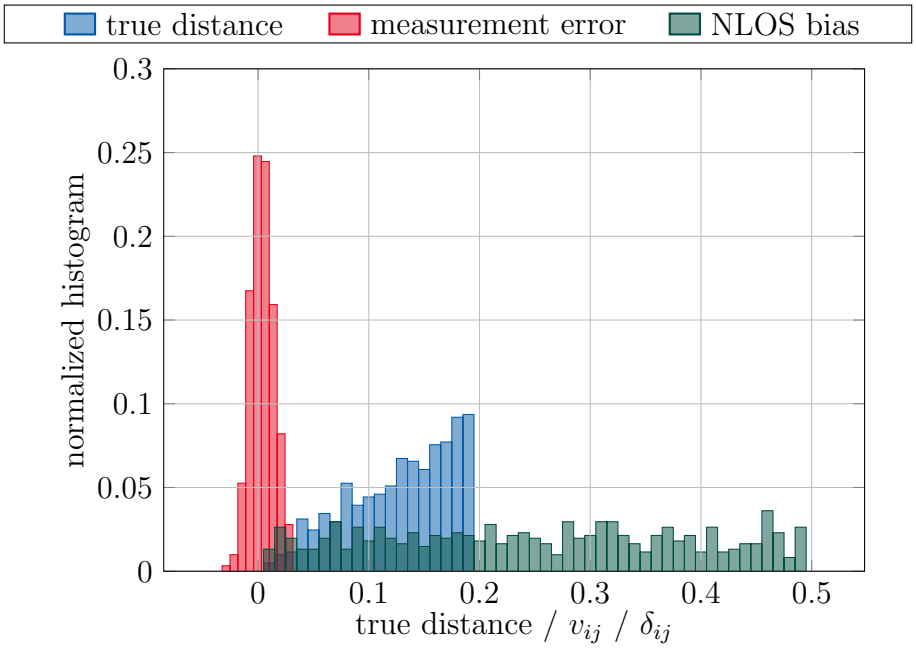
$$\mathrm{RMSE} \triangleq \sqrt{\frac{1}{N_{\mathrm{MC}}} \frac{1}{N_u} \sum_{j=1}^{N_{\mathrm{MC}}} \sum_{i=1}^{N_u} \left(\hat{\mathbf{x}}_i^j - \mathbf{x}_i^j\right)^T \left(\hat{\mathbf{x}}_i^j - \mathbf{x}_i^j\right)},$$

where $\mathbf{x}_i^j$ and $\hat{\mathbf{x}}_i^j$ are the location of node i and its estimate in the j-th Monte Carlo run, respectively, and N_{MC} is the total number of Monte Carlo runs. The performance of all methods is evaluated over $N_{\mathrm{MC}} = 500$ Monte Carlo runs.

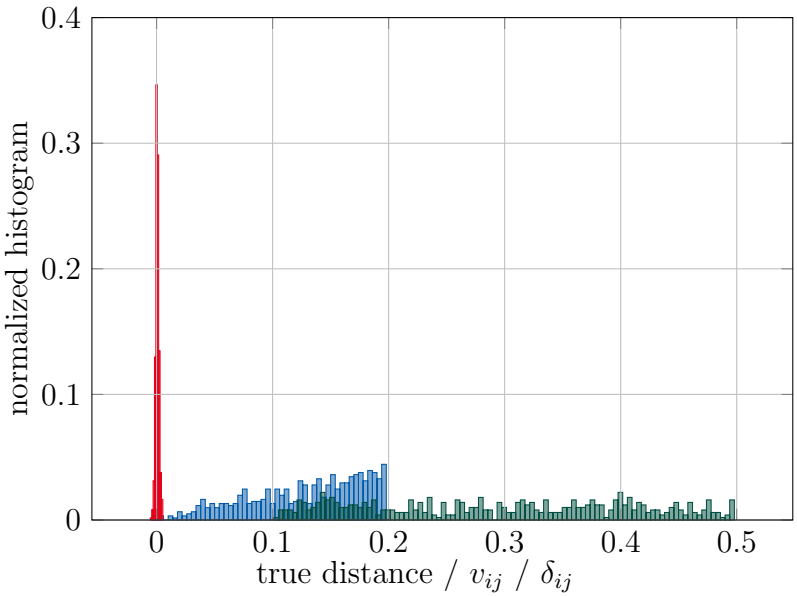
The MATLAB code that implements the proposed two SDPs and the data-driven strategy is available at: `https://gitlab.com/jindi.et/sdp_sparsity.git`.

5.6.1 Simulation Results: Regularization Parameter Selection

In this section, we investigate the influence of the regularization parameter (γ and σ_e) on localization performance and the efficacy of the proposed data-driven strategy (5.22). For this purpose, four different configurations are examined, including the proposed sparsity-promoting regularized SDP (5.11) with three representative candidates



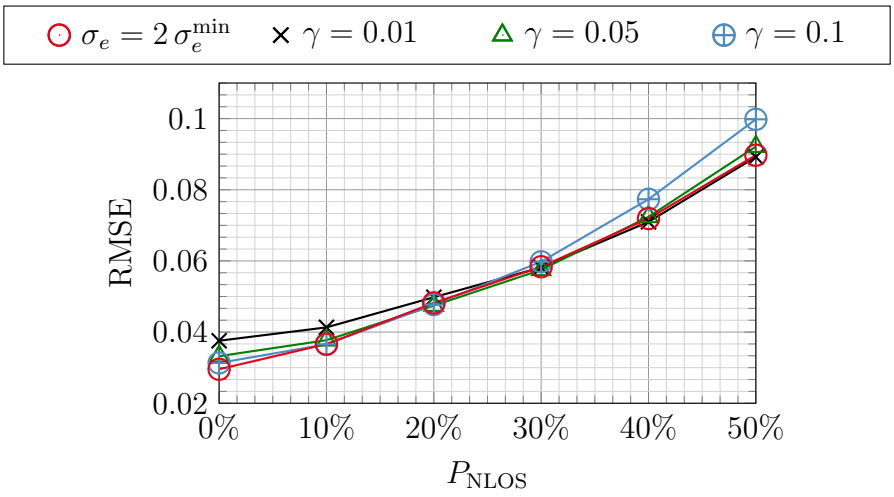
(a)



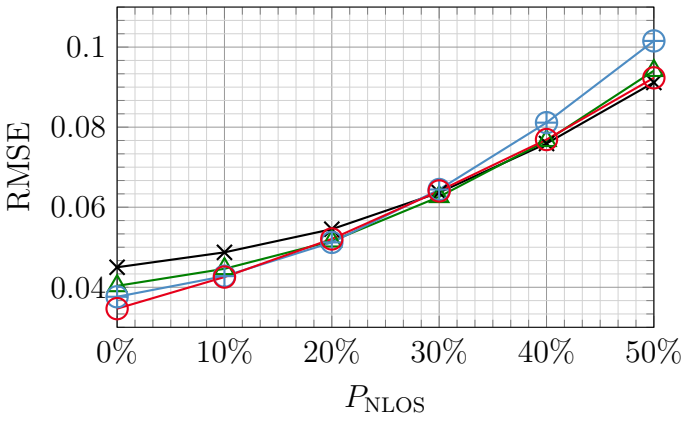
(b)

Figure 5.4: Histogram of true distance, measurement error of v_{ij} and NLOS biases δ_{ij}: (a) $v_{ij} \sim \mathcal{N}(0, 0.01^2)$, $\delta_{ij} \sim \mathcal{U}[0, 0.5]$ (b) $v_{ij} \sim \mathcal{N}(0, 0.01^2 \|\mathbf{x}_i - \mathbf{x}_j\|)$, $\delta_{ij} \sim \mathcal{U}[0.1, 0.5]$.

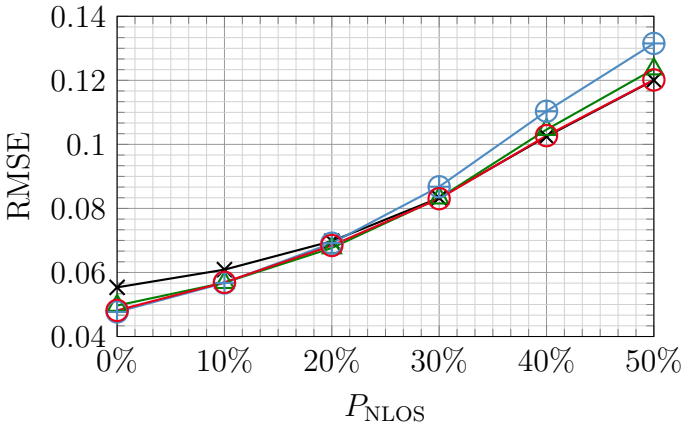
($\gamma = 0.01, 0.05, 0.1$) and the proposed data-driven strategy (5.22) with $c = 2$, i.e., $\sigma_e = 2\,\sigma_e^{\min}$. Note that these three values of γ are sensibly chosen and they provide satisfying localization accuracy.



(a) $v_{ij} \sim \mathcal{N}(0, 0.01^2)$ and $N_a = 8$

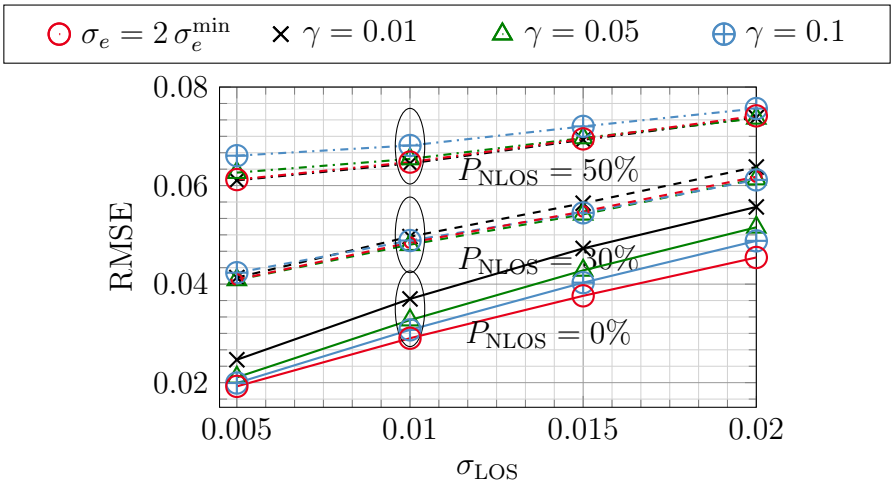


(b) $v_{ij} \sim \text{Laplace}(0, 0.01)$ and $N_a = 8$

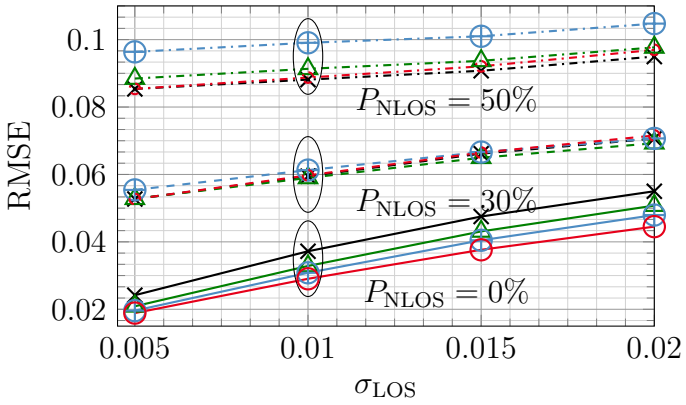


(c) $v_{ij} \sim \mathcal{N}(0, 0.01^2)$ and $N_a = 4$

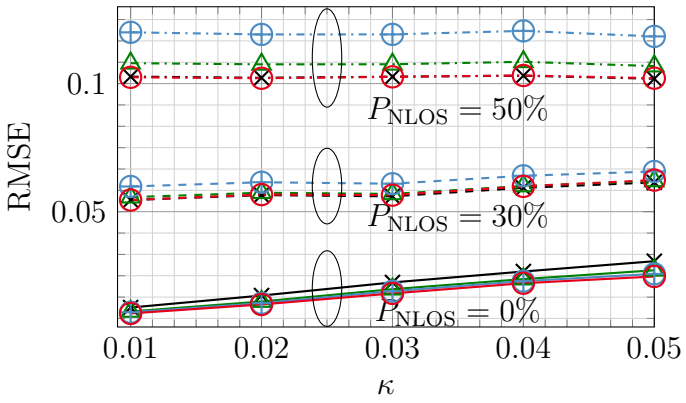
Figure 5.5: Localization accuracy of proposed methods at different values of P_{NLOS} with $\delta_{ij} \sim \mathcal{U}[0, 0.5]$.



(a) $v_{ij} \sim \mathcal{N}(0, \sigma^2_{\text{LOS}})$, $\delta_{ij} \sim \mathcal{U}[0, 0.2]$



(b) $v_{ij} \sim \mathcal{N}(0, \sigma^2_{\text{LOS}})$, $\delta_{ij} \sim \mathcal{U}[0, 0.5]$



(c) $v_{ij} \sim \mathcal{N}(0, \kappa^2 d^2_{ij})$, $\delta_{ij} \sim \mathcal{U}[0.1, 0.5]$

Figure 5.6: Localization accuracy of proposed methods at different variances of measurement error v_{ij} with $N_a = 8$.

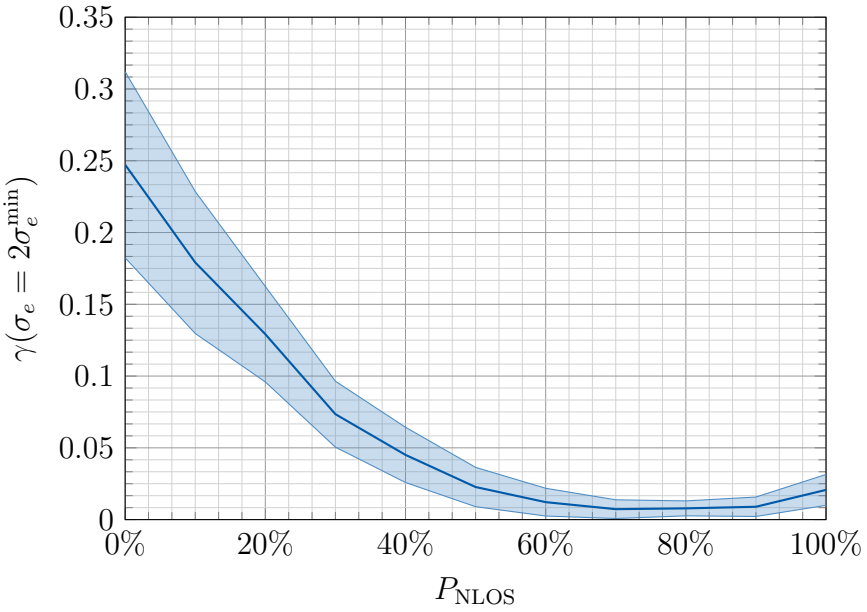


Figure 5.7: Equivalent γ of $\sigma_e = 2\,\sigma_e^{\min}$ at $v_{ij} \sim \mathcal{N}(0, 0.01^2),\ \delta_{ij} \sim \mathcal{U}[0, 0.5]$.

To mimic various practical localization scenarios, different measurement error distributions, e.g., $v_{ij} \sim \mathcal{N}(0, 0.01^2)$ or $v_{ij} \sim \text{Laplace}(0, 0.01)$, and different network layouts ($N_a = 4$ or $N_a = 8$) are considered. To illustrate the noise statistics, we depict the normalized histograms of v_{ij}, δ_{ij} versus the true distance in Fig. 5.4. The statistics of the LOS/NLOS measurement errors in Fig. 5.4-(a) corresponds to the scenarios in which the measurement error is moderate relative to the true distance. On the other hand, Fig. 5.4-(b) reflects a low noise scenario, which is typical for ultra-wideband technology [DCF+09].

Figure 5.5 compares the localization performance of all methods in terms of RMSE at different values of P_{NLOS}, where, for instance, $P_{\text{NLOS}} = 10\%$ corresponds to an LOS-heavy case and $P_{\text{NLOS}} = 90\%$ to an NLOS-heavy case. It can be seen that these four configurations perform similarly in these three scenarios and the performance difference of these four configurations is moderate. As expected, the localization performance of all methods degrade with an increasing P_{NLOS}, due to the adverse effect of NLOS measurements on localization accuracy. A closer inspection shows that for the proposed sparsity-promoting regularized SDP in (5.11), a fixed γ fails to achieve the highest localization accuracy in all scenarios. At a low P_{NLOS} ($0 \leq P_{\text{NLOS}} \leq 30\%$), $\gamma = 0.1$ seems to be the most appropriate parameter. However, for $P_{\text{NLOS}} > 30\%$, the highest localization accuracy is provided by the lower value, $\gamma = 0.01$. Among these four configurations, the proposed data-driven strategy keeps attaining the lowest localization error at different P_{NLOS} in all scenarios. Taking Fig. 5.5-(b) as an example, with respect to the red curve ($\sigma_e = 2\,\sigma_e^{\min}$), the black curve ($\gamma = 0.01$) leads to a performance loss

around 31% when $P_{\mathrm{NLOS}} = 0\%$, and the blue curve ($\gamma = 0.1$) leads to a performance loss around 9% when $P_{\mathrm{NLOS}} = 50\%$. These results demonstrate the efficacy of the proposed data-driven strategy for choosing an appropriate σ_e. It is important to point out that this parameter selection strategy is not sensitive to c and the localization performance remains comparable with c in the range of $[1.5, 2.5]$.

To make the comparison more comprehensive, the localization scenarios are further extended and the proposed methods are evaluated at varying noise variance for different combinations of LOS/NLOS distributions. The results are depicted in Fig. 5.6. Here, we have similar observations as in Fig. 5.5. Among all four curves, $\sigma_e = 2\,\sigma_e^{\min}$ achieves the lowest RMSE in almost all tested scenarios. This result again shows the ability of the proposed data-driven strategy in providing an appropriate σ_e. By contrast, a fixed regularization parameter, γ, fails to achieve the highest localization accuracy when the statistics of v_{ij} and δ_{ij} vary or when P_{NLOS} changes.

Next, we aim to analyze the remarkable localization performance of the proposed data-driven parameter selection strategy. Recall that Theorem 4 states that under certain conditions, there exists for each σ_e one γ such that Problem (5.11) and Problem (5.13) have equivalent optimal solutions. This implies that for one localization problem, each σ_e has an equivalent γ. As shown in the proof of Theorem 4, this equivalent value is $\gamma = \frac{1}{\lambda_\sigma^*}$ with λ_σ^* being the optimal KKT multiplier associated with the inequality constraint $\sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \mathrm{tr}\big(\mathbf{E}_{ij}\,\mathbf{Z}\big) - \epsilon_{ij} \right| \leq \sigma_e$. Relying on this theorem, the equivalent γ of $\sigma_e = c \cdot \sigma_e^{\min}$, denoted by $\gamma(\sigma_e = c \cdot \sigma_e^{\min})$, can be numerically calculated. Figure 5.7 depicts the numerical results of $\gamma(\sigma_e = 2\,\sigma_e^{\min})$ in terms of its sample mean and its sample standard deviation. It is shown that at a low P_{NLOS}, the equivalent γ of $\sigma_e = 2\,\sigma_e^{\min}$ is relatively high. As P_{NLOS} increases, the equivalent γ decreases gradually. This trend is in agreement with the requirement of an appropriate γ at varying P_{NLOS}: a large γ is more suitable for an LOS-heavy scenario and a small γ is more appropriate for an NLOS-heavy scenario. This analysis reflects that the proposed parameter selection strategy is able to dynamically adjust its sparsity weightings. This ability is attributed to the fact that other than using one fixed parameter σ_e, the proposed parameter selection strategy dynamically determines one σ_e; more precisely, it determines a unique $\sigma_e^{\min}$ for each localization problem.

5.6.2 Simulation Results: Comparison of Localization Performance

In this section, we compare the localization performance of the proposed framework with several state-of-the-art methods, including the centralized expectation-conditional

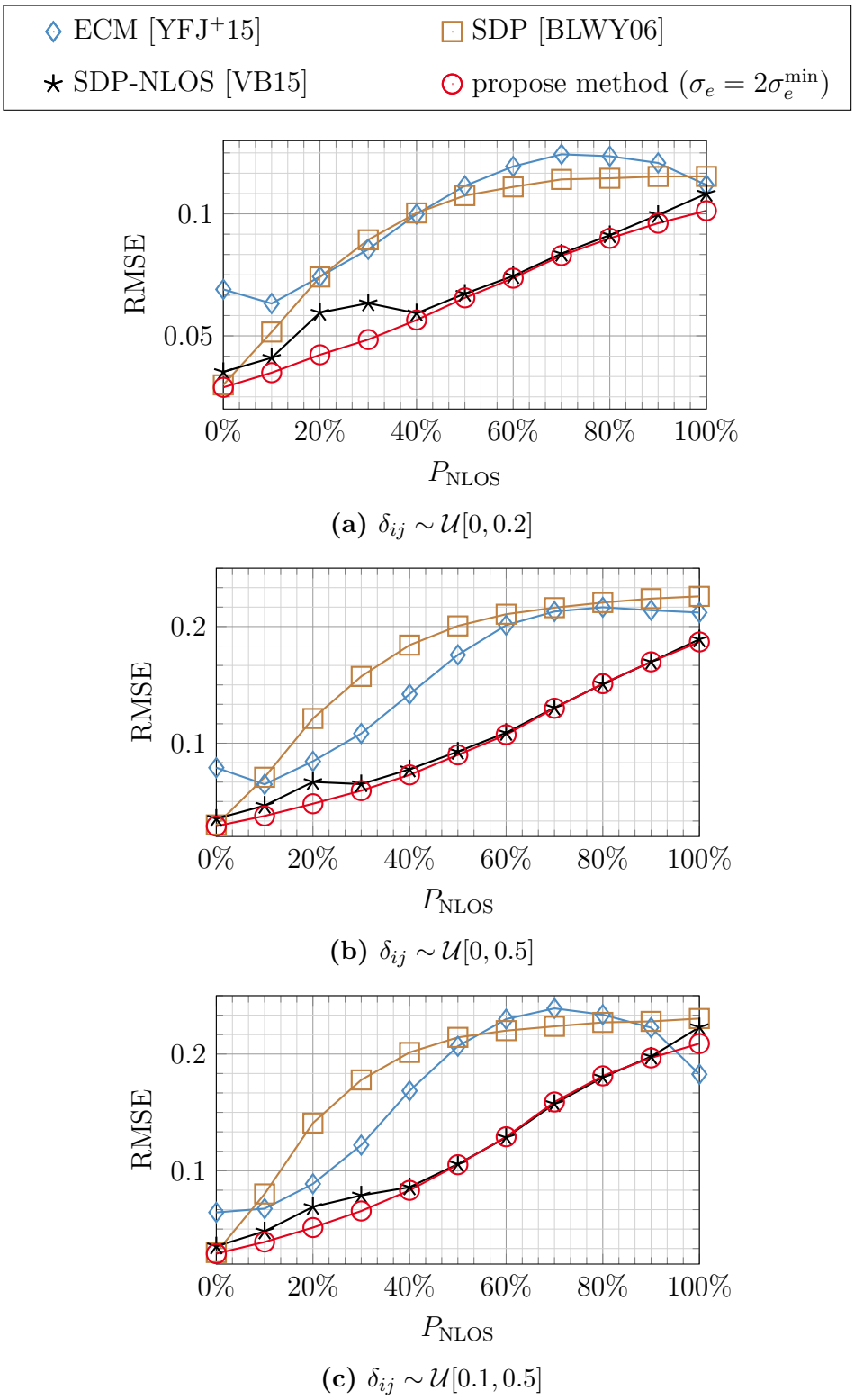
maximization (ECM) [YFJ+15], the SDP [BLWY06] and its variant designed to be robust against NLOS measurements in [VB15], referred to as SDP-NLOS. Here, the bound-constrained regularized SDP with $\sigma_e = 2\,\sigma_e^{\min}$ is selected to represent the proposed framework. Regarding the ECM method, the distribution of v_{ij} is initialized as $(1-\alpha_{\text{NLOS}})\mathcal{N}(0, 0.01^2) + \alpha_{\text{NLOS}}\mathcal{N}(\frac{a+b}{2}, \left(\frac{b-a}{4}\right)^2)$, where $\alpha_{\text{NLOS}} = P_{\text{NLOS}}$ when $P_{\text{NLOS}} \in [10\%, 90\%]$, $\alpha_{\text{NLOS}} = 0.95$ when $P_{\text{NLOS}} = 100\%$ and $\alpha_{\text{NLOS}} = 0.05$ when $P_{\text{NLOS}} = 0$. The position parameters are initialized as the true values plus a perturbation generated in $\mathcal{U}[-0.2, 0.2]$.

Figure 5.8 depicts the localization accuracy of different methods versus P_{NLOS} in scenarios where the statistics of the ranging bias δ_{ij} vary. It is observed that, compared with the SDP and the ECM methods, the proposed method ($\sigma_e = 2\,\sigma_e^{\min}$) achieves a significant performance improvement in almost all scenarios, not only in LOS-heavy scenarios but also in NLOS-heavy scenarios. This result indicates that our "sparsity" does not always mean "zero" entries overwhelmingly dominate "non-zero" entries, but can also mean that there is only a small fraction of "zero" entries. In both cases our proposed algorithm can exploit the "sparsity" property to improve the positioning accuracy. An additional observation is that the localization accuracy of the SDP-NLOS method is close to our method. Nevertheless, note that the SDP-NLOS method has two critical limitations: it is developed for Gaussian distributed measurement error and it requires the knowledge about the variance of the measurement error. By contrast, our method does not have these limitations.

Next, we evaluate how these methods perform at varying variance of measurement error. The localization RMSE curves are depicted in Fig. 5.9 for $P_{\text{NLOS}} = 0\%$, 20%, 40%. Once again, we see that our method attains the highest localization accuracy in almost all scenarios. The only exception occurs at $P_{\text{NLOS}} = 0\%$ (a pure LOS scenario), at which our method has a slight performance loss in comparison with the SDP method. This result is expected since this SDP method was developed for the pure LOS scenario. The simulation results presented so far demonstrate the ability of the proposed framework for NLOS effect mitigation in various scenarios.

5.6.3 Experimental Results

In addition to the synthetic data, it is interesting to see how does our method perform in a real sensor network. For this purpose, we evaluate all methods using the real sensor network and the TOA measurements provided in [PHP+03b]. The experimental setup is briefly reviewed below. The network consisted of 44 sensors in total, among

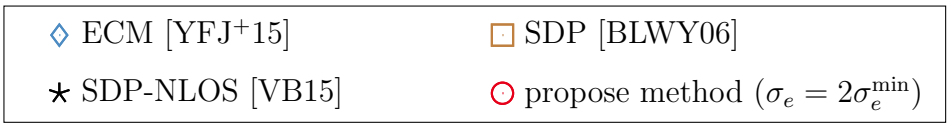


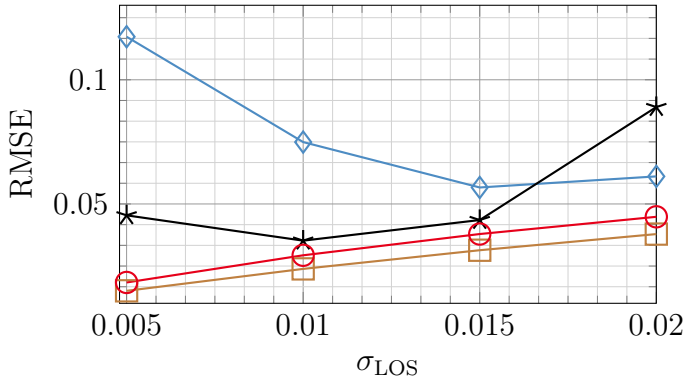
(a) $\delta_{ij} \sim \mathcal{U}[0, 0.2]$

(b) $\delta_{ij} \sim \mathcal{U}[0, 0.5]$

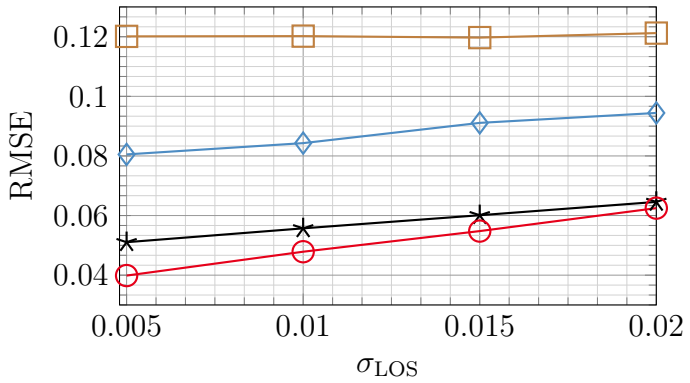
(c) $\delta_{ij} \sim \mathcal{U}[0.1, 0.5]$

Figure 5.8: Localization accuracy of different methods at different values of P_{NLOS} with $v_{ij} \sim \mathcal{N}(0, 0.01^2)$ and $N_a = 8$.

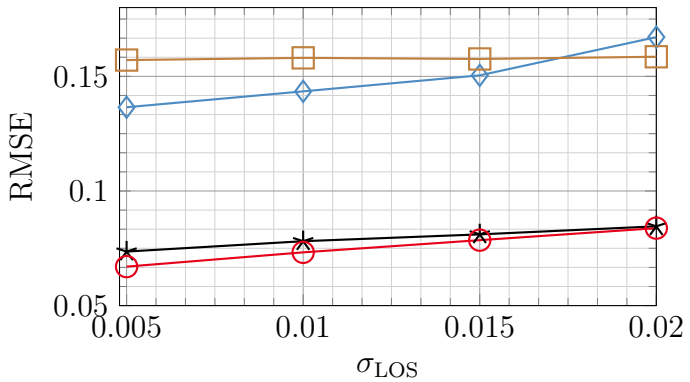




(a) $P_{\text{NLOS}} = 0\%$



(b) $P_{\text{NLOS}} = 20\%$



(c) $P_{\text{NLOS}} = 40\%$

Figure 5.9: Localization accuracy of different methods at varying σ_{LOS} with $v_{ij} \sim \mathcal{N}(0, \sigma_{\text{LOS}}^2)$, $\delta_{ij} \sim \mathcal{U}[0, 0.5]$ and $N_a = 8$.

which 4 anchors and 40 agents, and was deployed over an area of 14 m $\times$ 14 m. Figure 5.10 shows the deployment of the real sensor network, along with the position estimates obtained using our method with $\sigma_e = 2\,\sigma_e^{\min}$. There was a TOA measurement between each pair of sensors. The measurements were collected using a wideband direct-sequence spread-spectrum (DS-SS) transmitter and receiver (Sigtek model ST-515). For more details about the experiment, please see [PHP+03b].

To examine networks of different connectivity rates, we manually vary the communication range from 8 m to 14 m with a step size of 2 m. To test the fully connected scenario, a communication range of 20 m is considered as well. According to the Kolmogorov-Smirnov test, the distance measurement error fits well the model of $\mathcal{N}(0, 1.83^2)$, which is used as the initial model parameters for the ECM method. The initial position parameters are set to the true agent positions for the ECM method. This means that the most appropriate initialization parameters are provided for the ECM method. We stress that it is unrealistic to obtain such ideal initialization parameters in practice, since the true agent positions and measurement error model are rarely available.

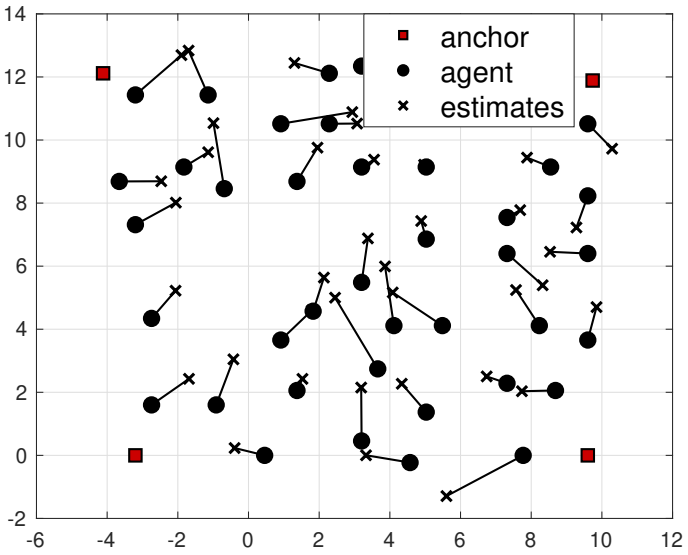


Figure 5.10: Deployment of the real sensor network and the position estimates obtained using the proposed framework with $\sigma_e = 2\,\sigma_e^{\min}$.

The result is shown in Fig. 5.11 and there are several observations. First, the localization error (RMSE $\geq$ 2.5 m) of the SDP-NLOS method is considerably larger than the other methods. We show the RMSE only for a range lower than 2.5 m so that the difference between the other methods becomes more evident. This poor performance

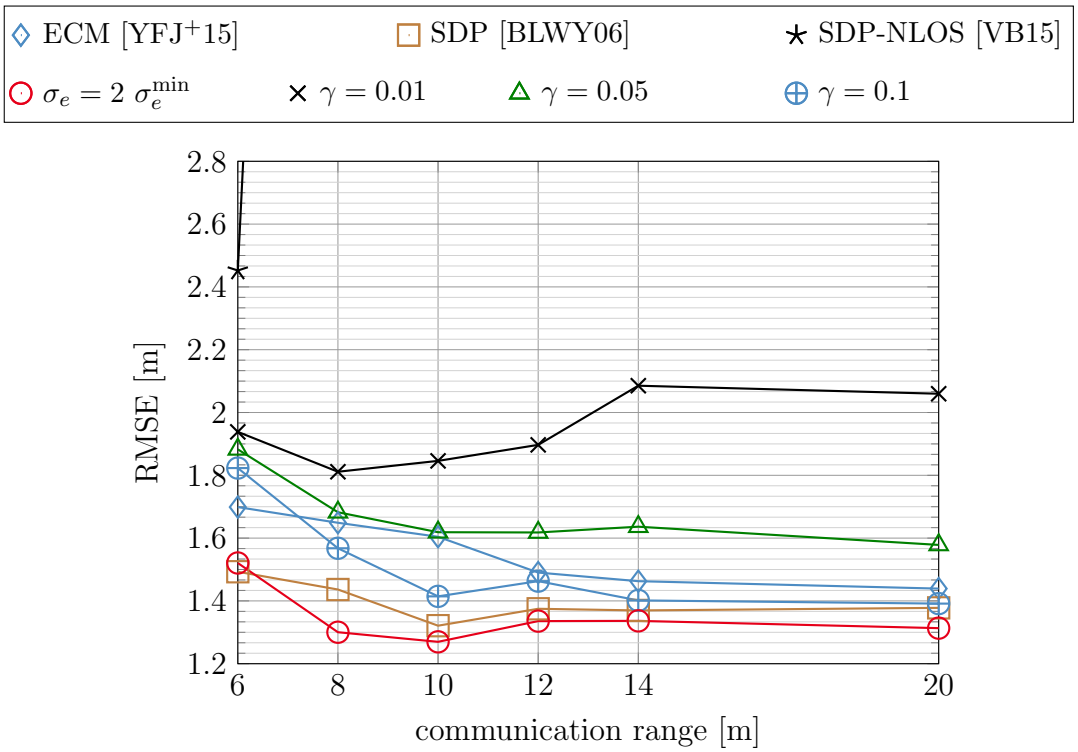


Figure 5.11: Localization accuracy of different methods at different communication ranges using real data.

of the SDP-NLOS method can be explained as follows. When there are excessively many connections, corresponding to the many constraints in the SDP-NLOS method, it becomes extremely challenging for the SDP-NLOS method to find a solution close to the true parameters. Second, the proposed approach ($\sigma_e = 2\,\sigma_e^{\min}$) once again retains the highest localization accuracy, followed by the SDP method. The good performance of the SDP method is due to the fact that the underlying scenario is close to a pure-LOS scenario. Regarding the sparsity-promoting regularized SDP, among these three parameters, $\gamma = 0.1$ seems to be the most appropriate one, followed by $\gamma = 0.05$, with $\gamma = 0.01$ being the least favorable one. However, all these three RMSE curves are above the $\sigma_e = 2\,\sigma_e^{\min}$ curve. This is because for this localization scenario, which is close to the pure-LOS scenario, a relatively large γ is more suitable. Evidently, $\gamma = 0.1$ is still not large enough. On the other hand, $\sigma_e^{\min}$ is dynamically determined for this special scenario, leading to the reliable outstanding localization performance of the method with $\sigma_e = 2\,\sigma_e^{\min}$. This example highlights the potential benefit of the proposed framework in conjunction with the proposed data-driven strategy in practical localization scenarios.

Methods:	ECM	SDP	SDP-NLOS	$\gamma = 0.05$	$\sigma_e = 2\sigma_e^{\min}$
Time [s]:	10.94	1.31	12.00	1.70	3.54

Table 5.2: Comparison of computational complexity in terms of simulation time. Here, the localization scenario is the same as that in Fig. 5.8-(b) with $P_{\text{NLOS}} = 20\%$.

5.6.4 Computational Complexity

In addition to the localization accuracy, computational efficiency is another important measure for practical localization problems. Here, we compare the computational efficiency of different methods in terms of simulation time. All SDPs are solved directly using the SeDuMi solver. The average simulation time, including preprocessing, interior-point iterations and postprocessing, is tabulated in Table 5.2. There are two main observations. First, the simulation time of $\sigma_e = c \cdot \sigma_e^{\min}$ is approximately twice that of the method with a predefined γ. This is an expected result because for the proposed parameter selection strategy, not only Problem (5.22) but also Problem (5.21) should be solved. This results in a doubled computational load. Second, all proposed methods, including the method with a predetermined γ and also the method with $\sigma_e = c \cdot \sigma_e^{\min}$, are computationally more efficient than the SDP-NLOS and the ECM methods. This is because the SDP-NLOS has many range-related constraints, which correspond to many semi-definite cones, which are quite time consuming to handle. The ECM method requires many iterations to achieve its local optimum, leading to its comparatively high computational load.

5.7 Summary

In this chapter, the general localization scenario, where range measurements are available but statistical knowledge of LOS/NLOS measurement errors is completely unknown, was considered. It was shown that the ranging biases are sparse in LOS-dominating environments. This sparsity was exploited so that the localization problem was formulated as two generic regularized optimization problems. To avoid the problem of local optima solutions, a specific SDP-based example was developed for each regularized formulation. It was theoretically shown, for certain conditions, that these two SDPs are equivalent in the sense that they share the same optimal solution. To select an appropriate regularization parameter, an efficient data-driven strategy was

devised which exploits the special structure of the optimization problem. It was shown in the numerical examples that our two SDP solutions, in particular, the proposed data-driven parameter selection method, provide overall good localization performance. In contrast to techniques with non convex objective functions, initialization of the proposed methods is not an issue because they are guaranteed to converge to the global optima.

Chapter 6

Conclusions and Outlook

The major advance detailed in this dissertation is that of cooperative localization methods for the practical case where full knowledge of the statistical measurement model is not available and, in particular, where experimental campaigns are not affordable and NLOS identification is not performed. The methods developed have a rigorous mathematical base and avoid the common problem of solutions that are local, not global.

For RSS measurements, and for the case where the model parameter is unknown, the cooperative localization problem was formulated as a statistical inference problem in a Bayesian framework with the positions and the unknown model parameter being treated as random variables. It was shown that message passing methods, with appropriate numerical strategies, are suitable for solving the statistical inference problem. In addition, the more general scenario, where range measurements are available but statistical knowledge of LOS/NLOS measurement errors is completely unknown, was considered. It was shown, first, that the bias parameters associated with measurement errors are sparse in LOS-dominating environments. Second, it was shown that this sparsity can be exploited and the localization problem can be formulated as two generic regularized optimization problems. To avoid the problem of local optima solutions, a specific SDP-based example was developed for each regularized formulation. To select an appropriate regularization parameter, an efficient data-driven strategy was utilized which exploits the special structure of the optimization problem.

The research documented in this thesis has led to some important open research questions and possible extensions. These are summarized below.

6.1 Inhomogeneous Environments

In Chapter 4, it was assumed that the localization environment is homogeneous with all links having a common path loss exponent. This is a simplification as practical localization environments are most likely inhomogeneous with each propagation link, e.g., connection link between node i and node j, having an individual path loss exponent. However, incorporating such inhomogeneity results in the measurement model

being under-determined. The most straightforward methods that are well suited for RSS-based localization in inhomogeneous environments are fingerprinting methods, e.g. [BP00]. Fingerprinting methods require an RSS map that is constructed based on experimental campaigns conducted in advance and such campaigns are likely to be not affordable in many applications. This necessitates the development of RSS-based cooperative localization methods for inhomogeneous environments when prior experimental campaigns are not available. This problem has only rarely been studied, is quite challenging and an open problem.

One promising research direction is to develop appropriate regularization methods dealing with the under-determined problem. Consider the physical properties of the path loss exponent:

$$\begin{aligned} &\alpha_{ij} = 2, && \text{if link } (i,j) \text{ is LOS,} \\ &\alpha_{ij} \in (2,6], && \text{if link } (i,j) \text{ is NLOS.} \end{aligned} \tag{6.1}$$

This implicitly indicates that vector $\boldsymbol{\alpha} = \{\alpha_{ij}\}_{(i,j)\in\Gamma}$ possesses the following two properties:

- Vector $\boldsymbol{\alpha} - 2$ is inherently element wise non-negative, meaning that $\alpha_{ij} - 2 \geq 0,\ \forall (i,j) \in \Gamma$;
- Vector $\boldsymbol{\alpha} - 2$ possesses sparsity property in LOS-dominating environments, since there are many zero elements in vector $\boldsymbol{\alpha} - 2$.

By utilizing the principles detailed in Chapter 5, it should be possible to develop appropriate regularization methods by exploiting the sparsity of vector $\boldsymbol{\alpha} - 2$.

6.2 Further Investigation of Regularized Methods

Chapter 5 detailed two generic methods to exploit the sparsity in the bias vector: the sparsity-promoting regularized method (5.3) and the bound-constrained regularized method (5.4). Two specific examples of these two generic methods, the sparsity-promoting regularized SDP (5.11) and the bound-constrained regularized SDP (5.13), were developed by specifying the choice for the residual error function and the sparsity-inducing function. The theoretical analysis and the data-driven parameter selection method given in Chapter 5 are limited to these two SDP methods.

In principle, other examples of these two generic methods can be constructed by choosing different residual error functions and different sparsity-inducing functions. As a

starting point, the residual error functions detailed in Table 5.1 can be considered and the l_0 norm can be considered as a sparsity-inducing function. In addition, the following warrants investigating. First, appropriate algorithms for solving the resulting optimization problems. Such algorithms would enable performance evaluation, and thereby providing an answer to the question as to whether introducing the sparsity term mitigates the effects of NLOS. In addition, a theoretical analysis on the relation between the resulting two examples of (5.3) and (5.4) is warranted. Another issue that warrants further investigation is how to select an appropriate regularization parameter. With respect to this problem, the sensitivity of localization performance with respect to the regularization parameter needs to be analyzed and the possibility of developing a data-driven parameter selection method needs to be investigated.

6.3 Computational Efficiency

Two core requirements for real-world wireless networks are high localization accuracy and robustness against NLOS measurements. Two additional, and significant, requirements are a light computational load and scalability as defined by the capability to process large networks. SDP methods are computationally much lighter than particle-based message passing methods. For example, for a moderately connected network with 100 agents, SDP methods require between 1.3 and 3.5 seconds to determine position estimates while particle-based methods can easily take several hours. The main reason for the heavy computational load of particle-based methods is that an analytical solution for the message-updating rule does not exist and numerical approximate solutions have to be utilized. Appropriate parametric forms have the potential to alleviate this problem by approximating the beliefs and/or messages. Further, variational Bayesian methods have the potential to provide approximate analytical solutions and, thus, have the potential to avoid computationally expensive Monte Carlo sampling methods.

Though the processing time of SDP methods is acceptable for moderately connected networks with of the order of 100 agents, the computational load becomes rapidly prohibitive for large and densely connected networks. The issue as to whether it is possible to achieve a reduction in computational load, but without a sacrifice of localization accuracy, warrants further investigation. Notably, the problem data associated with standard SDP problems, e.g., (11), is sparse. Accordingly, one possible research direction is to exploit the special structure of the SDP problems as in [FKMN00,NFF+03,ZFP+17]. In addition, the recently proposed first-order operator-splitting method, see [OCPB16], has potential for efficiently solving homogeneous self-dual embedding problems. Thus,

an alternative approach is to reformulate the problem as a homogeneous self-dual embedding problem and then solve it using the first-order operator-splitting method, e.g. [SZOL15].

6.4 Sources of Uncertainty

In real-world applications, the assumed statistical models commonly used do not perfectly match the underlying physical models. Two representative sources of uncertainty are associated with the reference power and the location of the anchor positions. For example, in Chapter 4, only the path loss exponent in the log-distance path loss propagation model (4.1) was assumed unknown. In real-world scenarios, the reference power may deviate from the reported value, e.g., caused by battery consumption. Further, the reference power may even be completely unknown. An extension of the proposed message passing methods to this harsh scenario would be useful. With respect to anchor positions, in practice, the locations of anchors are either manually calibrated or provided by certain GNSS receivers. As a result, anchor positions are only known with limited accuracy. When uncertainty information of the anchor positions is available, the following two approaches have potential. First, message passing methods can be extended in a straightforward manner by including the uncertainty information of anchor positions in the Bayesian framework. Alternatively, the sparsity-promoting regularized method (5.3) could be extended to take this uncertainty into account. More precisely, it can be adapted to the following minimax problem

$$\min_{\mathbf{X},\, \boldsymbol{\delta}\geq 0} \quad \max_{\{\mathbf{x}_{N_u+1},\ldots,\mathbf{x}_N\}\in S_a} \rho\big(\mathbf{X}, \boldsymbol{\delta}\big) + \gamma\, g(\boldsymbol{\delta}), \tag{6.2}$$

where S_a is the uncertainty set of the anchor positions $\{\mathbf{x}_{N_u+1}, \ldots, \mathbf{x}_N\}$. The key idea in this extension is to minimize the objective function in the least favorable case among all possible locations of anchors.

Appendix

A.1 Derivation of (4.12c)

We consider the calculation of the integral in (4.12c). For notational simplicity, we ignore the superscript $l, n-1$ during this derivation. The integral is calculated as follows:

$$\begin{aligned} Z_{ij} &= \int f(r_{ij}|\mathbf{x}_i, \mathbf{x}_j, \alpha)\ \mathrm{d}\mathbf{x}_i \\ &= \int f_{\mathcal{N}}\left(r_{ij} - A_i + 10\alpha \log_{10}\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{d_0}\right)\ \mathrm{d}\mathbf{x}_i \\ &\overset{①}{=} \int f_{\mathcal{N}}\left(r_{ij} - A_i + 10\alpha \log_{10}\frac{\|\mathbf{x}_{ij}\|}{d_0}\right)\ \mathrm{d}\mathbf{x}_{ij} \\ &\overset{②}{=} \int_0^{2\pi}\int_0^{\infty} f_{\mathcal{N}}\left(r_{ij} - A_i + 10\alpha \log_{10}\frac{d_{ij}}{d_0}\right)\cdot d_{ij}\ \mathrm{d}d_{ij}\ \mathrm{d}\theta_{ij} \\ &= 2\pi\frac{\log 10}{10\alpha}\ \mathbb{E}_{d_{ij}\sim\log\mathcal{N}\left(\mu_d,\sigma_d^2\right)}\left[d_{ij}^2\right] \\ &= 2\pi\frac{\log 10}{10\alpha}\cdot\exp(2\sigma_d^2 + 2\mu_d) \end{aligned} \tag{3}$$

with

$$\mu_d = \frac{\log 10}{10\alpha}(A_i - r_{ij}) + \log d_0, \qquad \sigma_d^2 = \sigma^2\frac{(\log 10)^2}{(10\alpha)^2}. \tag{4}$$

Here, $f_{\mathcal{N}}(\cdot)$ stands for the pdf of the Gaussian distribution $\mathcal{N}(0, \sigma^2)$, ① stands for $\mathbf{x}_{ij} = \mathbf{x}_i - \mathbf{x}_j$, and from ① to ② is achieved by transforming the Cartesian coordinate $\mathbf{x}_{ij}$ to the polar coordinate $[d_{ij}, \theta_{ij}]^T$.

A.2 Derivation of (4.30a)

For the measurement model in (4.1), the distance sample d_{ij}^l generated according to (4.28c)-(4.28b) fulfills the relation

$$\log\frac{d_{ij}^l}{d_0} = \underbrace{\frac{\log 10}{10\alpha^{l'}}\cdot(A_i - r_{ij}) + \frac{\log 10}{10\alpha^{l'}}\cdot v^l}_{\tilde{v}}. \tag{5}$$

Given α, A_i and r_{ij}, the variable $\tilde{v}$ is Gaussian distributed, namely, $\tilde{v} \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$ with

$$\tilde{\mu} = \frac{\log 10}{10\alpha^{l'}}\cdot(A_i - r_{ij}), \qquad \tilde{\sigma}^2 = \left(\frac{\log 10}{10\alpha^{l'}}\right)^2\cdot\sigma^2. \tag{6}$$

It follows that d_{ij}/d_0 is log-normal distributed, namely,

$$d_{ij}/d_0 \sim \text{Log}\mathcal{N}\big(\tilde{\mu}, \tilde{\sigma}^2\big). \tag{7}$$

Furthermore, it is given that the pdf of a log-normal-distributed random variable $a \sim \text{Log}\mathcal{N}(\mu_a, \sigma_a^2)$ is in the form of

$$f(a) = \frac{1}{\sqrt{2\pi}a\sigma_a}\exp\left(-\frac{(\log a - \mu_a)^2}{2\sigma_a^2}\right). \tag{8}$$

Finally, substituting a, μ_a and σ_a in (8) with d_{ij}/d_0, $\tilde{\mu}$ and $\tilde{\sigma}^2$, respectively, concludes the derivation.

A.3 Proof of Theorem 2

The proof is based on the KKT conditions of Problems (5.11) and (5.13). To derive the KKT conditions, we first reformulate both problems to the standard form of a linear conic program. We start with Problem (5.18), which is the equivalent form of Problem (5.11) without the absolute value operator.

Definition 1 *We define a bijection*

$$f_\Gamma : \Gamma \to \mathbb{Z}_\Gamma = \{1, \ldots, |\Gamma|\} \tag{9}$$

that is a one-to-one map between each pair of connection $(i, j) \in \Gamma$ *and one integer* $l \in \mathbb{Z}_\Gamma$.

By using the bijection defined above and dividing the objective function by factor γ, Problem (5.18) is further reformulated to

$$\begin{aligned}
\min_{\boldsymbol{\epsilon},\, \boldsymbol{\beta}^+,\, \boldsymbol{\beta}^-,\, \mathbf{Z}} \quad & \frac{1}{\gamma}\sum_{l=1}^{|\Gamma|}\big(\beta_l^+ + \beta_l^-\big) + \sum_{l=1}^{|\Gamma|}\epsilon_l, \\
\text{subject to} \quad & \epsilon_l \geq 0,\ \beta_l^+ \geq 0,\ \beta_l^- \geq 0, \quad l = 1, \ldots, \big|\Gamma\big|, \\
& \mathbf{Z} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{X} \\ \mathbf{X}^T & \mathbf{Y} \end{bmatrix} \succeq \mathbf{0}, \\
& \epsilon_l + \beta_l^+ - \beta_l^- + \text{tr}\big(\mathbf{E}_l\mathbf{Z}\big) = r_l^2,\ l = 1, \ldots, \big|\Gamma\big|.
\end{aligned} \tag{10}$$

Then, Problem (10) can be written in the standard form of a linear conic program, which is given by

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^T\mathbf{x} \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{K} \\ & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned} \tag{11}$$

Here, $\mathbf{x} = [\boldsymbol{\epsilon}; \boldsymbol{\beta}^+; \boldsymbol{\beta}^-; \mathrm{vec}(\mathbf{Z})]$ with $\mathrm{vec}(\mathbf{Z})$ being the vector constructed by the concatenation of all columns of $\mathbf{Z}$, $\mathbf{c} = \left[\mathbf{1}_{|\Gamma|}; \frac{1}{\gamma}\mathbf{1}_{|\Gamma|}; \frac{1}{\gamma}\mathbf{1}_{|\Gamma|}; \mathbf{0}_{n^2}\right]$ with $\mathbf{1}_{|\Gamma|}$ denoting the $|\Gamma| \times 1$ vector whose entries are all equal to one and $n = N_u + 2$, and cone $\mathcal{K}$ is the Cartesian product $\mathbb{R}_+^{|\Gamma|} \times \mathbb{R}_+^{|\Gamma|} \times \mathbb{R}_+^{|\Gamma|} \times \mathcal{S}_+^n$ with $\mathbb{R}_+^{|\Gamma|}$ being the $|\Gamma|$-dimensional non-negative orthant and $\mathcal{S}_+^n$ being the n-dimensional positive semi-definite cone. For brevity, the constraint matrix $\mathbf{A}$ and vector $\mathbf{b}$ are not explicitly given here. The *Lagrangian* $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ associated with Problem (11) is defined as

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \mathbf{c}^T\mathbf{x} - \boldsymbol{\lambda}^T\mathbf{x} + \boldsymbol{\nu}^T(\mathbf{b} - \mathbf{A}\mathbf{x}) \tag{12}$$

with $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_{\boldsymbol{\epsilon}}; \boldsymbol{\lambda}_{\boldsymbol{\beta}^+}; \boldsymbol{\lambda}_{\boldsymbol{\beta}^-}; \boldsymbol{\lambda}_{\mathbf{Z}}]$ and $\boldsymbol{\nu}$ being the KKT multipliers.

It is guaranteed that the objective function is non-negative in Problem (5.11) for any $\gamma > 0$, meaning that the optimal objective function cannot be $-\infty$. Slater's condition guarantees the strong duality, i.e., the zero-duality gap. Moreover, it also implies that there exists one pair of dual optimal points $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ when the dual optimal value is not $-\infty$, which is true due to the zero duality gap. Moreover, Condition 1) tells us that the optimal primal point $\mathbf{x}^*$ exists as well. Hence, there exists at least one pair of optimal primal/dual points $\mathbf{x}^* = \left[\boldsymbol{\epsilon}^*; \boldsymbol{\beta}^{+,*}; \boldsymbol{\beta}^{-,*}; \mathrm{vec}\left(\mathbf{Z}^*\right)\right], \boldsymbol{\lambda}^* = \left[\boldsymbol{\lambda}_{\boldsymbol{\epsilon}}^*; \boldsymbol{\lambda}_{\boldsymbol{\beta}^+}^*; \boldsymbol{\lambda}_{\boldsymbol{\beta}^-}^*; \boldsymbol{\lambda}_{\mathbf{Z}}^*\right]$ and $\boldsymbol{\nu}^*$ with zero duality gap. According to the KKT optimality, the following several conditions are sufficient and necessary for the primal/dual optimal points. Before listing the KKT conditions, several notations are first introduced. $\mathbf{A}_{(l)}$ and $\mathbf{A}_{(l)}^T$ denote the l-th column of matrix $\mathbf{A}$ and $\mathbf{A}_{(l)}$'s transpose, respectively. $\mathbf{A}_{(a:b)}$ and $\mathbf{A}_{(a:b)}^T$ denote the submatrix of $\mathbf{A}$ constructed by its columns a to b and $\mathbf{A}_{(a:b)}$'s transpose, respectively. $\mathbf{A}_{[c:d,a:b]}$ denotes the submatrix of $\mathbf{A}$ constructed by its rows c to d and columns a to b.

- Primal feasibility

$$\boldsymbol{\epsilon}^* \geq 0,\ \boldsymbol{\beta}^{+,*} \geq 0,\ \boldsymbol{\beta}^{-,*} \geq 0,\ \mathrm{vec}\left(\mathbf{Z}^*\right) \in \mathcal{S}_+^n; \tag{13}$$

- Primal feasibility

$$\begin{aligned} & \mathbf{A}\mathbf{x}^* = \mathbf{b} \\ \Longrightarrow & \begin{cases} \mathbf{Z}^*_{[1:2,1:2]} = \mathbf{I}_2 \\ \epsilon_l^* + \beta_l^{+,*} - \beta_l^{-,*} + \mathrm{tr}\left(\mathbf{E}_l\mathbf{Z}^*\right) = r_l^2, \qquad l = 1, \ldots, |\Gamma|; \end{cases} \end{aligned} \tag{14}$$

- Stationarity

$$\begin{aligned}
&\nabla_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) = \mathbf{0} \\
&\Longrightarrow \boldsymbol{\lambda}^* = \mathbf{c} - \mathbf{A}^T \boldsymbol{\nu}^*, \\
&\Longrightarrow \begin{cases}
\boldsymbol{\lambda}^*_{\boldsymbol{\epsilon}} & = \mathbf{1}_{|\Gamma|} - \mathbf{A}^T_{(1:|\Gamma|)} \boldsymbol{\nu}^*, \\
\boldsymbol{\lambda}^*_{\boldsymbol{\beta}^+} & = \frac{1}{\gamma} \mathbf{1}_{|\Gamma|} - \mathbf{A}^T_{(|\Gamma|+1:2|\Gamma|)} \boldsymbol{\nu}^*, \\
\boldsymbol{\lambda}^*_{\boldsymbol{\beta}^-} & = \frac{1}{\gamma} \mathbf{1}_{|\Gamma|} - \mathbf{A}^T_{(2|\Gamma|+1:3|\Gamma|)} \boldsymbol{\nu}^*, \\
\boldsymbol{\lambda}^*_{\mathbf{Z}} & = -\mathbf{A}^T_{(3|\Gamma|+1:3|\Gamma|+n^2)} \boldsymbol{\nu}^*;
\end{cases}
\end{aligned} \tag{15}$$

- Dual feasibility

$$\begin{aligned}
&\boldsymbol{\lambda}^*_{\boldsymbol{\epsilon}} \geq \mathbf{0}, \quad \boldsymbol{\lambda}^*_{\boldsymbol{\beta}^+} \geq \mathbf{0}, \quad \boldsymbol{\lambda}^*_{\boldsymbol{\beta}^-} \geq \mathbf{0}, \quad \boldsymbol{\lambda}^*_{\mathbf{Z}} \in \mathcal{S}^n_+ \\
&\Longrightarrow \begin{cases}
\mathbf{A}^T_{(l)} \boldsymbol{\nu}^* \leq 1, & l = 1, \ldots, |\Gamma|, \\
\mathbf{A}^T_{(l)} \boldsymbol{\nu}^* \leq \frac{1}{\gamma}, & l = |\Gamma|+1, \ldots, 2|\Gamma|, \\
\mathbf{A}^T_{(l)} \boldsymbol{\nu}^* \leq \frac{1}{\gamma}, & l = 2|\Gamma|+1, \ldots, 3|\Gamma|, \\
-\mathbf{A}^T_{(3|\Gamma|+1:3|\Gamma|+n^2)} \boldsymbol{\nu}^* \in \mathcal{S}^n_+;
\end{cases}
\end{aligned} \tag{16}$$

- Complementary conditions

$$\begin{aligned}
&\boldsymbol{\lambda}^{*T} \mathbf{x}^* = 0 \\
&\overset{①}{\Longrightarrow} \begin{cases}
\lambda^*_{\epsilon,\, l}\, \epsilon^*_l = 0, & l = 1, \ldots, |\Gamma|, \\
\lambda^*_{\beta^+,\, l}\, \beta^{+,*}_l = 0, & l = 1, \ldots, |\Gamma|, \\
\lambda^*_{\beta^-,\, l}\, \beta^{-,*}_l = 0, & l = 1, \ldots, |\Gamma|, \\
\lambda^{*T}_{\mathbf{Z}} \mathrm{vec}\big(\mathbf{Z}^*\big) = \Big[\mathbf{A}^T_{(3|\Gamma|+1:3|\Gamma|+n^2)} \boldsymbol{\nu}^*\Big]^T \mathrm{vec}\big(\mathbf{Z}^*\big) = 0.
\end{cases}
\end{aligned} \tag{17}$$

Note that Step ① follows from the fact that if the sum of several non-negative terms is zero, then each non-negative term must be zero.

Next, the KKT conditions of Problem (5.13) are derived. Problem (5.13) is first reformulated as

$$\begin{aligned}
\min_{\boldsymbol{\epsilon},\, \boldsymbol{\beta}^+,\, \boldsymbol{\beta}^-,\, \mathbf{Z}} \quad & \sum_{l=1}^{|\Gamma|} \epsilon_l \\
\text{subject to} \quad & \epsilon_l \geq 0,\ \beta^+_l \geq 0,\ \beta^-_l \geq 0, \quad l = 1, \ldots, |\Gamma|, \\
& \mathbf{Z} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{X} \\ \mathbf{X}^T & \mathbf{Y} \end{bmatrix} \succeq \mathbf{0}, \\
& \epsilon_l + \beta^+_l - \beta^-_l + \mathrm{tr}\big(\mathbf{E}_l \mathbf{Z}\big) = r_l^2, \quad l = 1, \ldots, |\Gamma|, \\
& \sum_{l=1}^{|\Gamma|} \big(\beta^+_l + \beta^-_l\big) \leq \sigma_e
\end{aligned} \tag{18}$$

and then put in the following form:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^T\mathbf{x} \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{K}, \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, \\ & \sum_{l=1}^{|\Gamma|} \left(\beta_l^+ + \beta_l^-\right) \leq \sigma_e. \end{aligned} \tag{19}$$

Here, $\mathbf{x} = [\boldsymbol{\epsilon}; \boldsymbol{\beta}^+; \boldsymbol{\beta}^-; \mathrm{vec}(\mathbf{Z})]$, $\mathbf{c} = \left[\mathbf{1}_{|\Gamma|}; \mathbf{0}_{|\Gamma|}; \mathbf{0}_{|\Gamma|}; \mathbf{0}_{n^2}\right]$, cone $\mathcal{K}$ denotes the Cartesian product $\mathbb{R}_+^{|\Gamma|} \times \mathbb{R}_+^{|\Gamma|} \times \mathbb{R}_+^{|\Gamma|} \times \mathcal{S}_+^n$. $\mathbf{A}$ and $\mathbf{b}$ are the same as in (11).

The Lagrangian $L(\mathbf{x}, \boldsymbol{\lambda}, \lambda_\sigma, \boldsymbol{\nu})$ associated with Problem (19) is defined as

$$L(\mathbf{x}, \boldsymbol{\lambda}, \lambda_\sigma, \boldsymbol{\nu}) = \mathbf{c}^T\mathbf{x} - \boldsymbol{\lambda}^T\mathbf{x} - \lambda_\sigma\left(\sigma_e - \sum_{l=1}^{|\Gamma|} \left(\beta_l^+ + \beta_l^-\right)\right) + \boldsymbol{\nu}^T(\mathbf{b} - \mathbf{A}\mathbf{x}) \tag{20}$$

with $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_{\boldsymbol{\epsilon}}; \boldsymbol{\lambda}_{\boldsymbol{\beta}^+}; \boldsymbol{\lambda}_{\boldsymbol{\beta}^-}; \boldsymbol{\lambda}_{\mathbf{Z}}]$, λ_σ and $\boldsymbol{\nu}$ being the KKT multipliers.

Let $\mathbf{x}^* = \left[\boldsymbol{\epsilon}^*; \boldsymbol{\beta}^{+,*}; \boldsymbol{\beta}^{-,*}; \mathrm{vec}\left(\mathbf{Z}^*\right)\right]$, $\boldsymbol{\lambda}^* = \left[\boldsymbol{\lambda}_{\boldsymbol{\epsilon}}^*; \boldsymbol{\lambda}_{\boldsymbol{\beta}^+}^*; \boldsymbol{\lambda}_{\boldsymbol{\beta}^-}^*; \boldsymbol{\lambda}_{\mathbf{Z}}^*\right]$, λ_σ^* and $\boldsymbol{\nu}^*$ be the optimal primal/dual points. According to the KKT optimality, the following several conditions are sufficient and necessary for the primal/dual optimal points.

- Primal feasibility

$$\boldsymbol{\epsilon}^* \geq \mathbf{0},\ \boldsymbol{\beta}^{+,*} \geq \mathbf{0},\ \boldsymbol{\beta}^{-,*} \geq \mathbf{0},\ \mathrm{vec}\left(\mathbf{Z}^*\right) \in \mathcal{S}_+^n; \tag{21}$$

- Primal feasibility

$$\begin{aligned} & \mathbf{A}\mathbf{x}^* = \mathbf{b} \\ \Longrightarrow & \begin{cases} \mathbf{Z}^*_{[1:2,1:2]} = \mathbf{I}_2 \\ \epsilon_l^* + \beta_l^{+,*} - \beta_l^{-,*} + \mathrm{tr}\left(\mathbf{E}_l\mathbf{Z}^*\right) = r_l^2, \qquad l = 1, \ldots, |\Gamma|; \end{cases} \end{aligned} \tag{22}$$

- Primal feasibility

$$\sum_{l=1}^{|\Gamma|} \left(\beta_l^{+,*} + \beta_l^{-,*}\right) \leq \sigma_e; \tag{23}$$

- Stationarity

$$\begin{aligned} & \nabla_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\lambda}^*, \lambda_\sigma^*, \boldsymbol{\nu}^*) = \mathbf{0} \\ & \Longrightarrow \boldsymbol{\lambda}^* = \mathbf{c} - \mathbf{A}^T\boldsymbol{\nu}^* + \lambda_\sigma^*\left[\mathbf{0}_{|\Gamma|}; \mathbf{1}_{|\Gamma|}; \mathbf{1}_{|\Gamma|}; \mathbf{0}_{n^2}\right], \\ & \Longrightarrow \begin{cases} \boldsymbol{\lambda}_{\boldsymbol{\epsilon}}^* & = \mathbf{1}_{|\Gamma|} - \mathbf{A}^T_{(1:|\Gamma|)}\boldsymbol{\nu}^*, \\ \boldsymbol{\lambda}_{\boldsymbol{\beta}^+}^* & = \lambda_\sigma^*\mathbf{1}_{|\Gamma|} - \mathbf{A}^T_{(|\Gamma|+1:2|\Gamma|)}\boldsymbol{\nu}^*, \\ \boldsymbol{\lambda}_{\boldsymbol{\beta}^-}^* & = \lambda_\sigma^*\mathbf{1}_{|\Gamma|} - \mathbf{A}^T_{(2|\Gamma|+1:3|\Gamma|)}\boldsymbol{\nu}^*, \\ \boldsymbol{\lambda}_{\mathbf{Z}}^* & = -\mathbf{A}^T_{(3|\Gamma|+1:3|\Gamma|+n^2)}\boldsymbol{\nu}^*; \end{cases} \end{aligned} \tag{24}$$

- Dual feasibility

$$\begin{aligned}&\boldsymbol{\lambda}_{\boldsymbol{\epsilon}}^* \geq \mathbf{0}, \quad \boldsymbol{\lambda}_{\boldsymbol{\beta}^+}^* \geq \mathbf{0}, \quad \boldsymbol{\lambda}_{\boldsymbol{\beta}^-}^* \geq \mathbf{0}, \quad \boldsymbol{\lambda}_{\mathbf{Z}}^* \in \mathcal{S}_+^n \\ \Longrightarrow &\begin{cases} \mathbf{A}_{(l)}^T \boldsymbol{\nu}^* \leq 1, & l = 1, \ldots, |\Gamma|, \\ \mathbf{A}_{(l)}^T \boldsymbol{\nu}^* \leq \lambda_\sigma^*, & l = |\Gamma|+1, \ldots, 2|\Gamma|, \\ \mathbf{A}_{(l)}^T \boldsymbol{\nu}^* \leq \lambda_\sigma^*, & l = 2|\Gamma|+1, \ldots, 3|\Gamma|, \\ -\mathbf{A}_{(3|\Gamma|+1:3|\Gamma|+n^2)}^T \boldsymbol{\nu}^* \in \mathcal{S}_+^n; \end{cases}\end{aligned} \tag{25}$$

- Dual feasibility

$$\lambda_\sigma^* \geq 0; \tag{26}$$

- Complementary conditions

$$\begin{aligned}&\boldsymbol{\lambda}^{*T} \mathbf{x}^* = 0 \\ \Longrightarrow &\begin{cases} \lambda_{\epsilon,\,l}^* \, \epsilon_l^* = 0, & l = 1, \ldots, |\Gamma|, \\ \lambda_{\beta^+,\,l}^* \, \beta_l^{+,*} = 0, & l = 1, \ldots, |\Gamma|, \\ \lambda_{\beta^-,\,l}^* \, \beta_l^{-,*} = 0, & l = 1, \ldots, |\Gamma|, \\ \boldsymbol{\lambda}_{\mathbf{Z}}^{*T} \mathrm{vec}\big(\mathbf{Z}^*\big) = \Big[\mathbf{A}_{(3|\Gamma|+1:3|\Gamma|+n^2)}^T \boldsymbol{\nu}^*\Big]^T \mathrm{vec}\big(\mathbf{Z}^*\big) = 0; \end{cases}\end{aligned} \tag{27}$$

- Complementary condition

$$\lambda_\sigma^* \Bigg(\sigma_e - \sum_{l=1}^{|\Gamma|} \big(\beta_l^{+,*} + \beta_l^{-,*}\big) \Bigg) = 0; \tag{28}$$

Proof: Under Condition 1), which guarantees the strong duality, the KKT optimality conditions are sufficient and necessary for the optimal solution. Let $\mathbf{x}^* = \big[\boldsymbol{\epsilon}^*; \boldsymbol{\beta}^{+,*}; \boldsymbol{\beta}^{-,*}; \mathrm{vec}\big(\mathbf{Z}^*\big)\big]$, $\boldsymbol{\lambda}^* = \Big[\boldsymbol{\lambda}_{\boldsymbol{\epsilon}}^*; \boldsymbol{\lambda}_{\boldsymbol{\beta}^+}^*; \boldsymbol{\lambda}_{\boldsymbol{\beta}^-}^*; \boldsymbol{\lambda}_{\mathbf{Z}}^*\Big]$, λ_σ^* and $\boldsymbol{\nu}^*$ be the optimal primal/dual points of Problem (5.13). This pair of primal/dual points $\mathbf{x}^*$, $\boldsymbol{\lambda}^*$, λ_σ^* and $\boldsymbol{\nu}^*$ must fulfill the KKT Conditions (21) - (28).

Suppose that we have $\gamma = \frac{1}{\lambda_\sigma^*}$. Condition 2) guarantees that $\lambda_\sigma^* \neq 0$, i.e., $\lambda_\sigma^* > 0$, thereby $\gamma > 0$ and finite. Then, it is apparent that the primal/dual points $\mathbf{x}^*$, $\boldsymbol{\lambda}^*$ and $\boldsymbol{\nu}^*$ fulfill the KKT Conditions (13) - (17) as well. Hence, $\mathbf{x}^*$, more specifically $\boldsymbol{\epsilon}^*$ and $\mathbf{Z}^*$, must be the optimal solution of Problem (5.11) with $\gamma = \frac{1}{\lambda_\sigma^*}$. □

A.4 Proof of Theorem 3

Proof: This proof is again based on the KKT conditions. Let $\mathbf{x}^* = \big[\boldsymbol{\epsilon}^*; \boldsymbol{\beta}^{+,*}; \boldsymbol{\beta}^{-,*}; \mathrm{vec}\big(\mathbf{Z}^*\big)\big], \boldsymbol{\lambda}^* = \Big[\boldsymbol{\lambda}_{\boldsymbol{\epsilon}}^*; \boldsymbol{\lambda}_{\boldsymbol{\beta}^+}^*; \boldsymbol{\lambda}_{\boldsymbol{\beta}^-}^*; \boldsymbol{\lambda}_{\mathbf{Z}}^*\Big]$ and $\boldsymbol{\nu}^*$ be one pair of optimal primal/dual points of Problem (5.11) with $\gamma = A$. This pair of primal/dual points $\mathbf{x}^*, \boldsymbol{\lambda}^*$ and $\boldsymbol{\nu}^*$ must fulfill the KKT Conditions (13) - (17).

Suppose that we have parameter $\sigma_e = \sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \mathrm{tr}\left(\mathbf{E}_{ij}\mathbf{Z}^*\right) - \epsilon_{ij}^* \right| = B$. Our purpose is to show that $\mathbf{x}^*, \boldsymbol{\lambda}^*$, $\lambda_\sigma^* = \frac{1}{\gamma} = \frac{1}{A}$ and $\boldsymbol{\nu}^*$ fulfill the KKT Conditions (21) - (28).

It is apparent that $\mathbf{x}^*, \boldsymbol{\lambda}^*$, $\lambda_\sigma^* = \frac{1}{\gamma} = \frac{1}{A}$ and $\boldsymbol{\nu}^*$ fulfill the KKT Conditions (21)-(22), (24)-(25) and (27). We now have to show that the other KKT conditions, including (23), (26) and (28), are fulfilled as well. Since we have $\sigma_e = \sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \mathrm{tr}\left(\mathbf{E}_{ij}\mathbf{Z}^*\right) - \epsilon_{ij}^* \right| = B$, it follows that both (23) and (28) hold. Moreover, since we have $\lambda_\sigma^* = \frac{1}{\gamma} = \frac{1}{A}$ and $A > 0$, it follows that $\lambda_\sigma^* \geq 0$, i.e., Condition (26) is fulfilled as well.

We have shown that $\mathbf{x}^*, \boldsymbol{\lambda}^*$, $\lambda_\sigma^* = \frac{1}{\gamma} = \frac{1}{A}$ and $\boldsymbol{\nu}^*$ must be the optimal solution of Problem (5.13) with $\sigma_e = \sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \mathrm{tr}\left(\mathbf{E}_{ij}\mathbf{Z}^*\right) - \epsilon_{ij}^* \right|$. Hence, $\boldsymbol{\epsilon}^*$ and $\mathbf{Z}^*$ must be the optimal solution of Problem (5.13) with $\sigma_e = \sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \mathrm{tr}\left(\mathbf{E}_{ij}\mathbf{Z}^*\right) - \epsilon_{ij}^* \right|$.

□

A.5 Proof of Theorem 4

Proof: (a). Let $\mathbf{x}^* = \left[\boldsymbol{\epsilon}^*; \boldsymbol{\beta}^{+,*}; \boldsymbol{\beta}^{-,*}; \mathrm{vec}\left(\mathbf{Z}^*\right)\right]$, $\boldsymbol{\lambda}^* = \left[\boldsymbol{\lambda}_{\boldsymbol{\epsilon}}^*; \boldsymbol{\lambda}_{\boldsymbol{\beta}^+}^*; \boldsymbol{\lambda}_{\boldsymbol{\beta}^-}^*; \boldsymbol{\lambda}_{\mathbf{Z}}^*\right]$, λ_σ^* and $\boldsymbol{\nu}^*$ be the optimal primal/dual points of Problem (5.13) with $\sigma_e = B$. We have to show that there exists a unique A such that $\boldsymbol{\epsilon}^*$ and $\mathbf{Z}^*$ are the optimal solution of Problem (5.11) with $\gamma = A$.

Suppose $A = \frac{1}{\lambda_\sigma^*}$. The proof in Section A.3 has shown that $\mathbf{x}^*$, $\boldsymbol{\lambda}^*$ and $\boldsymbol{\nu}^*$ must fulfill the KKT conditions of Problem (5.11) with $\gamma = \frac{1}{\lambda_\sigma^*} = A$, stating that $\boldsymbol{\epsilon}^*$ and $\mathbf{Z}^*$ are the optimal solutions of Problem (5.11) with $\gamma = A$. The uniqueness of A is guaranteed by the uniqueness of λ_σ^*, which is true under Condition 3).

(b). Let $\bar{\mathbf{x}}^* = \left[\bar{\boldsymbol{\epsilon}}^*; \bar{\boldsymbol{\beta}}^{+,*}; \bar{\boldsymbol{\beta}}^{-,*}; \mathrm{vec}\left(\bar{\mathbf{Z}}^*\right)\right], \bar{\boldsymbol{\lambda}}^* = \left[\bar{\boldsymbol{\lambda}}_{\boldsymbol{\epsilon}}^*; \bar{\boldsymbol{\lambda}}_{\boldsymbol{\beta}^+}^*; \bar{\boldsymbol{\lambda}}_{\boldsymbol{\beta}^-}^*; \bar{\boldsymbol{\lambda}}_{\mathbf{Z}}^*\right]$ and $\bar{\boldsymbol{\nu}}^*$ be the optimal primal/dual points of Problem (5.11) with $\gamma = A$. We have to show that $\bar{\boldsymbol{\epsilon}}^*$ and $\bar{\mathbf{Z}}^*$ are the optimal solutions of Problem (5.13) with $\sigma_e = B$.

Our proof is based on the KKT optimality conditions as well. Based on the proof in Section A.4, along with $\bar{\lambda}_\sigma^* = \frac{1}{\gamma} = \frac{1}{A}$, then $\bar{\mathbf{x}}^*, \bar{\boldsymbol{\lambda}}^*$ and $\bar{\boldsymbol{\nu}}^*$ are the optimal primal/dual points of Problem (5.13) with $\sigma_e = \sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \mathrm{tr}\left(\mathbf{E}_{ij}\bar{\mathbf{Z}}^*\right) - \bar{\epsilon}_{ij}^* \right|$. Next, we have to show that

$$\sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \mathrm{tr}\left(\mathbf{E}_{ij}\bar{\mathbf{Z}}^*\right) - \bar{\epsilon}_{ij}^* \right| = B \tag{29}$$

holds.

Let $\mathbf{x}^* = \left[\boldsymbol{\epsilon}^*; \boldsymbol{\beta}^{+,*}; \boldsymbol{\beta}^{-,*}; \mathrm{vec}\left(\mathbf{Z}^*\right)\right]$, $\boldsymbol{\lambda}^* = \left[\boldsymbol{\lambda}^*_{\boldsymbol{\epsilon}}; \boldsymbol{\lambda}^*_{\boldsymbol{\beta}^+}; \boldsymbol{\lambda}^*_{\boldsymbol{\beta}^-}; \boldsymbol{\lambda}^*_{\mathbf{Z}}\right]$, λ^*_σ and $\boldsymbol{\nu}^*$ be the optimal primal/dual points of Problem (5.13) with $\sigma_e = B$. Under Condition 2), the inequality constraint of problem (5.13) with $\sigma_e = B$ must be active, meaning that

$$\sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \mathrm{tr}\left(\mathbf{E}_{ij}\mathbf{Z}^*\right) - \epsilon^*_{ij} \right| = B. \tag{30}$$

Based on Proof (a), the optimal solutions, $\boldsymbol{\epsilon}^*$ and $\mathbf{Z}^*$, must be the optimal solutions of Problem (5.11) with $\gamma = A$ as well. This tells us that for Problem (5.11) with $\gamma = A$, there exists at least one pair of optimal solutions that fulfill

$$\sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \mathrm{tr}\left(\mathbf{E}_{ij}\mathbf{Z}^*\right) - \epsilon^*_{ij} \right| = B. \tag{31}$$

Moreover, Condition 4) guarantees that there exists a unique $\sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \mathrm{tr}\left(\mathbf{E}_{ij}\bar{\mathbf{Z}}^*\right) - \bar{\epsilon}^*_{ij} \right|$ for all optimal solutions of Problem (5.11) with $\gamma = A$. Hence, we have

$$\begin{aligned} &\sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \mathrm{tr}\left(\mathbf{E}_{ij}\bar{\mathbf{Z}}^*\right) - \bar{\epsilon}^*_{ij} \right| \\ =&\sum_{(i,j)\in\Gamma} \left| r_{ij}^2 - \mathrm{tr}\left(\mathbf{E}_{ij}\mathbf{Z}^*\right) - \epsilon^*_{ij} \right| \\ =&B. \end{aligned} \tag{32}$$

(c). The proof for the reverse statement follows the same principle. □

A.6 Proof of Theorem 5

Proof: The proof is based on the KKT Conditions (21)-(28) given in Section A.3. First, both $\mathbf{A}$ and $\mathbf{b}$ are explicitly given: $\mathbf{b} = \left[1; 0; 1; r_1^2, \ldots, r_{|\Gamma|}^2\right]$ and

$$\mathbf{A}_{(|\Gamma|+3)\times(3|\Gamma|+n^2)} = \begin{bmatrix} \overbrace{0,\ldots,0}^{\boldsymbol{\epsilon}} & \overbrace{0,\ldots,0}^{\boldsymbol{\beta}^+} & \overbrace{0,\ldots,0}^{\boldsymbol{\beta}^-} & \overbrace{1,\mathbf{0}^T}^{\mathrm{vec}(\mathbf{Z})} \\ 0,\ldots,0 & 0,\ldots,0 & 0,\ldots,0 & 0,1,\mathbf{0}^T \\ 0,\ldots,0 & 0,\ldots,0 & 0,\ldots,0 & \mathbf{0}^T_{n+1},1,\mathbf{0}^T \\ 1,0,\ldots,0 & 1,0,\ldots,0 & -1,0,\ldots,0 & \mathrm{vec}(\mathbf{E}_1) \\ \ddots & \ddots & \ddots & \vdots \\ 0,\ldots,0,1 & 0,\ldots,0,1 & 0,\ldots,0,-1 & \mathrm{vec}(\mathbf{E}_{|\Gamma|}) \end{bmatrix}. \tag{33}$$

From the dual feasibility Condition (25) and $\lambda_\sigma^* = 0$, we have

$$\begin{aligned}
\mathbf{A}_{(l)}^T \boldsymbol{\nu}^* &= \nu_{l+3}^* \leq 1, && l = 1, \ldots, |\Gamma|, \\
\mathbf{A}_{(l)}^T \boldsymbol{\nu}^* &= \nu_{l-|\Gamma|+3}^* \leq 0, && l = |\Gamma|+1, \ldots, 2|\Gamma|, \\
\mathbf{A}_{(l)}^T \boldsymbol{\nu}^* &= -\nu_{l-2|\Gamma|+3}^* \leq 0, && l = 2|\Gamma|+1, \ldots, 3|\Gamma|.
\end{aligned} \tag{34}$$

Combining the last two results, we have $0 \leq \nu_{l+3}^* \leq 0, \quad l = 1, \ldots, |\Gamma|$, leading to $\nu_{l+3}^* = 0$ for $l = 1, \ldots, |\Gamma|$. Therefore, we have $\boldsymbol{\nu}^* = [\nu_1^*; \nu_2^*; \nu_3^*; \mathbf{0}_{|\Gamma|}]$.

Next, we consider which values $\nu_1^*, \nu_2^*, \nu_3^*$ may take. Since

$$\begin{aligned}
\boldsymbol{\lambda}_{\mathbf{Z}}^* &= -\mathbf{A}_{(3|\Gamma|+1:3|\Gamma|+n^2)}^T \boldsymbol{\nu}^* \\
&= [\underbrace{-\nu_1^*, -\nu_2^*, 0, \ldots, 0}_{n+1}, -\nu_3^*, 0, \ldots, 0]
\end{aligned} \tag{35}$$

and $\boldsymbol{\lambda}_{\mathbf{Z}}^* \in \mathcal{S}_+^n$, we have

$$\begin{bmatrix} -\nu_1^* & 0 & \mathbf{0}_{n-2}^T \\ -\nu_2^* & -\nu_3^* & \mathbf{0}_{n-2}^T \\ \mathbf{0}_{n-2} & \mathbf{0}_{n-2} & \mathbf{0}_{(n-2)\times(n-2)} \end{bmatrix} \succeq \mathbf{0} \tag{36}$$

and therewith

$$\begin{bmatrix} -\nu_1^* & 0 \\ -\nu_2^* & -\nu_3^* \end{bmatrix} \succeq \mathbf{0}. \tag{37}$$

Due to symmetry, we have $\nu_2^* = 0$, and due to the positive semi-definiteness, we have $\nu_1^* \nu_3^* \geq 0$. Furthermore, from the complementary condition in (27), we have

$$\boldsymbol{\lambda}_{\mathbf{Z}}^{*\,T} \mathrm{vec}(\mathbf{Z}^*) = -\big(\nu_1^* \mathbf{Z}_{11}^* + \nu_3^* \mathbf{Z}_{22}^*\big) \overset{\mathbf{Z}_{11}^* = \mathbf{Z}_{22}^* = 1}{=} -\big(\nu_1^* + \nu_3^*\big) = 0. \tag{38}$$

If $\nu_1^* + \nu_3^* = 0$ and $\nu_1^* \nu_3^* \geq 0$, we have $\nu_1^* = \nu_3^* = 0$. Therefore, we have $\boldsymbol{\nu}^* = \mathbf{0}_{3+|\Gamma|}$.

Taking $\lambda_\sigma^* = 0$ and $\boldsymbol{\nu}^* = \mathbf{0}_{3+|\Gamma|}$ into account, from the stationarity Conditions (24), we have

$$\begin{aligned}
\boldsymbol{\lambda}_{\boldsymbol{\epsilon}}^* &= \mathbf{1}_{|\Gamma|} - \mathbf{A}_{(1:|\Gamma|)}^T \boldsymbol{\nu}^* = \mathbf{1}_{|\Gamma|}, \\
\boldsymbol{\lambda}_{\boldsymbol{\beta}^+}^* &= \lambda_\sigma^* \mathbf{1}_{|\Gamma|} - \mathbf{A}_{(|\Gamma|+1:2|\Gamma|)}^T \boldsymbol{\nu}^* = \mathbf{0}_{|\Gamma|}, \\
\boldsymbol{\lambda}_{\boldsymbol{\beta}^-}^* &= \lambda_\sigma^* \mathbf{1}_{|\Gamma|} - \mathbf{A}_{(2|\Gamma|+1:3|\Gamma|)}^T \boldsymbol{\nu}^* = \mathbf{0}_{|\Gamma|}, \\
\boldsymbol{\lambda}_{\mathbf{Z}}^* &= -\mathbf{A}_{(3|\Gamma|+1:3|\Gamma|+n^2)}^T \boldsymbol{\nu}^* = \mathbf{0}_{n^2}.
\end{aligned} \tag{39}$$

Moreover, due to the complementary condition $\lambda_{\epsilon,\,l}^* \, \epsilon_l^* = 0$ and $\boldsymbol{\lambda}_{\boldsymbol{\epsilon}}^* = \mathbf{1}_{|\Gamma|}$, it follows that $\boldsymbol{\epsilon}^* = \mathbf{0}_{|\Gamma|}$.

□

A.7 Proof of Corollary 1

Proof: According to Theorem 5, when the inequality constraint, $\sum_{(i,j)\in\Gamma}\left|r_{ij}^2 - \mathrm{tr}\big(\mathbf{E}_{ij}\mathbf{Z}\big) - \epsilon_{ij}\right| \leq \sigma_e$, is inactive at the optimal points $\boldsymbol{\epsilon} = \boldsymbol{\epsilon}^*$ and $\mathbf{Z} = \mathbf{Z}^*$, it holds that $\boldsymbol{\epsilon}^* = \mathbf{0}_{|\Gamma|}$. In this case, the optimization problem in (5.13) is equivalent to the following form:

$$\begin{aligned} \min_{\boldsymbol{\epsilon},\,\mathbf{Z}} \quad & 0 \\ \text{subject to} \quad & \epsilon_{ij} \geq 0, \quad \forall (i,j) \in \Gamma, \\ & \mathbf{Z} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{X} \\ \mathbf{X}^T & \mathbf{Y} \end{bmatrix} \succeq \mathbf{0}, \\ & \sum_{(i,j)\in\Gamma} \left|r_{ij}^2 - \mathrm{tr}\big(\mathbf{E}_{ij}\mathbf{Z}\big) - 0\right| \leq \sigma_e \end{aligned} \tag{40}$$

It is obvious that $\boldsymbol{\epsilon}$ is useless in (40). Discarding $\boldsymbol{\epsilon}$ gives rise to the following equivalent feasibility problem:

$$\begin{aligned} \text{find} \quad & \mathbf{Z} \\ \text{subject to} \quad & \mathbf{Z} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{X} \\ \mathbf{X}^T & \mathbf{Y} \end{bmatrix} \succeq \mathbf{0}, \\ & \sum_{(i,j)\in\Gamma} \left|r_{ij}^2 - \mathrm{tr}\big(\mathbf{E}_{ij}\mathbf{Z}\big)\right| \leq \sigma_e. \end{aligned} \tag{41}$$

□

List of Symbols

The most important symbols in the dissertation are listed in alphabetical order.

$\mathbf{0}_n$	$n \times 1$ all-zero vector
$\mathbf{1}_n$	$n \times 1$ all-one vector
A_0, A_i	reference power of all nodes, of node i
$B_i(\mathbf{x}_i)$	belief of position $\mathbf{x}_i$ at node i
$\mathbf{e}_i$	a vector with its i-th entry being one and the rest being zero
$\mathbf{E}_{ij}$	weighting matrix
$f(r_{ij}\|\mathbf{x}_i, \mathbf{x}_j, \alpha)$	likelihood function
$\tilde{f}(r_{ij}\|\mathbf{x}_i, \mathbf{x}_j, \alpha)$	normalized likelihood function
$f_i(\mathbf{x_i})$	prior probability of $\mathbf{x}_i$
$f_{\mathcal{N}}(\cdot)$	probability density function of normal distribution
$f_{\mathcal{U}[a,b)}(\cdot)$	probability density function of the uniform distribution $\mathcal{U}[a, b)$
$f_v(v)$	probability density function of v
$f_{\Delta(a,b,c)}(\cdot)$	probability density function of the triangle distribution $\Delta(a, b, c)$
$g(\cdot)$	sparsity-promoting function
(i, j)	one pair of neighbors
$\mathbf{I}_n$	$n \times n$ identity matrix
$\mathcal{K}$	cone
L	total number of particles
$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$	Lagrangian
$m_{f_{ij}\to i}(\mathbf{x}_i)$	message from factor f_{ij} to node i
$m_{f_{ij}\to\alpha}(\alpha)$	message from factor f_{ij} to node α
$m_{ij}(\mathbf{x}_i)$	message from node j to node i and shorthand-notation of $m_{f_{ij}\to i}(\mathbf{x}_i)$
$m_{ij}(\alpha)$	shorthand-notation of $m_{f_{ij}\to\alpha}(\alpha)$
$m_{s\to t}(x_t)$	message from node s to node t
N	total number of nodes
N_a	total number of anchors
$N_{\max}$	maximal number of iterations
N_u	total number of agents
$\mathcal{N}(\mu, \sigma^2)$	Gaussian distribution with mean μ and variance σ^2
$p(\cdot)$	probability mass function
P_{NLOS}	probability of NLOS
P_{pinf}	probability of primal infeasibility

$\Pr(\cdot)$	cumulative distribution function
$q(\cdot)$	proposal distribution
r_{ij}	RSS/range measurement between node i and node j
$\mathbf{r}$	collection of all measurements r_{ij}
R	total number of grid points
$\mathbb{R}$	real numbers
$\mathbb{R}_+^n$	n-dimensional non-negative orthant
S	index set of all nodes
S_a	index set of anchors
S_u	index set of agents
$\mathcal{S}^n$	set of $n \times n$ symmetric matrices
$\mathcal{S}_+^n$	set of $n \times n$ symmetric positive semi-definite matrices
$\mathcal{U}[a, b)$	uniform distribution on the interval $[a, b)$
v_{ij}	measurement error between node i and node j
$\left\{w_i^l, \mathbf{x}_i^l\right\}_{l=1}^L$	set of weighted particles representing $B_i(\mathbf{x}_i)$
$\left\{w_{ij}^l, \mathbf{x}_{ij}^l\right\}_{l=1}^L$	set of weighted particles representing $m_{f_{ij}\to i}(\mathbf{x}_i)$
$\left\{w_{j\to i}^l, \mathbf{x}_{j\to i}^l\right\}_{l=1}^L$	set of weighted particles representing $m_{ij}(\mathbf{x}_i)$
x_i	x-coordinate of node i
$\mathbf{x}_i$	position of node i
$\mathbf{X}$	position matrix
y_i	y-coordinate of node i
z_{ij}	quantized RSS measurement between node i and node j
$\mathbf{z}$	collection of all measurements z_{ij}
$\mathbf{Z}, \mathbf{Z}_{ij}$	positive semi-definite matrices containing all positions, containing $\mathbf{x}_i$ and $\mathbf{x}_j$
α, α_{ij}	path loss exponent of all links, of link (i, j)
$\boldsymbol{\beta}^-,\ \boldsymbol{\beta}^+$	slack variables
γ	regularization parameter
Γ	set of all pairs of neighbors
Γ_a	set of all pairs of agent-anchor neighbors
Γ_i	index set of node i's neighbors
Γ_u	set of all pairs of agent-agent neighbors
δ	bias parameter of range measurement
ϵ_{ij}	transformed bias associated with link (i, j)
$\boldsymbol{\epsilon}$	collection of all transformed bias ϵ_{ij}
$\boldsymbol{\lambda}$	KKT multiplier associated with the inequality constraints

$\boldsymbol{\mu}^k$	mean parameter of a Gaussian mixture
$\boldsymbol{\nu}$	KKT multiplier associated with the equality constraints
$\rho(\cdot)$	cost function
σ_e	residual error bound
$\sigma_e^{\min}$	the minimal value of residual error bound
$\boldsymbol{\Sigma}^k$	covariance matrix of a Gaussian mixture
ϕ^k	component coefficient of a Gaussian mixture
$\Phi(\cdot)$	cumulative distribution function of normal distribution
$\psi_{s,t}(x_s, x_t)$	potential function associated with edge (s, t)
$(\cdot)^*$	optimal solution

References

[BDS05] M. Briers, A. Doucet, and S. S. Singh, "Sequential Auxiliary Particle Belief Propagation," in *Proceedings of the International Conference on Information Fusion*, vol. 1, Philadelphia, USA, July 2005.

[Ber09] D. P. Bertsekas, *Convex Optimization Theory*. Belmont, Massachusetts, USA: Athena Scientific, 2009.

[Bis06] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York City, New York, USA: Springer, 2006.

[BLWY06] P. Biswas, T.-C. Lian, T.-C. Wang, and Y. Ye, "Semidefinite Programming Based Algorithms for Sensor Network Localization," *ACM Transactions on Sensor Networks*, vol. 2, no. 2, pp. 188–220, May 2006.

[BP00] P. Bahl and V. N. Padmanabhan, "RADAR: An In-building RF-based User Location and Tracking System," in *Proceedings of the IEEE Conference on Computer Communications*, vol. 2, Tel Aviv, Israel, March 2000, pp. 775–784.

[BTN01] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, Engineering Applications*. Philadelphia, PA, USA: MPS-SIAM Series on Optimization, 2001.

[BV04] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[CDG+14] A. Conti, D. Dardari, M. Guerra, L. Mucchi, and M. Z. Win, "Experimental Characterization of Diversity Navigation," *IEEE Systems Journal*, vol. 8, no. 1, pp. 115–124, March 2014.

[CLIC11] Y. T. Chan, B. H. Lee, R. Inkol, and F. Chan, "Received Signal Strength Localization With an Unknown Path Loss Exponent," in *24th Canadian Conference on Electrical and Computer Engineering (CCECE)*, Niagara Falls, Canada, May 2011, pp. 456–459.

[CMB+19] A. Conti, S. Mazuelas, S. Bartoletti, W. C. Lindsey, and M. Z. Win, "Soft Information for Localization-of-Things," *Proceedings of the IEEE*, vol. 107, no. 11, pp. 2240–2264, November 2019.

[CMS04] K. W. Cheung, W. K. Ma, and H. C. So, "Accurate Approximation Algorithm for TOA-Based Maximum Likelihood Mobile Location Using Semidefinite Programming," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, Montreal, Canada, May 2004.

[CPH06] J. Costa, N. Patwari, and A. Hero, "Distributed Weighted-Multidimensional Scaling for Node Localization in Sensor Networks," *ACM Transactions on Sensor Networks*, vol. 2, no. 1, pp. 39–64, February 2006.

[CPWG10] M. A. Caceres, F. Penna, H. Wymeersch, and R. Garello, "Hybrid GNSS-Terrestrial Cooperative Positioning via Distributed Belief Propagation," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM)*, Miami, USA, December 2010, pp. 1–5.

[CWA19] H. Chen, G. Wang, and N. Ansari, "Improved Robust TOA-Based Localization via NLOS Balancing Parameter Estimation," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 6177–6181, June 2019.

[CWW+12] H. Chen, G. Wang, Z. Wang, H. C. So, and H. V. Poor, "Non-Line-of-Sight Node Localization Based on Semi-Definite Programming in Wireless Sensor Networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 1, pp. 108–116, January 2012.

[Dat05] J. Dattorro, "Convex Optimization & Euclidean Distance Geometry," 2005.

[DCF+09] D. Dardari, A. Conti, U. Ferner, A. Giorgetti, and M. Z. Win, "Ranging With Ultrawide Bandwidth Signals in Multipath Environments," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 404–426, February 2009.

[DpG01] L. Doherty, K. S. J. pister, and L. E. Ghaoui, "Convex Position Estimation in Wireless Sensor Networks," in *Proceedings of the IEEE Conference on Computer Communications*, vol. 3, Anchorage, USA, April 2001.

[FHKZ09] C. Fritsche, U. Hammes, A. Klein, and A. M. Zoubir, "Robust Mobile Terminal Tracking in NLOS Environments Using Interacting Multiple Model Algorithm," in *Proceedings of the IEEE International Conference on Acoustics, Speech Signal Processing (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 3049–3052.

[FKMN00] M. Fukuda, M. Kojima, K. Murota, and K. Nakata, "Exploiting Sparsity in Semidefinite Programming via Matrix Completion I: General Framework," *SIAM Journal on Optimization*, vol. 11, no. 3, pp. 647–674, March 2000. [Online]. Available: https://doi.org/10.1137/S1052623400366218

[GB17] M. Grant and S. Boyd, "CVX: Matlab Software for Disciplined Convex Programming, version 2.1," 2017. [Online]. Available: http://cvxr.com/cvx

[GGL12] F. Gustafsson, F. Gunnarsson, and D. Lindgren, "Sensor Models and Localization Algorithms for Sensor Networks Based on Received Signal Strength," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, p. 16, January 2012.

[Gha12] L. E. Ghaoui, "Lecture Notes in Optimality Conditions for Conic Problems," March 2012.

[GSG19] P. M. Ghari, R. Shahbazian, and S. A. Ghorashi, "Maximum Entropy-Based Semi-Definite Programming for Wireless Sensor Network Localization," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3480–3491, 2019.

[HSG06] J. D. Hol, T. B. Schön, and F. Gustafsson, "On Resampling Algorithms for Particle Filters," in *Proceedings of the IEEE Nonlinear Statistical Signal Processing Workshop*, Cambridge, UK, September 2006, pp. 79–82.

[HWZ09] U. Hammes, E. Wolsztynski, and A. M. Zoubir, "Robust Tracking and Geolocation for Wireless Networks in NLOS Environments," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 5, pp. 889–901, October 2009.

[HZ11] U. Hammes and A. M. Zoubir, "Robust MT Tracking Based on M-Estimation and Interacting Multiple Model Algorithm," *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3398–3409, July 2011.

[IFMW05] A. T. Ihler, J. W. Fisher, R. L. Moses, and A. S. Willsky, "Nonparametric Belief Propagation for Self-localization of Sensor Networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 809–819, April 2005.

[IM09] A. Ihler and D. McAllester, "Particle Belief Propagation," in *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics*, Clearwater, USA, April 2009.

[ISFW03] A. T. Ihler, E. B. Sudderth, W. T. Freeman, and A. S. Willsky, "Efficient Multiscale Sampling From Products of Gaussian Mixtures," in *Proceedings of the 16th International Conference on Neural Information Processing Systems (NIPS)*, Vancouver, Canada, December 2003, pp. 1–8.

[JYF+15] D. Jin, F. Yin, C. Fritsche, A. M. Zoubir, and F. Gustafsson, "Efficient Cooperative Localization Algorithm In LOS/NLOS Environments," in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Nice, France, August 2015.

[JYF+16] ——, "Cooperative Localization Based on Severely Quantized RSS Measurements in Wireless Sensor Network," in *Proceedings of the IEEE Conference on Acoustic, Speech and Signal Processing (ICASSP)*, Shanghai, China, March 2016.

[JYF+20a] D. Jin, F. Yin, M. Fauß, M. Muma, and A. M. Zoubir, "Exploiting Sparsity for Robust Sensor Network Localization in Mixed LOS/NLOS Environments," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, May 2020, pp. 5915–5915.

[JYF+20b] D. Jin, F. Yin, C. Fritsche, F. Gustafsson, and A. M. Zoubir, "Bayesian Cooperative Localization Using Received Signal Strength With Unknown

Path Loss Exponent: Message Passing Approaches," *IEEE Transactions on Signal Processing*, vol. 68, no. 1, pp. 1120–1135, 2020.

[JYZS21] D. Jin, F. Yin, A. M. Zoubir, and H. C. So, "Exploiting Sparsity of Ranging Biases for NLOS Mitigation," *IEEE Transactions on Signal Processing*, vol. 69, pp. 3782–3795, 2021.

[KC76] C. Knapp and G. Carter, "The Generalized Correlation Method for Estimation of Time Delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, August 1976.

[Kv09] S. Korkmaz and A. van der Veen, "Robust Localization in Sensor Networks With Iterative Majorization Techniques," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Taipei, Taiwan, September 2009, pp. 2049–2052.

[LFS+12] J. Lien, U. J. Ferner, W. Srichavengsup, H. Wymeersch, and M. Z. Win, "A Comparison of Parametric and Sample-Based Message Representation in Cooperative Localization," *Hindawi Publishing Corporation*, September 2012.

[Li06] X. Li, "RSS-Based Location Estimation with Unknown Pathloss Model," *IEEE Transactions on Wireless Communications*, vol. 5, no. 12, pp. 3626–3633, December 2006.

[Li07] ——, "Collaborative Localization With Received-Signal Strength in Wireless Sensor Networks," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 6, pp. 3807–3817, November 2007.

[LZJ+19] B. Liu, X. Zhu, Y. Jiang, Z. Wei, and Y. Huang, "UAV and Piecewise Convex Approximation Assisted Localization With Unknown Path Loss Exponents," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 12, pp. 12 396–12 400, December 2019.

[MAF07] G. Q. Mao, B. D. O. Anderson, and B. Fidan, "Path Loss Exponent Estimation for Wireless Sensor Network Localization," *Computer Networks*, vol. 51, no. 10, pp. 2467 – 2483, July 2007.

[Man16] M. F. Mansour, "Channel Learning in Indoor Localization," in *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, March 2016, pp. 3886–3890.

[Mau12] R. Mautz, "Indoor Positioning Technologies," habilitation, ETH, 2012.

[MBL+09] S. Mazuelas, A. Bahillo, R. M. Lorenzo, P. Fernandez, F. A. Lago, E. Garcia, J. Blas, and E. J. Abril, "Robust Indoor Positioning Provided by Real-Time RSSI Values in Unmodified WLAN Networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 5, pp. 821–831, October 2009.

[MCAW18] S. Mazuelas, A. Conti, J. C. Allen, and M. Z. Win, "Soft Range Information for Network Localization," *IEEE Transactions on Signal Processing*, vol. 66, no. 12, pp. 3155–3168, June 2018.

[MFA07] G. Mao, B. Fidan, and B. D. O. Anderson, "Wireless Sensor Network Localization Techniques," *Computer Networks*, vol. 51, no. 10, p. 2529–2553, July 2007. [Online]. Available: https://doi.org/10.1016/j.comnet.2006.11.018

[MGWW10] S. Marano, W. M. Gifford, H. Wymeersch, and M. Z. Win, "NLOS Identification and Mitigation for Localization Based on UWB Experimental Data," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 7, pp. 1026–1035, September 2010.

[Mur12] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, Massachusetts London, UK: The MIT Press, 2012.

[NFF+03] K. Nakata, K. Fujisawa, M. Fukuda, M. Kojima, and K. Murota, "Exploiting Sparsity in Semidefinite Programming via Matrix Completion II: Implementation and Numerical Results," *Mathematical Programming*, vol. 95, no. 2, pp. 303–327, February 2003. [Online]. Available: https://doi.org/10.1007/s10107-002-0351-9

[NN03] D. Niculescu and B. Nath, "DV Based Positioning in Ad Hoc Networks," *Telecommunication Systems*, vol. 22, p. 267–280, January 2003.

[OCPB16] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, "Conic Optimization via Operator Splitting and Homogeneous Self-Dual Embeding," *Journal of Optimization Theory and Applications*, vol. 163, no. 3, pp. 1042–1068, June 2016.

[OFW11] P.-A. Oikonomou-Filandras and K.-K. Wong, "Hybrid Non-parametric Belief Propagation for Localization in Wireless Networks," in *IEEE Sensor Signal Processing Defence*, London, UK, September 2011.

[OWL10] R. W. Ouyang, A. K.-S. Wong, and C.-T. Lea, "Received Signal Strength-Based Wireless Localization via Semidefinite Programming: Noncooperative and Cooperative Schemes," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 3, pp. 1307–1318, March 2010.

[Pea88] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, California, USA: Morgan Kaufmann Publishers, 1988.

[PHP+03a] N. Patwari, A. O. Hero, M. Perkins, N. S. Correal, and R. J. O'Dea, "Relative Location Estimation in Wireless Sensor Networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2137–2148, August 2003.

[PHP+03b] ——, "Relative location estimation in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2137–2148, August 2003.

[Rap01] T. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, New York, USA: Prentice Hall PTR, 2001.

[RK89] R. Roy and T. Kailath, "ESPRIT-estimation of Signal Parameters via Rotational Invariance Techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, July 1989.

[RRW96] T. S. Rappaport, J. H. Reed, and B. D. Woerner, "Position Location Using Wireless Communications on Highways of the Future," *IEEE Communications Magazine*, vol. 34, no. 10, pp. 33–41, October 1996.

[RW07] D. Rudoy and P. J. Wolfe, "Multi-Scale MCMC Methods for Sampling from Products of Gaussian Mixtures," in *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 3, Honolulu, USA, April 2007, pp. 1201–1204.

[Sch86] R. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, March 1986.

[SGK12] N. Salman, M. Ghogho, and A. H. Kemp, "On the Joint Estimation of the RSS-Based Location and Path-loss Exponent," *IEEE Wireless Communications Letters*, vol. 1, no. 1, pp. 34–37, February 2012.

[SIFW03] E. B. Sudderth, A. T. Ihler, W. T. Freeman, and A. S. Willsky, "Nonparametric Belief Propagation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, Madison, USA, June 2003.

[SII+10] E. B. Sudderth, A. T. Ihler, M. Isard, W. T. Freeman, and A. S. Willsky, "Nonparametric Belief Propagation," *Communications of the ACM*, vol. 53, no. 10, pp. 95–103, October 2010.

[Sil86] B. W. Silvermann, *Density Estimation for Statistical and Data Analysis*, 1st ed. London, UK: Chapman and Hall, 1986.

[SJMZ16] F. Scheidt, D. Jin, M. Muma, and A. M. Zoubir, "Fast and Accurate Cooperative Localization in Wireless Sensor Networks," in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Budapest, Hungary, August 2016.

[SKG12] N. Salman, A. H. Kemp, and M. Ghogho, "Low Complexity Joint Estimation of Location and Path-Loss Exponent," *IEEE Wireless Communications Letters*, vol. 1, no. 4, pp. 364–367, June 2012.

[SL11] H. C. So and L. Lin, "Linear Least Squares Approach for Accurate Received Signal Strength Based Source Localization," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 4035–4040, August 2011.

[SL14] A. Simonetto and G. Leus, "Distributed Maximum Likelihood Sensor Network Localization," *IEEE Transactions on Signal Processing*, vol. 62, no. 6, pp. 1424–1437, 2014.

[SPZG10] V. Savic, A. Poblacioón, S. Zazo, and M. García, "An Experimental Study of RSS-Based Indoor Localization Using Nonparametric Belief Propagation Based on Spanning Trees," in *Proceedings of the International Conference on Sensor Technologies and Applications (SENSORCOMM)*, Venice, Italy, July 2010, pp. 238–243.

[SRZF04] Y. Shang, W. Rumi, Y. Zhang, and M. Fromherz, "Localization From Connectivity in Sensor Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 15, no. 11, pp. 961–974, November 2004.

[Stu99] J. F. Sturm, "Using SeDuMi 1.02, A Matlab Toolbox for Optimization Over Symmetric Cones," *Optimization Methods and Software*, vol. 11, no. 12, pp. 625–653, 1999.

[SZ13] V. Savic and S. Zazo, "Nonparametric Generalized Belief Propagation Based on Pseudo-Junction Tree for Cooperative Localization in Wireless Networks," *EURASIP Journal on Advances in Signal Processing*, vol. 2013, no. 1, p. 16, February 2013. [Online]. Available: http://asp.eurasipjournals.com/content/2013/1/16

[SZOL15] Y. Shi, J. Zhang, B. O'Donoghue, and K. B. Letaief, "Large-Scale Convex Optimization for Dense Wireless Cooperative Networks," *IEEE Transactions on Signal Processing*, vol. 63, no. 18, pp. 4729–4743, September 2015.

[TBD15] S. Tomic, M. Beko, and R. Dinis, "RSS-Based Localization in Wireless Sensor Networks Using Convex Relaxation: Noncooperative and Cooperative Schemes," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 5, pp. 2037–2050, May 2015.

[Tse07] P. Tseng, "Second-Order Cone Programming Relaxation of Sensor Network Localization," *SIAM Journal on Optimization*, vol. 18, no. 1, pp. 156–185, February 2007.

[VB07] S. Venkatesh and R. M. Buehrer, "NLOS Mitigation Using Linear Programming in Ultrawideband Location-Aware Networks," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 5, pp. 3182–3198, 2007.

[VB15] R. M. Vaghefi and R. M. Buehrer, "Cooperative Localization in NLOS Environments Using Semidefinite Programming," in *IEEE Communications Letters*, vol. 19, no. 8, August 2015, pp. 1382–1385.

[VGBS13] R. M. Vaghefi, M. R. Gholami, R. M. Buehrer, and E. G. Strom, "Cooperative Received Signal Strength-Based Sensor Localization With Unknown Transmit Powers," *IEEE Transactions on Signal Processing*, vol. 61, no. 6, pp. 1389–1403, March 2013.

[VVK03] J. J. Verbeek, N. Vlassis, and B. Kröse, "Efficient Greedy Learning of Gaussian Mixture Models," *Neural Computation*, vol. 15, p. 469–485, 2003.

[WCLA14] G. Wang, H. Chen, Y. Li, and N. Ansari, "NLOS Error Mitigation for TOA-Based Localization via Convex Relaxation," *IEEE Transactions on Wireless Communications*, vol. 13, no. 8, pp. 4119–4131, August 2014.

[WCLJ12] G. Wang, H. Chen, Y. Li, and M. Jin, "On Received-Signal-Strength Based Localization with Unknown Transmit Power and Path Loss Exponent," *IEEE Wireless Communications Letters*, vol. 1, no. 5, pp. 536–539, October 2012.

[Wei11] A. J. Weiss, "Direct Geolocation of Wideband Emitters Based on Delay and Doppler," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2513–2521, June 2011.

[WGW+19] Y. Wang, K. Gu, Y. Wu, W. Dai, and Y. Shen, "Exploiting NLOS Bias Correlation in Cooperative Localization," in *Proceedings of the IEEE International Conference on Communications Workshops (ICC Workshops)*, Shanghai, China, May 2019, pp. 1–5.

[WGW+20] ——, "NLOS Effect Mitigation via Spatial Geometry Exploitation in Cooperative Localization," *IEEE Transactions on Wireless Communications*, vol. 19, no. 9, pp. 6037–6049, 2020.

[WL07] M. Welling and J. J. Lim, "A Distributed Message Passing Algorithm for Sensor Localization," in *Proceedings of the International Conference on Artificial Neural Networks (ICANN)*. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 767–775.

[WLW09] H. Wymeersch, J. Lien, and M. Z. Win, "Cooperative Localization in Wireless Networks," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 427–450, 2009.

[WML+18] M. Z. Win, F. Meyer, Z. Liu, W. Dai, S. Bartoletti, and A. Conti, "Efficient Multisensor Localization for the Internet of Things: Exploring a New Class of Scalable Localization Algorithms," *IEEE Signal Processing Magazine*, vol. 35, no. 5, pp. 153–167, September 2018.

[WSN] "Wireless Sensor Network Market - Growth, Trends, and Forecast (2020 - 2025)," https://www.mordorintelligence.com/industry-reports/wireless-sensor-networks-market, accessed: 2020-09-05.

[WZCL18] G. Wang, S. Zhang, H. Chen, and Y. Li, "Robust TOA-Based Cooperative Localization Under NLOS Conditions," in *Proceedings of the 24th Asia-Pacific Conference on Communications (APCC)*, Ningbo, China, November 2018, pp. 418–421.

[YAZ+13] F. Yin, L. A., A. M. Zoubir, C. Fritsche, and F. Gustafsson, "RSS-based Sensor Network Localization in Contaminated Gaussian Measurement Noise," in *5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Saint Martin, French West Indies, France, December 2013, pp. 121–124.

[YFGZ13a] F. Yin, C. Fritsche, F. Gustafsson, and A. M. Zoubir, "Received Signal Strength-Based Joint Parameter Estimation Algorithm for Robust Geolocation in LOS/NLOS Environments," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 6471–6475.

[YFGZ13b] F. Yin, C. Fritsche, F. Gustafsson, and A. M. Zoubir, "TOA-Based Robust Wireless Geolocation and Cramér-Rao Lower Bound Analysis in Harsh LOS/NLOS Environments," *IEEE Transactions on Signal Processing*, vol. 61, no. 9, pp. 2243–2255, May 2013.

[YFGZ14] ——, "EM- and JMAP-ML Based Joint Estimation Algorithms for Robust Wireless Geolocation in Mixed LOS/NLOS Environments," *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 168–182, January 2014.

[YFJ+15] F. Yin, C. Fritsche, D. Jin, F. Gustafsson, and A. M. Zoubir, "Cooperative Localization in WSNs Using Gaussian Mixture Modeling: Distributed ECM Algorithms," *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1448–1463, March 2015.

[YZ12] F. Yin and A. M. Zoubir, "Robust Positioning in NLOS Environments Using Nonparametric Adaptive Kernel Density Estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, March 2012, pp. 3517–3520.

[YZFG13] F. Yin, A. M. Zoubir, C. Fritsche, and F. Gustafsson, "Robust Cooperative Sensor Network Localization via the EM Criterion in LOS/NLOS Environments," in *IEEE 14th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Darmstadt, Germany, June 2013, pp. 505–509.

[ZB11] R. Zekavat and R. M. Buehrer, *Handbook of Position Location: Theory, Practice and Advances*, 1st ed. Wiley-IEEE Press, 2011.

[ZFP+17] Y. Zheng, G. Fantuzzi, A. Papachristodoulou, P. Goulart, and A. Wynn, "Fast ADMM for Semidefinite Programs with Chordal Sparsity," in *American Control Conference*, Seattle, USA, May 2017, pp. 3335–3340.

[ZKCM12] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust Estimation in Signal Processing: A Tutorial-Style Treatment of Fundamental Concepts," *IEEE Signal Processing Magazine*, vol. 29, no. 4, pp. 61–80, July 2012.

Erklärung laut Promotionsordnung

§8 Abs. 1 lit. c PromO Ich versichere hiermit, dass die elektronische Version meiner Dissertation mit der schriftlichen Version übereinstimmt.

§8 Abs. 1 lit. d PromO Ich versichere hiermit, dass zu einem vorherigen Zeitpunkt noch keine Promotion versucht wurde. In diesem Fall sind nähere Angaben über Zeitpunkt, Hochschule, Dissertationsthema und Ergebnis dieses Versuchs mitzuteilen.

§9 Abs. 1 PromO Ich versichere hiermit, dass die vorliegende Dissertation selbstständig und nur unter Verwendung der angegebenen Quellen verfasst wurde.

§9 Abs. 2 PromO Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

Darmstadt, 19. October 2020

M.Sc. Di Jin