WIRELESS GEOLOCATION IN A NON-LINE-OF-SIGHT ENVIRONMENT

Yihong Qi

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
ELECTRICAL ENGINEERING

November 2003

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Professor Hisashi Kobayashi
(Principal Adviser)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Approved for the Princeton University Graduate School:

Dean of the Graduate School

To my parents

Abstract

Positioning a mobile station (MS) with a wireless communication system in a non-line-of-sight (NLOS) environment, which is sometimes called "NLOS geolocation" for short, has become an important issue with the rapid development of mobile communications in recent years. This dissertation deals with certain problems regarding NLOS geolocation. Specifically, four topics are investigated: a unified analysis of geolocation in an NLOS enviroment, the relationship among major distance-based schemes, a geolocation algorithm using the sequential simplex method (SSM) and a wavelet-based approach to channel estimation in a multipath environment.

In the unified analysis for NLOS geolocation, we answer two basic questions: what is the best positioning accuracy; and how to achieve it? In addition, the relationship between the conventional geometry-based methods and a theoretically optimal receivers is clarified.

The relationship among the time-of-arrival (TOA), time-difference-of-arrival (TDOA) and signal-strength (SS) based methods, is explored. The tradeoff between the accuracy limits for the TOA and SS based methods leads to our proposal of a hybrid distance estimation scheme that can take advantage of both TOA and SS information.

By combining the analytical results obtained above, a new geolocation algorithm that can mitigate NLOS effects is devised. The geolocation problem is formulated as a constrained optimization problem. The SSM serves as an optimization tool, which can handle a non-differentiable objective function and complex boundary conditions

in a simple manner.

If multipath propagation is present, channel estimation becomes important for the time-delay based positioning methods. A simple wavelet-based approach is developed for channel estimation. This approach takes advantage of two properties of the scaling function associated with an orthonormal wavelet: the bandwidth efficiency and the orthogonality among the scaling function and its time-shift replicas. This approach can be applied to the identification of a large class of linear systems, not just a wireless communication channel.

Acknowledgments

I wish to express my sincere gratitude to my advisor Professor Hisashi Kobayashi for his guidance, motivation and support throughout my graduate career at Princeton. I have learned a lot from his scientific thinking methodology, rigorous work attitude and his personality of integrity and generosity, which will continue to be good examples in my future research and life.

I want to thank all the faculty members in the Information Sciences and Systems group for providing an excellent education and research environment. My special thanks go to Professor Sun-Yuan Kung, Professor Vicent H. Poor and Professor Stuart Schwartz for their careful reading of my thesis and for their thoughtful comments. I am also grateful to Professor Ingrid Daubechies who introduces me wavelet theory and its applications in engineering fields.

I would like to acknowledge Dr. Hirohito Suda in NTT DoCoMo Inc. for his valuable advice and suggestions on the subject in Chapter 2 of this thesis.

I thank my colleagues in Professor Kobayashi's research group and the fellow graduate students for their friendship and help. I also thank Ms. Cynthia Menkes and Ms. Karen Williams for their friendly assistance.

I wish to dedicate this thesis to my parents for their constant support and encouragement.

This thesis work has been supported, in part, by research grants from the New Jersey Center for Wireless Telecommunications (NJCWT) and NTT DoCoMo Inc.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The term "wireless geolocation" is defined as an action to locate a mobile station (MS) in a wireless communication system. Nowadays the demand for wireless geolocation techniques goes far beyond the military use for which the techniques were originally developed. Besides soldiers, pilots and sailors, ordinary citizens can benefit from the geolocation capability to make their lives and work more productive and safer. In the United States, the Federal Communications Commission (FCC) has mandated cellular system operators to estimate the position of an emergency caller with accuracy of less than 125 meters in Enhanced 911 (E-911) service [1]. Other important geolocation applications, some of which are yet to be developed, include navigational services such as Intelligent Transport System (ITS), road-side assistance, location-sensitive billing and "mobile yellow pages".

Conventional geolocation methods have been well deployed in earlier navigation systems such as Decca and Loran [2] which are terrestrial-based location systems, and in more recent Global Positioning System (GPS) which is a highly successful and accurate system using a constellation of 24 satellites [3]. The conventional methods usually take locations of base stations (BS) as reference points and adopt one or more types of measurements pertaining to position information, i.e., time of arrival (TOA),

time difference of arrivals (TDOA) and angle of arrival (AOA). The desired position is then calculated from triangulation.

With pervasive usage of cell phones and the expanding demand of the market, major challenges of the geolocation technology have two aspects: marketing and technology. From the marketing aspect, cellular system operators prefer to rely on their existing cellular communications infrastructure to provide geolocation related services in lieu of some designated positioning system such as GPS. It is also of interest to the consumers to keep modifications of cell phones at an affordable level. These requirements prompt system designers to look into the trade-offs among three basic geolocation architectures: mobile-based, network-based, and hybrid geolocations [4]. From the technical aspect, a main challenge is to deal with harsh mobile environments, especially non-line-of-sight (NLOS) propagation of a radio signal, which can severely degrade the positioning accuracy. Therefore, the conventional methods often become inadequate, since they function well only in line-of-sight (LOS) environments. Furthermore, these conventional schemes that apply a rather heuristic geometric argument to location-bearing measurements do not explicitly take into account such parameters as signal-to-noise ratio (SNR) and the signal bandwidth when combining and processing the measurements from different BSs. Thus, optimal solutions cannot be guaranteed.

Therefore, in this dissertation, we address the following two fundamental technical questions for NLOS geolocation. What is the best achievable geolocation accuracy in an NLOS environment? And how should the corresponding optimum receiver be structured?

1.1 Overview of the Dissertation

Major contributions of this dissertation are in the following areas:

- Unified analysis of NLOS geolocation (Chapter 2)
- Relationship among the distance-based positioning methods (Chapter 3)
- A geolocation algorithm using the sequential simplex method (Chapter 4)
- A wavelet-based approach to channel estimation (Chapter 5)

We provide here a brief overview of each chapter.

Chapter 2 presents a unified analysis of geolocation in an NLOS environment, where the TOA, TDOA and signal strength (SS) based positioning methods are investigated in an integrated manner. From the viewpoint of estimation theory [5], we answer the above two basic questions of the subject, i.e., what is the best positioning accuracy? and how to achieve it? In addition, the relationship between the conventional methods and the theoretically optimal receivers is established. Specifically, the conventional methods are reduced to special cases of the optimum solutions under certain conditions.

In **Chapter 3**, we explore the connections among the three distance based methods (TOA, TDOA and SS positioning), in contrast to the treatment in the previous chapter where these methods are separately investigated. An analytical explanation is provided for the argument that given a set of BS locations and an MS position, the TOA method should achieve higher positioning precision than the TDOA counterpart. Moreover, it is revealed that the two positioning methods may attain the same level of accuracy under certain conditions. On the other hand, we pursue the

tradeoff between the accuracy limits for the TOA and SS based methods, which leads to a hybrid geolocation scheme that combines both TOA and SS data.

The above two chapters represent the theoretical contributions of the dissertation. We turn to some practical issues in the remaining chapters.

A geolocation algorithm using the Sequential Simplex Method (SSM) is proposed in **Chapter 4**. The NLOS geolocation problem is formulated as a constrained optimization problem, in which the SSM serves as an optimization tool. It involves an iterative evaluation of an objective function itself in contrast to evaluation of derivatives of the objective function performed in typical gradient methods. The movement of an MS is often limited by geographic conditions (such as street and highway layouts) and a finite speed of the MS. The SSM can incorporate such complex boundary conditions in a simple manner.

When the signal used for geolocation is subject to multipath propagation, channel estimation becomes an important issue for TOA positioning, since we need to precisely locate the first component of the arriving signal. As an initial attempt in this direction, we develop a wavelet-based method for channel estimation in **Chapter 5**. To estimate the channel impulse response, a training sequence is constructed using a scaling function associated with an orthonormal wavelet. By appropriately sampling the received signal, we are able to obtain the projection of the channel impulse response onto the subspace spanned by the scaling functions, which then allows a simple yet accurate reconstruction of the channel response. This scheme takes advantage of two properties of the scaling function: orthogonality between the scaling function and its time-shift replica, and its bandwidth efficiency. The amount of computations required is proportional to the number of data samples used, which is much less than other known methods. The scheme is also robust to some severe channel conditions,

such as low SNR and long delay spread.

We begin each of the following four chapters with an introduction and a review of the state of the art in the subject area, and then present our development of analysis or algorithms.

Chapter 2

Unified Analysis of NLOS Geolocation

2.1 Introduction

Positioning using radio signals that are subject to non-line-of-sight (NLOS) propagation has long been recognized as a difficult task. Many research efforts have been devoted to finding better solutions in recent years, driven by the increasing demand for the geolocation function in cellular communications.

While the past experiences of geolocation in a line-of-sight (LOS) environment were successful, the common trend of the recent research is to modify the algorithms and analysis employed in LOS geolocation by accommodating NLOS factors. The application of estimation theory to radar systems in estimating the distance and velocity of moving objects is well discussed in [9]. Its extension to the LOS geolocation problem is rather straightforward [13]: the dimensionality is increased from one (in radar) to two or three (in geolocation), and multiple transmitters (or receivers) are considered due to multiple base stations (BSs). When propagation of radio signals involve NLOS paths, a new type of parameters, i.e., additional path lengths induced

by NLOS propagation, must be taken into account. These NLOS delays are often treated as small variations of LOS delays [11, 12]. In such treatments, most results obtained under the LOS assumption are adopted. However, they cannot take into account specific characteristics of NLOS delays, e.g., NLOS delays cannot be negative. Many algorithms, e.g., [14, 23, 24, 25], have been proposed to mitigate NLOS effects. The rationale of these schemes often comes from some ad hoc geometric or intuitive arguments. Thus, few of them can be verified to be optimal. We believe that a systematic understanding based on a comprehensive framework of the NLOS geolocation problem is called for.

This chapter presents our attempts along this direction. We pursue two fundamental questions of NLOS geolocation within a general framework that accommodates various geolocation methods: what is the best positioning accuracy in an NLOS environment? and how to achieve it? A closely related question is concerned with the relationship between theoretically optimum solutions and the conventional methods that are based on measurements pertaining to position information, i.e., time of arrival (TOA), time difference of arrival (TDOA) and signal strength.

Our formulation starts with a set of received signal waveforms themselves[1], instead of such intermediate statistics as TOA and TDOA data. NLOS effects are viewed as random variables. Depending on whether statistical information on the variables is available or not, we separate our investigation in two stages. In the first stage, we concentrate on a scenario where NLOS and LOS signals coexist, but with no knowledge of NLOS delays. We show that the best geolocation accuracy, represented by the Cramer-Rao Lower Bound (CRLB), is attained asymtotically by maximum likelihood estimation (MLE) made on the above intermediate statistics. Thus the

[1] We treat the signal strength based positioning method in a different way, which is to be discussed in Section 2.3.2

conventional methods can serve as an important component of the optimal receivers. Moreover, it turns out that the best location estimate should be *solely* based on LOS signals, by discarding NLOS signals entirely. This somewhat surprising result sheds some light on how to simplify geolocation algorithms in practice.

We then in the second stage pursue a more general and realistic scenario where *a priori* information of the NLOS delays is provided in terms of some parametric distributions such as Gamma distribution. Accordingly, we need to extend our previous result by incorporating the prior information. The best geolocation accuracy, given by what we term "Generalized CRLB (G-CRLB)", is shown to be attained by maximum *a posterior* (MAP) estimators. We further obtain tight lower and upper bounds for the G-CRLB. The lower bound is achieved when we know exactly the amount of NLOS delays. This is an ideal case where an NLOS signal can be treated as a LOS signal by subtracting the fixed amount from the overall time delay. In the other extreme, the upper bound is reached when the variances of the NLOS delays grow to infinity. Then the result reduces to the CRLB obtained in the first stage.

To emphasize the central ideas, we start with in Section 2.2 a detailed derivation of the above statement for TOA positioning, which outlines a typical procedure of our analysis. We then integrate the TDOA and signal strength (SS) based positioning into the framework developed for TOA positioning in the remaining sections.

2.2 TOA positioning

2.2.1 Problem Formulation

Consider a synchronous communication system, where clocks at an MS and BSs are strictly synchronized. A BS receives the radio signal transmitted from the MS[1], via a single (LOS or NLOS) propagation path. It is assumed that we have some pre-knowledge to determine whether the signal is via a LOS path or not. Several algorithms [16, 17, 18] have been proposed to perform such identification. Then, by processing the received signals at all the BSs, the MS position is to be estimated. Let $\mathcal{B} = \{1, 2, \cdots, B\}$ be the set of indices of all the base stations involved, whose locations $\left\{\boldsymbol{p}_b = (x_b, y_b)^T,\ b \in \mathcal{B}\right\}$ are known. Denote the set of M BSs that receive NLOS signals by $\mathcal{NL} = \{k_1, k_2, \cdots, k_M\}$. We can label the elements of $\mathcal{NL} = \{1, 2, \cdots, M\}$ without loss of generality. The complement of $\mathcal{NL}$, denoted as $\mathcal{L}(= \mathcal{B} \setminus \mathcal{NL})$, is the set of LOS stations, with its cardinality being $L = B - M$. The unknown parameter of our interest is the MS position $\boldsymbol{p} = (x, y)^T$, but there are M additional parameters that need to be estimated, i.e., NLOS propagation induced path lengths,

$$\boldsymbol{l} = \begin{pmatrix} l_1 \\ l_2 \\ \vdots \\ l_M \end{pmatrix}.$$

Thus, we define an $(M + 2)$-dimensional vector $\boldsymbol{\theta}$ by concatenating the unknown $\boldsymbol{p}$ and $\boldsymbol{l}$:

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{p} \\ \boldsymbol{l} \end{pmatrix}.$$

[1]Here we only consider the geolocation using up-link signals. Yet most results based on this assumption should hold for a positioning system that uses down-link signals.

Let τ_b be the time delay of the received signal at the b-th base station (BS_b):

$$\tau_b = \frac{1}{c}\left\{\sqrt{(x_b - x)^2 + (y_b - y)^2} + l_b\right\}, \tag{2.1}$$

where $l_b = 0$ if $b \in \mathcal{L}$, and $c = 3 \times 10^8$m/s is the speed of light. Then, the baseband representation of the received signal at BS_b is

$$r_b(t) = A_b s(t - \tau_b) + n_b(t), \quad \text{for } b \in \mathcal{B}, \tag{2.2}$$

where A_b is the signal amplitude, $s(t)$ is a known signal waveform, and $n_b(t)$'s are independent complex-valued white Gaussian noise processes with spectral density $N_0/2$. Hence, the joint probability density function (p.d.f.) of the observables $\{r_b(t),\ b \in \mathcal{B}\}$ conditioned on $\boldsymbol{\theta}$ is

$$f_{\boldsymbol{\theta}}(\boldsymbol{r}) \propto \prod_{b=1}^{B} \exp\left\{-\frac{1}{N_0}\int |r_b(t) - A_b s(t - \tau_b)|^2\, dt\right\}. \tag{2.3}$$

This p.d.f plays a key role in casting the NLOS geolocation into a multi-parameter estimation problem.

2.2.2 Cramer-Rao Lower Bound

It is well known that the Cramer-Rao Lower Bound (CRLB) sets a lower limit for the variance (or covariance matrix) of any unbiased estimates of an unknown parameter (or unknown parameters) [5, 6]. For the TOA positioning method, we will show in the next section that the maximum likelihood (ML) estimate for the MS position can be approximated as an unbiased estimate under certain conditions. Thus, the CRLB is an appropriate formula to provide the best geolocation accuracy when we do not have any prior information of the NLOS delays $\boldsymbol{l}$. The CRLB is given as follows. Recall $f_{\boldsymbol{\theta}}(\boldsymbol{r})$ is the p.d.f. of observations $\boldsymbol{r}$ conditioned on $\boldsymbol{\theta}$. The *Fisher information*

matrix (FIM) is determined by

$$\mathbf{J}_{\boldsymbol{\theta}} = E_{\boldsymbol{\theta}} \left[\frac{\partial}{\partial \boldsymbol{\theta}} \ln f_{\boldsymbol{\theta}}(\boldsymbol{r}) \cdot \left(\frac{\partial}{\partial \boldsymbol{\theta}} \ln f_{\boldsymbol{\theta}}(\boldsymbol{r}) \right)^T \right], \tag{2.4}$$

which is an $(M+2) \times (M+2)$ matrix. $E_{\boldsymbol{\theta}}[\cdot]$ stands for the expectation conditioned on $\boldsymbol{\theta}$ and symbol "T" is for transpose. Denote $\hat{\boldsymbol{\theta}}$ as an estimate of the vector of parameters $\boldsymbol{\theta}$. The CRLB is then expressed as

$$E_{\boldsymbol{\theta}} \left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \right] \geq \mathbf{J}_{\boldsymbol{\theta}}^{-1}, \tag{2.5}$$

where "$\mathbf{A} \geq \mathbf{B}$" should be interpreted as matrix $(\mathbf{A} - \mathbf{B})$ is non-negative definite. The CRLB of Eq. (2.5) represented in matrix form provides a lower bound on the mean-square errors for the individual components of $\boldsymbol{\theta}$, i.e.,

$$E(\hat{\theta}_b - \theta_b)^2 \geq \left[\mathbf{J}_{\boldsymbol{\theta}}^{-1} \right]_{bb}, \quad \text{for } 1 \leq b \leq M+2, \tag{2.6}$$

which is due to the property of a non-negative definite matrix that all of its diagonal terms are non-negative.

To evaluate the CRLB, we first observe that $f_{\boldsymbol{\theta}}(\boldsymbol{r})$ of Eq. (2.3) is function of τ_b's that in turn are functions of the parameters $\boldsymbol{\theta}$ as stated in Eq. (2.1). Thus we are able to decompose $\mathbf{J}_{\boldsymbol{\theta}}$ of Eq. (2.4) using the *chain rule* as

$$\mathbf{J}_{\boldsymbol{\theta}} = \mathbf{H} \cdot \mathbf{J}_{\boldsymbol{\tau}} \cdot \mathbf{H}^T, \tag{2.7}$$

where

$$\mathbf{H} = \begin{pmatrix} \frac{\partial \tau_1}{\partial x} & \frac{\partial \tau_2}{\partial x} & \cdots & \frac{\partial \tau_M}{\partial x} & \cdots & \frac{\partial \tau_B}{\partial x} \\ \frac{\partial \tau_1}{\partial y} & \frac{\partial \tau_2}{\partial y} & \cdots & \frac{\partial \tau_M}{\partial y} & \cdots & \frac{\partial \tau_B}{\partial y} \\ \frac{\partial \tau_1}{\partial l_1} & \frac{\partial \tau_2}{\partial l_1} & \cdots & \frac{\partial \tau_M}{\partial l_1} & \cdots & \frac{\partial \tau_B}{\partial l_1} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial \tau_1}{\partial l_M} & \frac{\partial \tau_2}{\partial l_M} & \cdots & \frac{\partial \tau_M}{\partial l_M} & \cdots & \frac{\partial \tau_B}{\partial l_M} \end{pmatrix} \tag{2.8}$$

is an $(M+2)\times B$ matrix, and $\mathbf{J}_{\boldsymbol{\tau}}$ is the FIM conditioned on $\boldsymbol{\tau}$:

$$\mathbf{J}_{\boldsymbol{\tau}} = E_{\boldsymbol{\tau}}\left[\frac{\partial}{\partial \boldsymbol{\tau}} \ln f_{\boldsymbol{\theta}}(\boldsymbol{r}) \cdot \left(\frac{\partial}{\partial \boldsymbol{\tau}} \ln f_{\boldsymbol{\theta}}(\boldsymbol{r})\right)^T\right]. \tag{2.9}$$

The matrix $\mathbf{H}$ summarizes information regarding the geometric configuration among the MS and the BS's. It can be decomposed into NLOS and LOS components:

$$\mathbf{H} = \frac{1}{c}\begin{pmatrix} \mathbf{H}_{NL} & \mathbf{H}_{L} \\ \mathbf{I}_M & \mathbf{0} \end{pmatrix}, \tag{2.10}$$

where $\mathbf{I}_M$ is the identity matrix of order M, $\mathbf{H}_{NL}$ and $\mathbf{H}_L$ are $2\times M$ and $2\times(B-M)$ matrices, respectively, expressed as

$$\mathbf{H}_{NL} = \begin{pmatrix} \cos\phi_1 & \cos\phi_2 & \cdots & \cos\phi_M \\ \sin\phi_1 & \sin\phi_2 & \cdots & \sin\phi_M \end{pmatrix},$$

and

$$\mathbf{H}_{L} = \begin{pmatrix} \cos\phi_{M+1} & \cos\phi_{M+2} & \cdots & \cos\phi_B \\ \sin\phi_{M+1} & \sin\phi_{M+2} & \cdots & \sin\phi_B \end{pmatrix}.$$

Angle ϕ_b is determined by

$$\phi_b = \tan^{-1}\frac{y-y_b}{x-x_b},$$

which is the geometric angle between the positions of the MS and BS_b.

Subscript "NL" or "L" refers to NLOS or LOS stations, respectively. We shall adopt this notation throughout this chapter. Similar to $\mathbf{H}$ of Eq. (2.10), we can decompose $\mathbf{J}_{\boldsymbol{\tau}}$ as

$$\mathbf{J}_{\boldsymbol{\tau}} = \begin{pmatrix} \boldsymbol{\Lambda}_{NL} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{L} \end{pmatrix}, \tag{2.11}$$

where $\boldsymbol{\Lambda}_{NL}$ and $\boldsymbol{\Lambda}_L$ are diagonal matrices of order M and $(B-M)$, respectively, given by

$$\boldsymbol{\Lambda}_{NL} = \mathrm{diag}\left(\lambda_1,\ \lambda_2,\ \cdots, \lambda_M\right), \tag{2.12}$$

and

$$\mathbf{\Lambda}_L = \mathrm{diag}\left(\lambda_{M+1},\ \lambda_{M+2},\ \cdots, \lambda_B\right). \tag{2.13}$$

The diagonal term is

$$\lambda_b = 8\pi^2\beta^2 \cdot R_b, \tag{2.14}$$

where R_b is the SNR of the received signal at BS_b,

$$R_b = \frac{\int |A_b s(t)|^2 dt}{N_0} = \frac{A_b^2}{N_0}.$$

We assume the normalization condition

$$\int |s(t)|^2 dt = 1, \tag{2.15}$$

for simplicity. Parameter β is the effective bandwidth of the signal waveform $s(t)$, defined as

$$\beta^2 = \frac{\int f^2 |S(f)|^2 df}{\int |S(f)|^2 df} = \int f^2 |S(f)|^2 df,$$

where $S(f)$ is the Fourier transform of $s(t)$. Note that $\mathbf{J}_{\boldsymbol{\tau}}$ contains the system parameters.

By now we have calculated the two components of $\mathbf{J}_{\boldsymbol{\theta}}$ in Eq. (2.7). It is straightforward to compute

$$\mathbf{J}_{\boldsymbol{\theta}} = \frac{1}{c^2}\begin{pmatrix} \mathbf{H}_{NL}\mathbf{\Lambda}_{NL}\mathbf{H}_{NL}^T + \mathbf{H}_L\mathbf{\Lambda}_L\mathbf{H}_L^T & \mathbf{H}_{NL}\mathbf{\Lambda}_{NL} \\ \mathbf{\Lambda}_{NL}\mathbf{H}_{NL}^T & \mathbf{\Lambda}_{NL} \end{pmatrix}. \tag{2.16}$$

From the above equation, we see that $\mathbf{J}_{\boldsymbol{\theta}}$ depends on contributions from both NL (NLOS) and L (LOS) signals. The inverse of $\mathbf{J}_{\boldsymbol{\theta}}$, i.e., the CRLB, gives the best estimation accuracy of $\boldsymbol{\theta} = (\boldsymbol{p}, \boldsymbol{l})$. However, the accuracy of $\boldsymbol{p}$ is of our primary interest. Thus we shall concentrate on $\left[\mathbf{J}_{\boldsymbol{\theta}}^{-1}\right]_{2\times 2}$, which is the first 2×2 diagonal submatrix of $\mathbf{J}_{\boldsymbol{\theta}}^{-1}$. An interesting characteristic of this quantity is stated as below.

Proposition 2.1

When no prior information on $\boldsymbol{\theta}$ is available, the CRLB for the MS position is

$$\left[\mathbf{J}_{\boldsymbol{\theta}}^{-1}\right]_{2\times 2} = c^2 \left(\mathbf{H}_L \boldsymbol{\Lambda}_L \mathbf{H}_L^T\right)^{-1}, \tag{2.17}$$

which depends only on LOS signals.

Proof.

It suffices to show

$$\left[\mathbf{J}_{\boldsymbol{\theta}}^{-1}\right]_{11} = c^2 \left[\left(\mathbf{H}_L \Sigma_L \mathbf{H}_L^T\right)^{-1}\right]_{11}, \tag{2.18}$$

where $[\mathbf{A}]_{mn}$ refers to the element at the m-th row and n-th column of matrix $\mathbf{A}$. It can be done in the following four steps.

1. By definition,

$$\left[\mathbf{J}_{\boldsymbol{\theta}}^{-1}\right]_{11} = \frac{\left|\widetilde{\mathbf{J}}_{\boldsymbol{\theta}(1,1)}\right|}{\left|\mathbf{J}_{\boldsymbol{\theta}}\right|}, \tag{2.19}$$

   where $\widetilde{\mathbf{J}}_{\boldsymbol{\theta}(1,1)}$ is an $(M+1)\times(M+1)$ matrix obtained by removing the first row and the first column of $\mathbf{J}_{\boldsymbol{\theta}}$, and $|\mathbf{A}|$ is the determinant of matrix $\mathbf{A}$.

2. The denominator in the above equation can be expressed in terms of $\mathbf{H}_L$, $\boldsymbol{\Lambda}_L$ and $\boldsymbol{\Lambda}_{NL}$ by exploring submatrix manipulation:

$$\begin{aligned}
\left|\mathbf{J}_{\boldsymbol{\theta}}\right| &= \frac{1}{c^{2B}} \begin{vmatrix} \mathbf{H}_{NL}\boldsymbol{\Lambda}_{NL}\mathbf{H}_{NL}^T + \mathbf{H}_L\boldsymbol{\Lambda}_L\mathbf{H}_L^T & \mathbf{H}_{NL}\boldsymbol{\Lambda}_{NL} \\ \boldsymbol{\Lambda}_{NL}\mathbf{H}_{NL}^T & \boldsymbol{\Lambda}_{NL} \end{vmatrix} \\
&= \frac{1}{c^{2B}} \begin{vmatrix} \mathbf{H}_L\boldsymbol{\Lambda}_L\mathbf{H}_L^T & \mathbf{0} \\ \boldsymbol{\Lambda}_{NL}\mathbf{H}_{NL}^T & \boldsymbol{\Lambda}_{NL} \end{vmatrix} \qquad (2.20) \\
&= \frac{1}{c^{2B}} \left|\mathbf{H}_L\boldsymbol{\Lambda}_L\mathbf{H}_L^T\right| \cdot \left|\boldsymbol{\Lambda}_{NL}\right|. \qquad (2.21)
\end{aligned}$$

The second equality is obtained by subtracting the second row left multiplied by $\mathbf{H}_{NL}$ from the first row. The determinant of the matrix remains unchanged under such linear operation [33].

3. Similarly, we derive the nominator of Eq. (2.19):

$$\begin{aligned}
\left|\tilde{\mathbf{J}}_{\boldsymbol{\theta}(1,1)}\right| &= \frac{1}{c^{2B-2}}\left|\begin{pmatrix} \widetilde{\mathbf{H}}_{NL(1,0)} & \widetilde{\mathbf{H}}_{L(1,0)} \\ \mathbf{I}_M & \mathbf{0} \end{pmatrix}\begin{pmatrix} \boldsymbol{\Lambda}_{NL} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_L \end{pmatrix}\begin{pmatrix} \widetilde{\mathbf{H}}^T_{NL(1,0)} & \mathbf{I}_M \\ \widetilde{\mathbf{H}}^T_{L(1,0)} & \mathbf{0} \end{pmatrix}\right| \\
&= \frac{1}{c^{2B-2}}\begin{vmatrix} \widetilde{\mathbf{H}}_{NL(1,0)}\boldsymbol{\Lambda}_{NL}\widetilde{\mathbf{H}}^T_{NL(1,0)} + \widetilde{\mathbf{H}}_{L(1,0)}\boldsymbol{\Lambda}_L\widetilde{\mathbf{H}}^T_{L(1,0)} & \widetilde{\mathbf{H}}_{NL(1,0)}\boldsymbol{\Lambda}_{NL} \\ \boldsymbol{\Lambda}_{NL}\widetilde{\mathbf{H}}^T_{NL(1,0)} & \boldsymbol{\Lambda}_{NL} \end{vmatrix} \\
&= \frac{1}{c^{2B-2}}\left|\widetilde{\mathbf{H}}_{L(1,0)}\boldsymbol{\Lambda}_L\widetilde{\mathbf{H}}^T_{L(1,0)}\right| \cdot \left|\boldsymbol{\Lambda}_{NL}\right|, \qquad (2.22)
\end{aligned}$$

where $\widetilde{\mathbf{H}}_{NL(1,0)}$ and $\widetilde{\mathbf{H}}_{L(1,0)}$ are obtained by deleting the first row of $\mathbf{H}_{NL}$ and $\mathbf{H}_L$, respectively.

4. Substituting $\left|\mathbf{J}_{\boldsymbol{\theta}}\right|$ and $\left|\tilde{\mathbf{J}}_{\boldsymbol{\theta}(1,1)}\right|$ of Eqs. (2.21) and (2.22) into Eq. (2.19), we reach

$$\begin{aligned}
\left[\mathbf{J}^{-1}_{\boldsymbol{\theta}}\right]_{11} &= c^2 \cdot \frac{\left|\widetilde{\mathbf{H}}_{L(1,0)}\boldsymbol{\Lambda}_L\widetilde{\mathbf{H}}^T_{L(1,0)}\right|}{\left|\mathbf{H}_L\boldsymbol{\Lambda}_L\mathbf{H}_L^T\right|} \\
&= c^2\left[\left(\mathbf{H}_L\boldsymbol{\Lambda}_L\mathbf{H}_L^T\right)^{-1}\right]_{11}. \qquad (2.23)
\end{aligned}$$

By repeating the above steps for the other elements of $\left[\mathbf{J}^{-1}_{\boldsymbol{\theta}}\right]_{2\times 2}$, we obtain

$$\left[\mathbf{J}^{-1}_{\boldsymbol{\theta}}\right]_{mn} = c^2\left[\left(\mathbf{H}_L\boldsymbol{\Lambda}_L\mathbf{H}_L^T\right)^{-1}\right]_{mn}, \quad \text{for } m = 1, 2, \text{ and } n = 1, 2. \qquad (2.24)$$

□

From the above derivation, we are convinced that the CRLB for the MS position estimate depends only on LOS signals, and the contributions from the NLOS

signals are completely eliminated. This somewhat surprising and counter-intuitive consequence results from our assumption that no prior information is available concerning the NLOS delays $\boldsymbol{l}$. It is equivalent to assuming a prior distribution with *infinite variance* for $\boldsymbol{l}$. Then any possible improvement in the positioning precision to be made by including the NLOS signals is annulled by the infinite variance (or "total uncertainty") of $\boldsymbol{l}$. We shall revisit this argument from another perspective in Section 2.2.4.

Physical Interpretation

Now we are in a position to provide a physical interpretation of the CRLB in Proposition 2.1. It will be convenient for our discussion if we use the explicit expression of the minimum mean square error of the MS position estimate $\hat{\boldsymbol{p}} = (\hat{x}, \hat{y})$ instead:

$$\begin{aligned} \left[E_{\boldsymbol{\theta}}(\boldsymbol{p} - \widehat{\boldsymbol{p}})^2\right]_{\min} &= \left[E_{\boldsymbol{\theta}}(x - \hat{x})^2\right]_{\min} + \left[E_{\boldsymbol{\theta}}(y - \hat{y})^2\right]_{\min} \\ &= \left[\mathbf{J}_{\boldsymbol{\theta}}^{-1}\right]_{11} + \left[\mathbf{J}_{\boldsymbol{\theta}}^{-1}\right]_{22} \\ &= \frac{c^2}{8\pi^2\beta^2} \cdot \frac{\sum_{b\in\mathcal{L}} R_b}{\sum \sum_{b_1,b_2\in\mathcal{L}} R_{b_1} R_{b_2} \sin^2(\phi_{b_1} - \phi_{b_2})}. \end{aligned} \tag{2.25}$$

First, the exclusive dependency on the LOS signals of the positioning accuracy can be exploited to simplify NLOS geolocation algorithms. When no prior information on NLOS delays is available, an algorithm should identify and then reject all NLOS signals, instead of actually estimating the NLOS induced path length $\boldsymbol{l}$ along with $\boldsymbol{p} = (x, y)$. This approach usually reduces computational burden, as will be verified in Chapter 4 where we consider concrete examples.

Second, $\left[E_{\boldsymbol{\theta}}(\boldsymbol{p} - \widehat{\boldsymbol{p}})^2\right]_{\min}$ is inversely proportional to the square of the effective bandwidth, β^2, which is consistent with the MMSE (minimum mean square error) of a distance estimate in radar estimation (see e.g., [9]).

Third, the positioning accuracy is controlled by the geometric configuration of the LOS stations, only through sine functions of the angle differences $(\phi_{b_1} - \phi_{b_2}),\ b_1, b_2 \in \mathcal{L}$, seen by the MS. The estimation error grows to infinity when all the LOS stations and the MS are lined up on a straight line, i.e.,

$$\left[E_{\boldsymbol{\theta}}(\boldsymbol{p} - \widehat{\boldsymbol{p}})^2\right]_{\min} \longrightarrow +\infty, \quad \text{as} \quad \phi_{b_1} - \phi_{b_2} = 0, \pi, \quad \text{for all } b_1, b_2 \in \mathcal{L}. \tag{2.26}$$

Fortunately, such a case is rare in reality.

Fourth, we investigate the contribution of an individual SNR, say R_k, to an overall accuracy enhancement. To facilitate the investigation, we isolate R_k from the rest of $\left[E_{\boldsymbol{\theta}}(\boldsymbol{p} - \widehat{\boldsymbol{p}})^2\right]_{\min}$ given by Eq. (2.25) as

$$\left[E_{\boldsymbol{\theta}}(\boldsymbol{p} - \widehat{\boldsymbol{p}})^2\right]_{\min} = \frac{c^2}{8\pi^2\beta^2} \cdot \frac{1}{C}\left\{1 + \frac{AC - G}{G + CR_k}\right\}, \tag{2.27}$$

where

$$\begin{aligned} A &= \sum_{b \neq k,\ b \in \mathcal{L}} R_b \geq 0, \\ G &= \sum_{b_1, b_2 \neq k,\ b \in \mathcal{L}} \sum R_{b_1} R_{b_2} \sin^2(\phi_{b_1} - \phi_{b_2}) \geq 0, \\ C &= \sum_{b \neq k,\ b \in \mathcal{L}} R_b \sin^2(\phi_b - \phi_k) \geq 0. \end{aligned}$$

Observing the above expression, we notice that the sign of $(AC - G)$ is important. We claim $(AC - G) \geq 0$, where the equality holds if and only if $\phi_b - \phi_k = 0$ or π, for $b \neq k,\ b \in \mathcal{L}$. To see this, let

$$\xi_b = \phi_b - \phi_k, \quad \text{for } b \neq k.$$

Then,

$$AC - G = \sum_{b_1, b_2 \neq k} \sum R_{b_1} R_{b_2} \left(\sin^2 \xi_{b_2} - \sin^2(\xi_{b_1} - \xi_{b_2})\right)$$

$$
\begin{aligned}
&= \sum_{b\neq k} R_b^2 \sin^2\xi_b + 2\sum \sum_{b_1<b_2,\ b_1,b_2\neq k} R_{b_1}R_{b_2}\cdot(\sin^2\xi_{b_1}\sin^2\xi_{b_2} + \sin\xi_{b_1}\sin\xi_{b_2}\cos\xi_{b_1}\cos\xi_{b_2}) \\
&\geq \left(\sum_{b\neq k} R_b^2 \sin^2\xi_b\cos^2\xi_b + 2\sum \sum_{b_1<b_2,\ b_1,b_2\neq k} R_{b_1}R_{b_2}\sin\xi_{b_1}\sin\xi_{b_2}\cos\xi_{b_1}\cos\xi_{b_2}\right) \\
&\quad +2\sum \sum_{b_1<b_2,\ b_1,b_2\neq k} R_{b_1}R_{b_2}\sin^2\xi_{b_1}\sin^2\xi_{b_2} \\
&\quad \left(\text{using } \textstyle\sum_{b\neq k} R_b^2\sin^2\xi_b \geq \sum_{b\neq k} R_b^2\sin^2\xi_b\cos^2\xi_b\right) \\
&= \left(\frac{1}{2}\sum_{b\neq k} R_b\sin(2\xi_b)\right)^2 + 2\sum \sum_{b_1<b_2,\ b_1,b_2\neq k} R_{b_1}R_{b_2}\sin^2\xi_{b_1}\sin^2\xi_{b_2} \\
&\geq 0. \qquad (2.28)
\end{aligned}
$$

The above "if and only if" condition is not difficult to verify. Therefore, we are able to conclude from Eq. (2.27) that when $(AC-G)>0$, an increase in R_k should enhance the geolocation performance; however, no such gain is expected when $AC-G=0$, which corresponds to the case where Eq. (2.26) holds, i.e., when all the LOS stations and the MS align on a straight line.

Numerical Examples:

Consider a cellular CDMA system with seven BSs as shown in Figure 2.1, with cell radius of 2000m. A moving MS transmits signals. The SNR of a received signal at a BS is tuned to be $R_0 = 3$dB when the distance between the pair of MS and BS is 2000m. The chip rate of the CDMA signal is $W = 5$Mcps.

The relation between W and the effective bandwidth β can be approximated as

$$\beta \cong \frac{W}{\sqrt{3}}. \qquad (2.29)$$

It can be derived from the normalization condition of Eq. (2.15)

$$\int_{-\infty}^{\infty} |S(f)|^2\, df = 1,$$

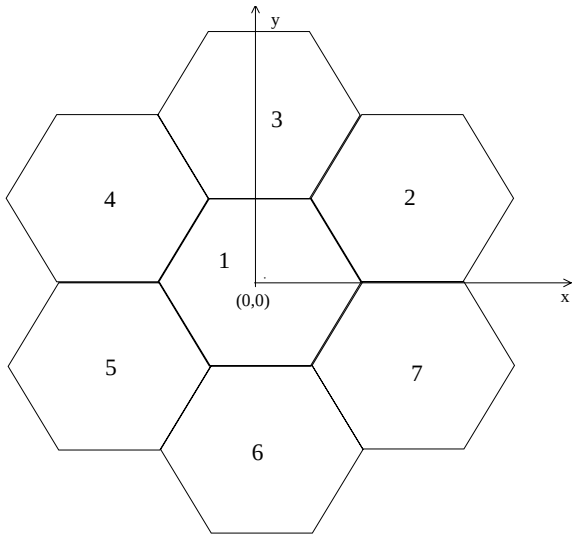


Figure 2.1: A Cellular system with seven base stations.

and approximation

$$|S(f)|^2 \cong |S(0)|^2, \quad \text{for } f \in [-W,\ W],$$

which yields

$$|S(0)|^2 \cong \frac{1}{2W}.$$

Hence,

$$\begin{aligned} \beta^2 &\cong \int_{-W}^{W} f^2 |S(f)|^2 \, df \\ &\cong |S(0)|^2 \cdot \frac{1}{3} \, f^3 \Big|_{-W}^{W} \\ &= \frac{W^2}{3}. \end{aligned}$$

Therefore, we are able to calculate the positioning accuracy (in meter) for the CDMA system, by using the square root of $\left[E_{\boldsymbol{\theta}}(\boldsymbol{p}-\widehat{\boldsymbol{p}})^2\right]_{\min}$ in Eq. (2.25), or equivalently,

$$\mathcal{P}_{CR} \stackrel{\text{def}}{=} \sqrt{\text{trace}\left(\left[\mathbf{J}_{\boldsymbol{\theta}}^{-1}\right]_{2\times 2}\right)}. \tag{2.30}$$

Figure 2.2 depicts a contour map of the accuracy $\mathcal{P}_{CR}$ for the CDMA system with all the seven BSs being LOS stations. The locations of the LOS stations are marked as "*", and the numbers labeled on the contour curves represent the positioning accuracy. It is observed that first, the contour pattern is symmetric due to the set of evenly distributed BSs; second, the better accuracy values appear in the central region that surrounds the central BS at (0,0).

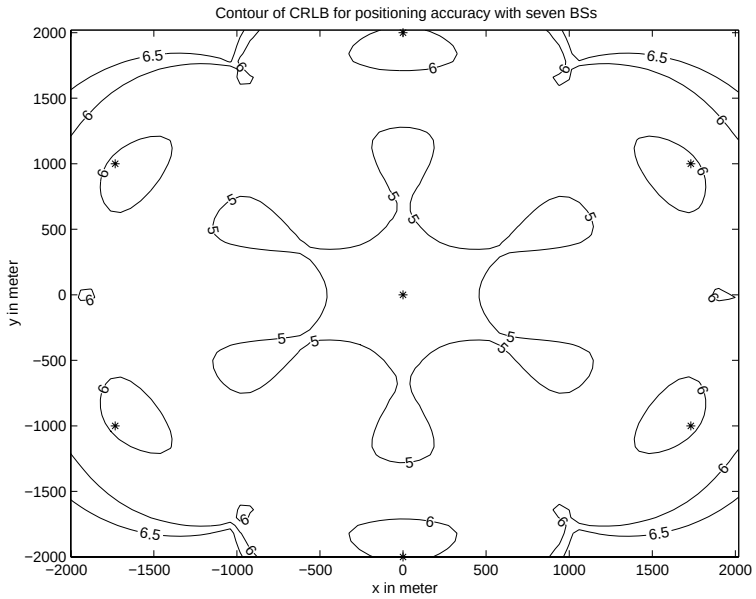


Figure 2.2: The contour of the accuracy $\mathcal{P}_{CR}$ (in meter) for the CDMA system with all the seven BSs being LOS stations.

Figures 2.3 and 2.4 show the accuracy $\mathcal{P}_{CR}$ vs. the chip rate and the SNR at a position 2000m from the MS, respectively. As expected, a better geolocation accuracy is achieved with a higher SNR, a higher chip rate and with more BSs involved. The performance may not further improve, however, beyond some threshold about 5Mcps in Figure 2.3.

At the end of this section, we briefly discuss the accuracy limit for the NLOS delay estimate, which may be useful to some applications such as channel estimation.

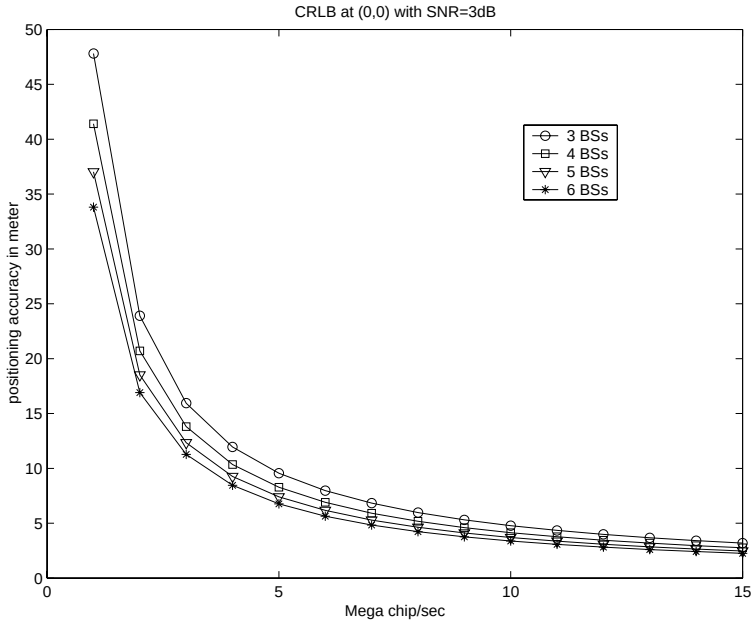


Figure 2.3: The numerical curves for the CRLB: $\mathcal{P}_{CR}$ vs. the chip rate with different numbers of BSs involved.

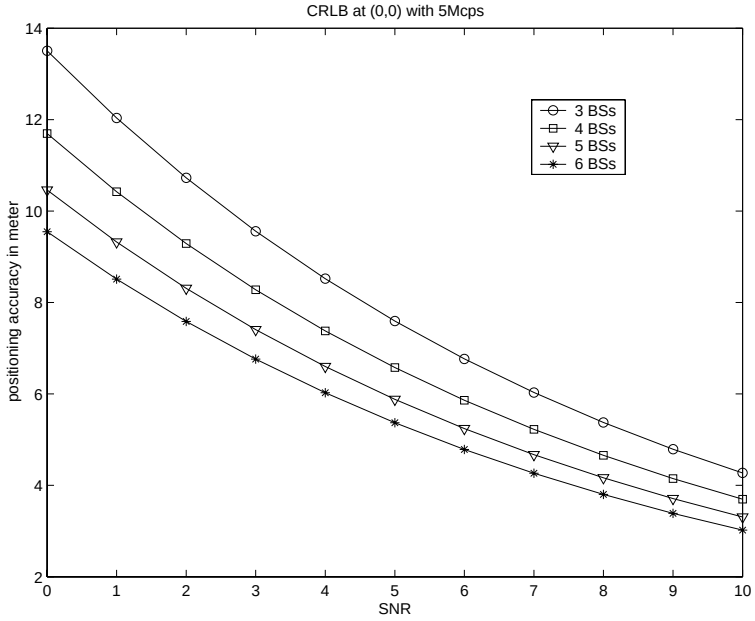


Figure 2.4: The numerical curves for the CRLB: $\mathcal{P}_{CR}$ vs. the SNR at a position 2000m from the MS with various numbers of BSs involved.

Analogous to the development of the CRLB for $\boldsymbol{p}$, we find the CRLB for estimating an l_m, the m-th NLOS induced path length, given by

$$\begin{aligned}\left[E_{\boldsymbol{\theta}}(l_m-\widehat{l}_m)^2\right]_{\min} &= \left[\mathbf{J}_{\boldsymbol{\theta}}^{-1}\right]_{(m+2),(m+2)} \\ &= \frac{c^2}{8\pi^2\beta^2 R_m}\cdot\frac{\sum\sum_{b_1,b_2\in\mathcal{L}\cup\{m\}} R_{b_1}R_{b_2}\sin^2(\phi_{b_1}-\phi_{b_2})}{\sum\sum_{b_1,b_2\in\mathcal{L}} R_{b_1}R_{b_2}\sin^2(\phi_{b_1}-\phi_{b_2})}, \\ &\geq \frac{c^2}{8\pi^2\beta^2 R_m}, \quad \text{for } m\in\mathcal{NL}. \end{aligned} \tag{2.31}$$

We notice that only the LOS signals and the NLOS signal associated with this specific BS_m matter, and the signals from the other $(M-1)$ NLOS BSs should be excluded from the estimation procedure. We also see that $\left[E_{\boldsymbol{\theta}}(l_m-\widehat{l}_m)^2\right]_{\min}$ is lower bounded by $c^2/8\pi^2\beta^2R_m$, which is the CRLB for estimating the overall propagation path from the MS to BS_m as we shall see in the next section.

2.2.3 Maximum Likelihood Estimation

So far we have investigated the CRLB for TOA positioning in an NLOS environment. We now raise the next question: how can we achieve or asymptotically achieve the accuracy suggested by the CRLB? In this section, we probe this question in two stages, which are based on the following known result[1].

If the time delay of the received signal of Eq. (2.2) is estimated at a matched filter output with large $R_b\cdot\beta^2$, it can be modeled as

$$\hat{\tau}_b = \tau_b + \xi_b, \quad \text{for } b\in\mathcal{B}, \tag{2.32}$$

where ξ_b is a Gaussian random variable $\mathcal{N}(0,\sigma_b^2)$ with $\sigma_b^2 = 1/8\pi^2\beta^2R_b$.

[1] A more general result was proved in the context of range and velocity estimation in a radar problem in [9].

A direct way to see this is to apply the Taylor series expansion to the matched filter output as shown in Appendix 2.1.

The two stages of exploring the achievability of the CRLB are represented by the following lemma and proposition.

Lemma 2.1

The CRLB for the MS position estimate, which is based on the time delay estimates from LOS stations, i.e., $\{\hat{\tau}_b,\ b \in \mathcal{L}\}$*, is equivalent to* $\left[\mathbf{J}_{\boldsymbol{\theta}}^{-1}\right]_{2\times 2}$ *of Eq. (2.17).*

Proof.

We compute the Fisher Information Matrix based on the p.d.f. of $\{\hat{\tau}_b,\ b \in \mathcal{L}\}$ conditioned on $\boldsymbol{p}$, i.e.,

$$f_{\boldsymbol{p}}\left(\hat{\boldsymbol{\tau}}\right) \propto \prod_{b\in\mathcal{L}} \exp\left\{-\frac{1}{2\sigma_b^2}(\hat{\tau}_b - \tau_b)^2\right\}, \tag{2.33}$$

with $\sigma_b^2 = 1/4\pi^2\beta^2 R_b$, as

$$\mathbf{J}_{\boldsymbol{p}} = \frac{1}{c^2}\mathbf{H}_L\boldsymbol{\Lambda}_L\mathbf{H}_L^T. \tag{2.34}$$

Compared the above expression with Eq. (2.17) of Proposition 2.1, it is straightforward to verify

$$\mathbf{J}_{\boldsymbol{p}}^{-1} = \left[\mathbf{J}_{\boldsymbol{\theta}}^{-1}\right]_{2\times 2}.$$

□

We now can answer the achievability question raised at the beginning of the section:

Proposition 2.2

The maximum likelihood estimator (MLE) based on the time delay estimates from LOS stations is unbiased, and can attain the CRLB for the MS position estimate in an NLOS environment, i.e., $\left[\mathbf{J}_{\boldsymbol{\theta}}^{-1}\right]_{2\times 2}$*, as the SNR and/or* β^2 *increase to infinity.*

Proof.

We first show there exists an unbiased estimator $\hat{\boldsymbol{p}}$ as the SNR and/or β^2 increase to infinity.

1. Recall the time delay is

$$\tau_b = \frac{1}{c}\left\{\sqrt{(x_b - x)^2 + (y_b - y)^2}\right\}, \quad b \in \mathcal{L}.$$

We differentiate both sides of the above equation, and express the result in matrix notation:

$$\Delta\boldsymbol{\tau} = \frac{1}{c}\mathbf{H}_L^T \, \Delta\boldsymbol{p}, \tag{2.35}$$

where

$$\Delta\boldsymbol{\tau} = \begin{pmatrix} \Delta\tau_1 \\ \Delta\tau_2 \\ \vdots \\ \Delta\tau_L \end{pmatrix}, \text{ and } \Delta\boldsymbol{p} = \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}.$$

Considering the physical meaning of $\Delta\tau_b$, we can interpret $\Delta\tau_b$ as the error of the TOA estimate in Eq. (2.32), i.e.,

$$\Delta\tau_b = \xi_b, \quad b \in \mathcal{L}.$$

Note the above expression is based on the assumption that the error ξ_b, or the variance of ξ_b, is small enough, which requires that $R_b \cdot \beta^2$ be sufficiently large (see Eq. (2.32)). Since

$$\Delta\boldsymbol{\tau} = \hat{\boldsymbol{\tau}} - \boldsymbol{\tau},$$

and

$$\Delta\boldsymbol{p} = \hat{\boldsymbol{p}} - \boldsymbol{p},$$

Eq. (2.35) becomes

$$\hat{\boldsymbol{\tau}} - \boldsymbol{\tau} = \frac{1}{c}\mathbf{H}_L^T\,(\hat{\boldsymbol{p}} - \boldsymbol{p}). \tag{2.36}$$

2. It is convenient for our further discussion if $f_{\boldsymbol{p}}(\hat{\boldsymbol{\tau}})$ of Eq. (2.34) is expressed in matrix notation:

$$f_{\boldsymbol{p}}(\hat{\boldsymbol{\tau}}) \propto \exp\left\{-\frac{1}{2}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau})^T \boldsymbol{\Lambda}_L (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau})\right\}, \tag{2.37}$$

where $\boldsymbol{\Lambda}_L$ was defined in Eq. (2.11). Substituting Eq. (2.36) into Eq. (2.37), we have

$$f_{\boldsymbol{p}}(\hat{\boldsymbol{\tau}}) \propto \exp\left\{-\frac{1}{2c^2}(\hat{\boldsymbol{p}} - \boldsymbol{p})^T \mathbf{H}_L \boldsymbol{\Lambda}_L \mathbf{H}_L^T (\hat{\boldsymbol{p}} - \boldsymbol{p})\right\}. \tag{2.38}$$

Note that the term

$$\frac{1}{c^2}\mathbf{H}_L \boldsymbol{\Lambda}_L \mathbf{H}_L^T$$

in the above equation is exactly the FIM, $[\mathbf{J}_{\boldsymbol{\theta}}]_{2\times 2}$, in Eq. (2.17).

3. Since $\hat{\boldsymbol{p}}$ is a multi-variate Gaussian variable with mean $\boldsymbol{p}$ as shown in Eq. (2.38), i.e.,

$$E_{\boldsymbol{p}}(\hat{\boldsymbol{p}}) = \boldsymbol{p},$$

$\hat{\boldsymbol{p}}$ is unbiased.

From estimation theory [5], the CRLB can be attained by the MLE, denoted as $\hat{\boldsymbol{p}}_{ML}$, if and only if there exists an unbiased estimator $\hat{\boldsymbol{p}}$ such that the i-th component of $\hat{\boldsymbol{p}}$ satisfies

$$\hat{p}_i - p_i = \sum_{j\in\mathcal{L}} k_{ij}(\boldsymbol{p}) \cdot \frac{\partial}{\partial p_j} \ln f_{\boldsymbol{p}}(\hat{\boldsymbol{\tau}}), \quad \text{for } i = 1 \text{ and } 2, \tag{2.39}$$

where $k_{ij}(\boldsymbol{p})$ is a quantity that depends on $\boldsymbol{p}$ only, and the estimator $\hat{\boldsymbol{p}}$ is exactly $\hat{\boldsymbol{p}}_{ML}$. It is straightforward to verify that $\hat{\boldsymbol{p}}$ in Eq. (2.38) satisfies the above sufficient

and necessary condition, since

$$\hat{\boldsymbol{p}} - \boldsymbol{p} = -c^2 \left(\mathbf{H}_L \boldsymbol{\Lambda}_L \mathbf{H}_L^T \right)^{-1} \cdot \frac{\partial}{\partial \boldsymbol{p}} \ln f_{\boldsymbol{p}}(\hat{\boldsymbol{\tau}}), \tag{2.40}$$

which is the vector version of Eq. (2.39). Thus, we can replace $\hat{\boldsymbol{p}}$ by $\hat{\boldsymbol{p}}_{ML}$. □

As we see, no NLOS propagation related terms are involved throughout the above proof. The reason is that the CRLB for the MS position estimate is independent of any NLOS signals, as stated in Proposition 2.1.

In essence, the optimum geolocation receiver given in Proposition 2.2 consists of three simple components: (1) estimating the time delays with the matched filter method at all the BSs, (2) identifying and removing the NLOS estimates, and (3) estimating the MS position using the MLE based on the LOS time delays.

We now can clarify the relationship between the optimal MLE and the conventional method which is directly formulated from TOA estimates: for the optimal solution, it is still necessary to extract the TOA estimates first, which has the following two advantages. The TOA estimates can be used not only for the geolocation purpose but also for other applications such as channel estimation and data extraction. On the other hand, many existing techniques for time delay estimation, e.g., the generalized cross-correlation methods [19, 20] and subspace-based estimation [21, 22], may be adopted in a geolocation scheme.

2.2.4 Generalized-CRLB

In practice, it may be possible to predict the distribution of the NLOS delays based on some statistical scattering models [41, 42]. With such information, a higher positioning accuracy should be expected. Accordingly, we need to extend our previous

results by incorporating such extra information. The logical structure of the following development parallels that of the previous two sections: first obtaining the best accuracy limit, and then looking for a receiver that can attain the limit.

The accuracy limit now is given by the Generalized-CRLB (G-CRLB) [5]. Analogous to the relationship between the CRLB and the Fisher information matrix, the G-CRLB is defined as

$$E\left\{(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\right\} \geq \mathbf{J}^{-1}. \tag{2.41}$$

Here matrix $\mathbf{J}$ is what may be called the "generalized Fisher information matrix" and consists of two components,

$$\mathbf{J} = \mathbf{J}_D + \mathbf{J}_P, \tag{2.42}$$

where subscripts "D" and "P" stand for "data" and "prior" information, respectively. The first component $\mathbf{J}_D$ is given by

$$\mathbf{J}_D = E\left[\frac{\partial}{\partial\boldsymbol{\theta}}\log f_{\boldsymbol{\theta}}(\boldsymbol{r})\cdot\left(\frac{\partial}{\partial\boldsymbol{\theta}}\log f_{\boldsymbol{\theta}}(\boldsymbol{r})\right)^T\right], \tag{2.43}$$

where the expectation $E[\cdot]$ is taken over both $\boldsymbol{r}$ and $\boldsymbol{\theta}$. It pertains to the observation $\boldsymbol{r} = \{r_b(t), b \in \mathcal{B}\}$ as well as the prior probability of $\boldsymbol{\theta}$, $p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$. The second component is defined as

$$\mathbf{J}_P = E\left[\frac{\partial}{\partial\boldsymbol{\theta}}\log p_{\boldsymbol{\theta}}(\boldsymbol{\theta})\cdot\left(\frac{\partial}{\partial\boldsymbol{\theta}}\log p_{\boldsymbol{\theta}}(\boldsymbol{\theta})\right)^T\right], \tag{2.44}$$

where the expectation is over $\boldsymbol{\theta}$. However, out of two types of components of the unknown vector

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{p} \\ \boldsymbol{l} \end{pmatrix},$$

the first type of components $\boldsymbol{p}$ are constant parameters (although unknown), for which we have no prior information. In other words, the probabilistic information about $\boldsymbol{\theta}$

is provided only for the NLOS delay parameters $\boldsymbol{l}$, which is equivalent to setting

$$p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = p_{\boldsymbol{l}}(\boldsymbol{l}). \tag{2.45}$$

Compared with the CRLB which is conditioned on specific values of the parameters to be estimated, the G-CRLB utilizes the *priori* probability density of the parameters and provides a bound that is averaged over sufficiently many trials.

A. NLOS delays of Gaussian distribution

We first assume that the random variables $\boldsymbol{l}$ are independently Gaussian distributed with mean $\boldsymbol{u}_l$ and covariance matrix

$$\boldsymbol{\Omega} = \mathrm{diag}(\omega_1^2, \omega_2^2, \cdots, \omega_M^2). \tag{2.46}$$

Although the independence assumption and the Gaussian distribution may not hold in actual environments, this simplified model allows us gain some insight on how the extra information on $\boldsymbol{l}$ enhances the geolocation precision. We shall see this point clearly when we consider some realistic distributions, such as Gamma distributions, next.

Inserting the Gaussian p.d.f. into the expression of $\mathbf{J}_P$ in Eq. (2.44) yields

$$\mathbf{J}_P = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}^{-1} \end{pmatrix}. \tag{2.47}$$

To evaluate $\mathbf{J}_D$, we modify its expression of Eq. (2.43) in terms of the FIM $\mathbf{J}_{\boldsymbol{\theta}}$ of Eq. (2.16), as

$$\mathbf{J}_D = E\left(\mathbf{J}_{\boldsymbol{\theta}}\right), \tag{2.48}$$

where the expectation is taken over $\boldsymbol{l}$. Noticing that

$$\mathbf{J}_{\boldsymbol{\theta}} = \frac{1}{c^2} \begin{pmatrix} \mathbf{H}_{NL}\boldsymbol{\Lambda}_{NL}\mathbf{H}_{NL}^T + \mathbf{H}_L\boldsymbol{\Lambda}_L\mathbf{H}_L^T & \mathbf{H}_{NL}\boldsymbol{\Lambda}_{NL} \\ \boldsymbol{\Lambda}_{NL}\mathbf{H}_{NL}^T & \boldsymbol{\Lambda}_{NL} \end{pmatrix}$$

is actually independent of $\boldsymbol{l}$, we reach

$$\mathbf{J}_D = \mathbf{J}_{\boldsymbol{\theta}}. \tag{2.49}$$

Hence, we have obtained both components of $\mathbf{J}$ in Eq. (2.42).

We shall now focus our discussion on $[\mathbf{J}^{-1}]_{2\times 2}$, which is the first 2×2 diagonal sub-matrix of $\mathbf{J}^{-1}$. The explicit expression of $[\mathbf{J}^{-1}]_{2\times 2}$ is rather complex (see the proof below), thus we will instead present its lower and upper bounds, from which we can gain a better intuitive understanding of the physical meaning on $[\mathbf{J}^{-1}]_{2\times 2}$:

Proposition 2.3

The G-CRLB for the MS position estimate, $[\mathbf{J}^{-1}]_{2\times 2}$, has the following upper and lower bounds

$$c^2 \left(\mathbf{H}_L \boldsymbol{\Lambda}_L \mathbf{H}_L^T\right)^{-1} \geq [\mathbf{J}^{-1}]_{2\times 2} \geq c^2 \left(\mathbf{H}_{NL} \boldsymbol{\Lambda}_{NL} \mathbf{H}_{NL}^T + \mathbf{H}_L \boldsymbol{\Lambda}_L \mathbf{H}_L^T\right)^{-1}. \tag{2.50}$$

The lower bound is attained when $\omega_m^2 \longrightarrow 0$, for $1 \leq m \leq M$, where ω_m^2's are the diagonal term of the covariance matrix $\boldsymbol{\Omega}$ of the Gaussian variables $\boldsymbol{l}$. The upper bound is achieved when all $\omega_m^2 \longrightarrow +\infty$.

Proof.

We expand the expression of $\mathbf{J}$ as

$$\mathbf{J} = \frac{1}{c^2} \begin{pmatrix} \mathbf{H}_{NL} \boldsymbol{\Lambda}_{NL} \mathbf{H}_{NL}^T + \mathbf{H}_L \boldsymbol{\Lambda}_L \mathbf{H}_L^T & \mathbf{H}_{NL} \boldsymbol{\Lambda}_{NL} \\ \boldsymbol{\Lambda}_{NL} \mathbf{H}_{NL}^T & \boldsymbol{\Lambda}_{NL} + c^2 \boldsymbol{\Omega}^{-1} \end{pmatrix} \tag{2.51}$$

$$\stackrel{\text{def}}{=} \frac{1}{c^2} \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{pmatrix}. \tag{2.52}$$

We employ the matrix inverse formula [10]

$$\mathbf{J}^{-1} = c^2 \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{F}\mathbf{W}^{-1}\mathbf{F}^T & -\mathbf{F}\mathbf{W}^{-1} \\ -\mathbf{W}^{-1}\mathbf{F}^T & \mathbf{W}^{-1} \end{pmatrix}, \tag{2.53}$$

where

$$\mathbf{W} = \mathbf{C} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}, \quad \mathbf{F} = \mathbf{A}^{-1}\mathbf{B},$$

and all the inverses that occur in the above expressions exist. Then, It is clear that

$$[\mathbf{J}^{-1}]_{2\times 2} = c^2\mathbf{A}^{-1} + c^2\mathbf{F}\mathbf{W}^{-1}\mathbf{F}^T, \tag{2.54}$$

A further discussion on the above equation proceeds as follows.

1. We first establish the non-negative definiteness of

$$\mathbf{W} = c^2\mathbf{\Omega}^{-1} + \left(\mathbf{\Lambda}_{NL} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}\right).$$

It is clear that $\mathbf{\Omega}^{-1} \geq 0$. We can also show

$$\begin{aligned}
\mathbf{\Lambda}_{NL} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B} &= \mathbf{\Lambda}_{NL} - \mathbf{B}^T\left(\mathbf{H}_{NL}\mathbf{\Lambda}_{NL}\mathbf{H}_{NL}^T + \mathbf{H}_L\mathbf{\Lambda}_L\mathbf{H}_L^T\right)^{-1}\mathbf{B} \\
&\geq \mathbf{\Lambda}_{NL} - \mathbf{\Lambda}_{NL}\mathbf{H}_{NL}^T\left(\mathbf{H}_{NL}\mathbf{\Lambda}_{NL}\mathbf{H}_{NL}^T\right)^{-1}\mathbf{H}_{NL}\mathbf{\Lambda}_{NL} \\
&\geq 0.
\end{aligned} \tag{2.55}$$

The last inequality is obtained by applying $\mathbf{Q} = \mathbf{\Lambda}_{NL}$ and $\mathbf{P} = \mathbf{\Lambda}_{NL}\mathbf{H}_{NL}^T$ in the known result (see pp. 49 of [10]) in matrix theory that

$$\mathbf{Q} \geq \mathbf{P}(\mathbf{P}^T\mathbf{Q}^{-1}\mathbf{P})^{-1}\mathbf{P}^T, \tag{2.56}$$

where $\mathbf{Q}$ is a positive definite $m \times m$ matrix, $\mathbf{P}$ is an $m \times k$ matrix.

2. The previous step verifies $\mathbf{W} \geq 0$. Thus, the lower bound stated in Eq. (2.50) is acquired:

$$\begin{aligned}
[\mathbf{J}^{-1}]_{2\times 2} &= c^2\mathbf{A}^{-1} + c^2\mathbf{F}\mathbf{W}^{-1}\mathbf{F}^T \\
&\geq c^2\mathbf{A}^{-1} \\
&= c^2\left(\mathbf{H}_{NL}\mathbf{\Lambda}_{NL}\mathbf{H}_{NL}^T + \mathbf{H}_L\mathbf{\Lambda}_L\mathbf{H}_L^T\right)^{-1}.
\end{aligned} \tag{2.57}$$

The equality holds when $\mathbf{W} = 0$, which is asymptotically achieved as $\mathbf{\Omega} \to 0$.

3. The upper bound is obtained by using $\mathbf{\Omega}^{-1} \geq 0$ as

$$\begin{aligned}
[\mathbf{J}^{-1}]_{2\times 2} &= c^2 \left[\mathbf{A}^{-1} + \mathbf{F}\left(c^2\mathbf{\Omega}^{-1} + \mathbf{\Lambda}_{NL} + -\mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{F}^T\right] \\
&\leq c^2 \left[\mathbf{A}^{-1} + \mathbf{F}\left(\mathbf{\Lambda}_{NL} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{F}^T\right] \qquad (2.58) \\
&= c^2 \left(\mathbf{H}_L\mathbf{\Lambda}_L\mathbf{H}_L^T\right)^{-1}. \qquad (2.59)
\end{aligned}$$

The last equality is derived by comparison between $\mathbf{J}_{\boldsymbol{\theta}}$ in Eq. (2.16) and $\mathbf{J}$ in Eq. (2.51) that

$$\mathbf{J}_{\boldsymbol{\theta}} = \mathbf{J}\Big|_{\mathbf{\Omega}^{-1}=0}.$$

Note the expression in Eq. (2.58) equals $[\mathbf{J}^{-1}]_{2\times 2}\Big|_{\mathbf{\Omega}^{-1}=0}$. Then, by utilizing the previous result of Proposition 2.1,

$$[\mathbf{J}_{\boldsymbol{\theta}}^{-1}]_{2\times 2} = \left(\mathbf{H}_L\mathbf{\Lambda}_L\mathbf{H}_L^T\right)^{-1},$$

we obtain the upper bound as in Eq. (2.59). In addition, it is not difficult to see that the upper bound is attained when $\mathbf{\Omega}^{-1} \to 0$, or the diagonal terms of $\mathbf{\Omega}$ increase to infinity.

□

At the lower bound, zero variances of NLOS delay estimates imply that we have acquired the exact NLOS path length, then the NLOS stations can be treated equivalent to the LOS ones by subtracting the known amount of NLOS delay l_b from the total time delay, τ_b. The analytical expression of the low bound, where the roles of LOS and NLOS quantities appear symmetric, confirms this deduction. At the other extreme, we simply have no idea of the NLOS distribution due to the infinite variances, which is reduced to the CRLB in the previous section. For a general $p_{\boldsymbol{l}}(\boldsymbol{l})$, the accuracy falls between the two limits. In other words, the lower and upper bounds

provide the range of the feasible performance improvement made by the extra NLOS information.

The restriction of the independence among the different NLOS delays, or that $\mathbf{\Omega}$ is a diagonal matrix, is nonessential. The proposition remains unchanged if we drop this constraint.

B. NLOS delays of Gamma distribution

Since an NLOS delay is always nonnegative in nature, l_b is often assumed to be Gamma distributed (subscript "b" will be omitted for a while):

$$G(l|\ \alpha, q) = \frac{\alpha^q}{\Gamma(q)} \exp(-\alpha \cdot l) \cdot l^{q-1}, \quad \text{for } l > 0, \tag{2.60}$$

where q and α are positive constants, and $\Gamma(\cdot)$ is the Gamma function. Roughly speaking, α in the exponential term determines the declining rate of the p.d.f as $l \to \infty$, while q in the polynomial is set to "balance" this trend. The "interaction" between the two parameters then controls the decay and spread pattern of the p.d.f. Figures 2.5 and 2.6 give two sets of density curves with fixed $\alpha = 0.5$ and $q = 0.5$, respectively. For the fixed α, the spreading of the p.d.f increases and the decay becomes slower as q becomes larger. An opposite tendency is observed for the density curves with fixed $q = 0.5$ in Figure 2.6. A well-known special case of the Gamma distribution is the exponential distribution, for which $q = 1$:

$$G(l|\ \alpha, 1) = \alpha \cdot \exp(-\alpha \cdot l), \quad \text{for } l > 0. \tag{2.61}$$

To evaluate $\mathbf{J}_P$ for the Gamma distribution, it is convenient for us first to obtain

$$\begin{aligned} E\left(\frac{\partial}{\partial l} \log G(l|\ \alpha, q)\right)^2 &= \alpha^2 + \frac{\alpha^q \cdot (q-1)^2}{\Gamma(q)} \int_0^{+\infty} \exp(-\alpha l) \cdot l^{q-3}\ dl \\ &\quad + \frac{2\alpha^{q+1} \cdot (q-1)}{\Gamma(q)} \int_0^{+\infty} \exp(-\alpha l) \cdot l^{q-2}\ dl. \end{aligned} \tag{2.62}$$

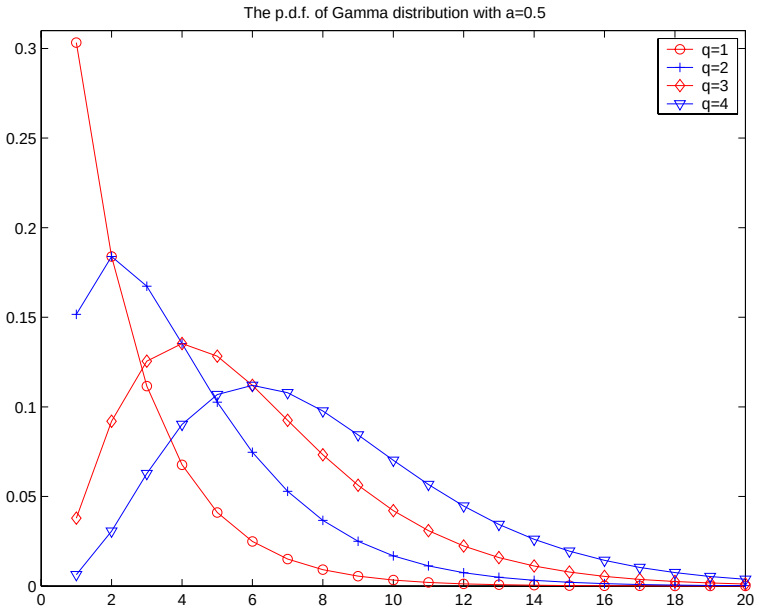


Figure 2.5: The p.d.f. of Gamma distribution with fixed $\alpha = 0.5$ and $q = 1, 2, 3, 4$.

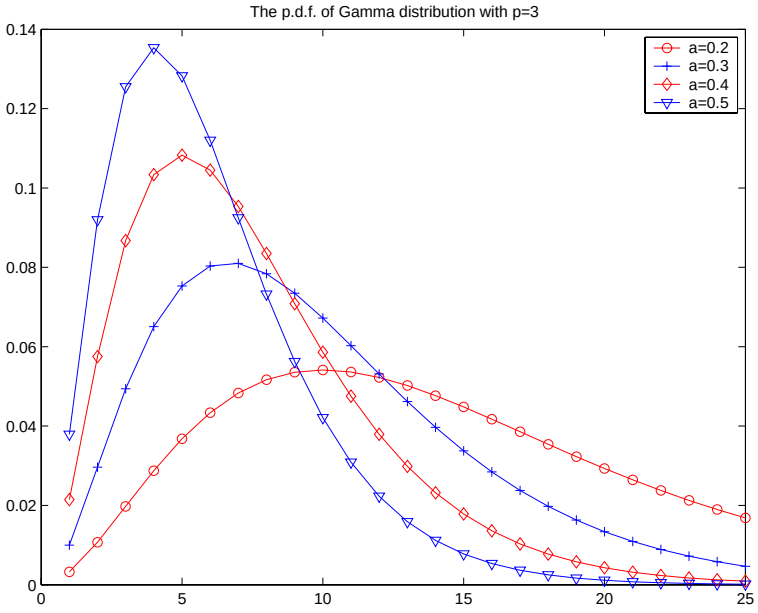


Figure 2.6: The p.d.f. of Gamma distribution with fixed $q = 3$ and $\alpha = 0.2, 0.3, 0.4, 0.5$.

The above expression looks complicated. Yet, it can be simplified in the following three cases.

- For $q = 1$, i.e., the exponential distribution,

$$E\left(\frac{\partial}{\partial l}\log G(l|\ \alpha, 1)\right)^2 = \alpha^2, \tag{2.63}$$

since the second and the third terms of Eq. (2.62) are zeros.

- For $q > 2$,

$$E\left(\frac{\partial}{\partial l}\log G(l|\ \alpha, q)\right)^2 = \frac{\alpha^2}{q-2}, \tag{2.64}$$

where we use the definition

$$\Gamma(q) = \int_0^{+\infty} \exp(-x)\cdot x^{q-1}\ dx, \quad \text{for } q > 0 \tag{2.65}$$

and the property

$$\Gamma(q+1) = q\Gamma(q) \tag{2.66}$$

to cancel out $\Gamma(\cdot)$'s in the denominators and the nominators in the second and the third terms of Eq. (2.62).

- For $q = 2$,

$$E\left(\frac{\partial}{\partial l}\log G(l|\ \alpha, 2)\right)^2 \to +\infty. \tag{2.67}$$

A direct way to get this is to let $q \to 2$ in Eq. (2.64). On the other hand, it can be derived from Eq. (2.62) as shown in Appendix 2.2.

For the other cases of $0 < q < 1$ and $1 < q < 2$, we may not be able to find a closed-form expression for Eq. (2.62). Hence, the discussion for these cases would be complex.

For simplicity, we assume that the NLOS delay variables l_b's are independent and Gamma distributed as

$$G(l_b|\ \alpha_b, q_b) = \frac{\alpha_b^{q_b}}{\Gamma(q_b)} \exp(-\alpha_b \cdot l_b) \cdot l_b^{q_b-1}, \tag{2.68}$$

for $l_b > 0$, $\alpha_b > 0$ and $q_b > 2$. Then,

$$\mathbf{J}_P = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Pi}^{-1} \end{pmatrix}, \tag{2.69}$$

where

$$\mathbf{\Pi}^{-1} = \text{diag}\left(\frac{\alpha_1^2}{q_1-2}, \frac{\alpha_2^2}{q_2-2}, \cdots, \frac{\alpha_M^2}{q_M-2}\right). \tag{2.70}$$

Comparing $\mathbf{J}_P$ of the Gaussian distribution in Eq. (2.47) and that of the Gamma distribution of Eq. (2.69), we recognize that $(q_b-2)/\alpha_b^2$ of the Gamma is comparable with the variance ω_b^2 of the Gaussian (See Eqs. (2.46) and (2.70)). In this way, the conclusion of Proposition 2.3 regarding the Gaussian case can be applied to the Gamma distribution. Specifically, when $q_b = 2$, for all $b \in \mathcal{NL}$, the quantity $(q_b-2)/\alpha_b^2$ of the Gamma corresponds to ω_b^2's$\to 0$ of the Gaussian. According to Proposition 2.3, it suggests that the lower bound of the G-CRLB concerning the Gamma distribution, i.e., $c^2\left(\mathbf{H}_{NL}\mathbf{\Lambda}_{NL}\mathbf{H}_{NL}^T + \mathbf{H}_L\mathbf{\Lambda}_L\mathbf{H}_L^T\right)^{-1}$, is attained when $q_b = 2$, for all $b \in \mathcal{NL}$.

When the NLOS delays may conform to some other distribution, we can go through steps similar to those for the Gamma distributions. Therefore, Proposition 2.3 holds for $p_{\boldsymbol{l}}(\boldsymbol{l})$ in general.

C. Lower Bounds with Prior Information on the MS Position

Suppose that we have some prior statistics of the MS position $\boldsymbol{p}$ as well as the NLOS variables $\boldsymbol{l}$. For example, we may assume that $\boldsymbol{p}$ is a bi-variate Gaussian variable with

known mean $\boldsymbol{u}_p$ and covariance $\boldsymbol{\Sigma}_p$. However, it is not easy to find the corresponding G-CRLB in a closed form, because of the difficulty in evaluating $\mathbf{J}_D$ which involves the expectation over $\boldsymbol{p}$ of $\mathbf{J}_{\boldsymbol{\theta}}$ given in Eq. (2.48). So we make a coarse approximation here: if the diagonal terms of the lower bound for the G-CRLB without the MS information (see Eq. (2.50)) are larger than those of $\boldsymbol{\Sigma}_p$, we take $\boldsymbol{\Sigma}_p$ as the positioning accuracy; if not, the extra MS information shall be ignored.

Numerical Examples

Based on the system specifications given in Section 2.2.2, we consider two numerical results of the positioning accuracy of the G-CRLB, which is given by

$$\mathcal{P}_{G-CR} \stackrel{\text{def}}{=} \sqrt{\text{trace}\left([\mathbf{J}^{-1}]_{2\times 2}\right)}. \tag{2.71}$$

Result 1.

We assume that BS_1 and BS_2 are NLOS BSs, while the remaining five stations receive LOS signals. The propagation loss factor is 2 for LOS (free space) and 4 for NLOS paths. The MS is located at (500,700). No prior information of the MS position is available here. The SNR is 0dB at a position 2000m from the MS. In Figure 2.7, the two middle curves represent $\mathcal{P}_{G-CR}$ (in meter) as a function of the mean value of NLOS induced path length (also in meter) with the standard deviations of the NLOS paths 5m and 15m. We also plot the upper bound (the top curve) and the lower bound (the bottom curve), which are associated with the cases where the deviations of the NLOS path are infinite and zero, respectively. It is observed that the lower and upper bounds become close to each other when the NLOS delay gets larger, because the NLOS signals are weaker and contain less "information" on the MS position. We also see that when the deviation of the NLOS delay is smaller, or

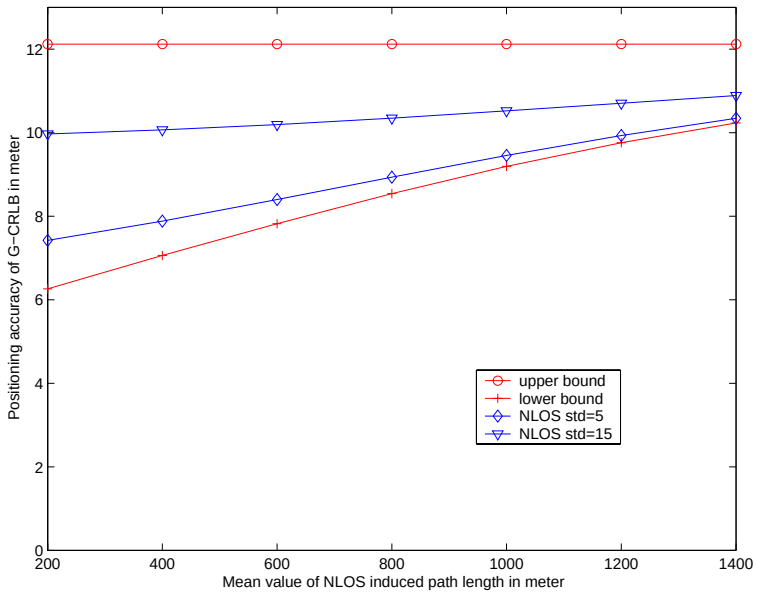


Figure 2.7: Numerical curves for the G-CRLB in a partial NLOS case: the accuracy $\mathcal{P}_{G-CR}$ vs. the mean of the NLOS induced path length.

equivalently, when we have more knowledge on $\boldsymbol{l}$, the corresponding curves converge to the lower bound curve.

Result 2.

The worst situation for TOA positioning occurs when no BSs receive LOS signals. Then, the G-CRLB of Eq. (2.50) becomes

$$\mathbf{J}^{-1} = c^2 \begin{pmatrix} \mathbf{H}_{NL}\mathbf{\Lambda}_{NL}\mathbf{H}_{NL}^T & \mathbf{H}_{NL}\mathbf{\Lambda}_{NL} \\ \mathbf{\Lambda}_{NL}\mathbf{H}_{NL}^T & \mathbf{\Lambda}_{NL} + \mathbf{\Omega}^{-1} \end{pmatrix}^{-1}. \tag{2.72}$$

Note no LOS related term is involved in the above expression. Figure 2.8 shows the accuracy $\mathcal{P}_{G-CR}$ vs. the standard deviation of NLOS induced path length with 3, 4, and 5 NLOS BSs. The mean of l_b, for $b \in \mathcal{NL}$, is set to be 100m, while the other model parameters are same as the previous example. We see that the accuracy is higher when signals from more BSs are processed and the deviation of the NLOS delays is smaller. The estimation error is less than 100m in most cases. Therefore,

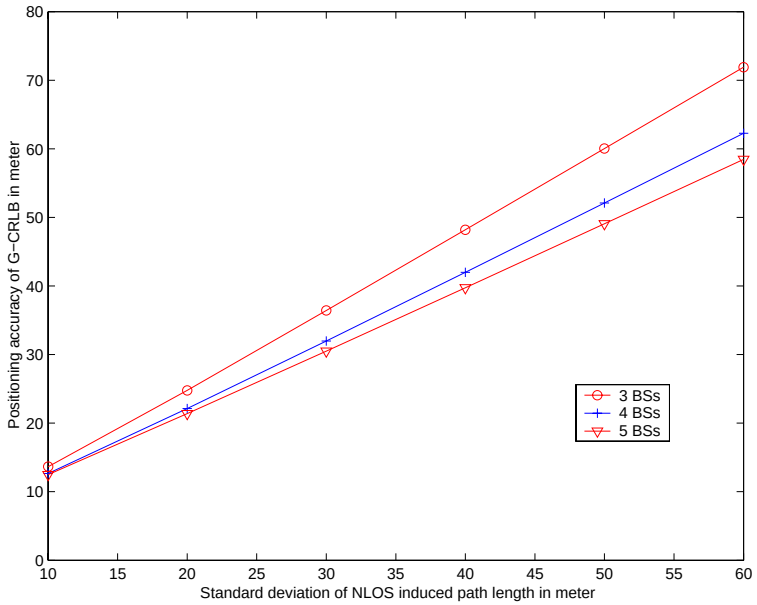


Figure 2.8: Numerical curves for the G-CRLB in a total NLOS case: the accuracy $\mathcal{P}_{G-CR}$ vs. the standard deviation of NLOS induced path length.

it is possible to locate the MS with an acceptable precision in a environment where no LOS signals present, as long as the deviation of NLOS delays is small enough and there is sufficient number of NLOS stations.

2.2.5 Maximum *a Posteriori* Probability Estimator

Now we move on to discuss the achievability of the G-CRLB and the optimum receiver that will asymptotically attain the bound.

Proposition 2.4

Given a prior p.d.f $p_{\boldsymbol{l}}(\boldsymbol{l})$ *of the NLOS delays, the G-CRLB for the MS position estimate,* $\left[\mathbf{J}^{-1}\right]_{2\times 2}$*, is asymptotically attained by the MAP (the maximum a posteriori probability) estimator based on the time delay estimates from all BSs. The achievability requires that the accuracy of the time delay estimates be sufficiently high, and*

$\boldsymbol{l}$ be in the neighborhood of a local maximum of $p_{\boldsymbol{l}}(\boldsymbol{l})$.

Proof.

Our objective is to prove the sufficient and necessary condition [5]: the G-CRLB is achievable by the MAP estimator, denoted as $\hat{\boldsymbol{\theta}}_{MAP}(\hat{\boldsymbol{\tau}})$, if and only if there exists estimator $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\tau}})$ such that the i-th component of $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\tau}})$ satisfies

$$[\hat{\theta}(\hat{\boldsymbol{\tau}})]_i - \theta_i = \sum_{j=1}^{M+2} d_{ij}(\boldsymbol{p}) \cdot \frac{\partial}{\partial \theta_j} \ln \left(f_{\boldsymbol{\theta}}(\hat{\boldsymbol{\tau}}) \cdot p_{\boldsymbol{l}}(\boldsymbol{l}) \right), \quad \text{for } i = 1, 2, \cdots, M+2, \tag{2.73}$$

where $d_{ij}(\boldsymbol{p})$ is a coefficient depending on $\boldsymbol{p}$ only, and $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\tau}})$ is exactly $\hat{\boldsymbol{\theta}}_{MAP}(\hat{\boldsymbol{\tau}})$. The proof of Eq. (2.73) proceeds as follows. We first show that the "if and only if" (iff) condition holds when $p_{\boldsymbol{l}}(\boldsymbol{l})$ is a multivariate Gaussian, by employing the same technique as in the proof of the achievability for the CRLB in Proposition 2.2. We then show that an arbitrary p.d.f can be approximated by a Gaussian p.d.f under certain conditions. Thus, the "iff" condition of Eq. (2.73) is valid for a general $p_{\boldsymbol{l}}(\boldsymbol{l})$.

1. Assume $p_{\boldsymbol{l}}(\boldsymbol{l}) = \mathcal{N}(\boldsymbol{u}_l, \boldsymbol{\Omega})$, where $\boldsymbol{\Omega}$ is not restricted to a diagonal matrix. Let

$$\boldsymbol{u}_\theta = \begin{pmatrix} 0 \\ 0 \\ \boldsymbol{u}_l \end{pmatrix}.$$

   We write $p_{\boldsymbol{l}}(\boldsymbol{l})$ in terms of $\mathbf{J}_P$ in Eq. (2.47) as

$$p_{\boldsymbol{l}}(\boldsymbol{l}) \;\propto\; \exp\left\{ -\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{u}_{\boldsymbol{\theta}})^T \cdot \mathbf{J}_P \cdot (\boldsymbol{\theta} - \boldsymbol{u}_{\boldsymbol{\theta}}) \right\}. \tag{2.74}$$

2. Recall the time delay of the received signal at BS_b is

$$\tau_b = \frac{1}{c} \left\{ \sqrt{(x_b - x)^2 + (y_b - y)^2} + l_b \right\}, b \in \mathcal{B}.$$

Differentiating both sides of the above equation yields

$$\Delta\boldsymbol{\tau} = \mathbf{H}^T \ \Delta\boldsymbol{\theta}, \tag{2.75}$$

or equivalently,

$$\hat{\boldsymbol{\tau}} - \boldsymbol{\tau} = \mathbf{H}^T \ (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}), \tag{2.76}$$

where

$$\Delta\boldsymbol{\tau} = \begin{pmatrix} \Delta\tau_1 \\ \Delta\tau_2 \\ \vdots \\ \Delta\tau_B \end{pmatrix}, \quad \Delta\boldsymbol{\theta} = \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta l_1 \\ \vdots \\ \Delta l_M \end{pmatrix} = \begin{pmatrix} \Delta\boldsymbol{p} \\ \Delta\boldsymbol{l} \end{pmatrix},$$

and $\mathbf{H}$ was given in Eq. (2.10). Therefore, $f_{\boldsymbol{\theta}}(\hat{\boldsymbol{\tau}})$ can be derived as

$$\begin{aligned} f_{\boldsymbol{\theta}}(\hat{\boldsymbol{\tau}}) &\propto \exp\left\{-\frac{1}{2}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau})^T \mathbf{J}_{\boldsymbol{\tau}} (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau})\right\} \\ &= \exp\left\{-\frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \cdot \mathbf{H}\mathbf{J}_{\boldsymbol{\tau}}\mathbf{H}^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\right\} \\ &= \exp\left\{-\frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \cdot \mathbf{J}_D \cdot (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\right\}, \end{aligned} \tag{2.77}$$

where $\mathbf{J}_{\boldsymbol{\tau}}$ and $\mathbf{J}_D$ were defined in Eqs. (2.11) and (2.49), respectively. Due to the same argument as in Proposition 2.2, we replace $\hat{\boldsymbol{\theta}}$ by $\hat{\boldsymbol{\theta}}_{ML}$, the MLE of $\boldsymbol{\theta}$ based on the time delay estimates from all the BSs.

3. Here comes the decisive step. It is not difficult to verify that

$$\begin{aligned} & f_{\boldsymbol{\theta}}(\hat{\boldsymbol{\tau}}) \cdot p_{\boldsymbol{l}}(\boldsymbol{l}) \\ \propto\ & \exp\left\{-\frac{1}{2}(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta})^T \cdot \mathbf{J}_D \cdot (\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta})\right\} \cdot \exp\left\{-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{u}_{\boldsymbol{\theta}})^T \cdot \mathbf{J}_P \cdot (\boldsymbol{\theta} - \boldsymbol{u}_{\boldsymbol{\theta}})\right\} \\ =\ & \exp\left\{-\frac{1}{2}(\hat{\boldsymbol{\theta}}_{MAP} - \boldsymbol{\theta})^T \cdot \mathbf{J} \cdot (\hat{\boldsymbol{\theta}}_{MAP} - \boldsymbol{\theta})\right\} \end{aligned} \tag{2.78}$$

where

$$\hat{\boldsymbol{\theta}}_{MAP} = \mathbf{J}^{-1}(\mathbf{J}_D \cdot \hat{\boldsymbol{\theta}}_{ML} + \mathbf{J}_P \cdot \boldsymbol{u}_\theta). \tag{2.79}$$

Then, it is straightforward to show

$$\hat{\boldsymbol{\theta}}_{MAP} - \boldsymbol{\theta} = -\mathbf{J}^{-1} \cdot \frac{\partial}{\partial \boldsymbol{\theta}} \ln \left(f_{\boldsymbol{\theta}}(\hat{\boldsymbol{\tau}}) \cdot p_{\boldsymbol{l}}(\boldsymbol{l}) \right), \tag{2.80}$$

which is the vector version of Eq. (2.73). Thus we have proved that the sufficient and necessary condition of Eq. (2.73) holds for Gaussian distributions.

4. For an arbitrary $p_{\boldsymbol{l}}(\boldsymbol{l})$, consider the Taylor series expansion of $\ln p_{\boldsymbol{l}}(\boldsymbol{l})$ around a local maximum point $\boldsymbol{l}_m$:

$$\begin{aligned} \ln p_{\boldsymbol{l}}(\boldsymbol{l}) &= \ln p_{\boldsymbol{l}}(\boldsymbol{l}_m) + (\boldsymbol{l} - \boldsymbol{l}_m)^T \frac{\partial}{\partial \boldsymbol{l}} \ln p_{\boldsymbol{l}}(\boldsymbol{l})\Big|_{\boldsymbol{l}=\boldsymbol{l}_m} \\ &\quad + \frac{1}{2}(\boldsymbol{l} - \boldsymbol{l}_m)^T \cdot \frac{\partial}{\partial \boldsymbol{l}} \left(\frac{\partial}{\partial \boldsymbol{l}} \right)^T \ln p_{\boldsymbol{l}}(\boldsymbol{l})\Big|_{\boldsymbol{l}=\boldsymbol{l}_m} \cdot (\boldsymbol{l} - \boldsymbol{l}_m) + \cdots \end{aligned} \tag{2.81}$$

Noting

$$\frac{\partial}{\partial \boldsymbol{l}} \ln p_{\boldsymbol{l}}(\boldsymbol{l})\Big|_{\boldsymbol{l}=\boldsymbol{l}_m} = 0,$$

we approximate $p_{\boldsymbol{l}}(\boldsymbol{l})$ by a Gaussian distribution in the local region around $\boldsymbol{l} = \boldsymbol{l}_m$:

$$p_{\boldsymbol{l}}(\boldsymbol{l}) \propto \exp\left\{ -\frac{1}{2}(\boldsymbol{l} - \boldsymbol{l}_m)^T \cdot \frac{\partial}{\partial \boldsymbol{l}} \left(\frac{\partial}{\partial \boldsymbol{l}} \right)^T \ln p_{\boldsymbol{l}}(\boldsymbol{l})\Big|_{\boldsymbol{l}=\boldsymbol{l}_m} \cdot (\boldsymbol{l} - \boldsymbol{l}_m) \right\}. \tag{2.82}$$

The conclusion for the Gaussian therefore can be applied to a broad class of prior distribution $p_{\boldsymbol{l}}(\boldsymbol{l})$.

□

It is the right time for us to review the justification of a frequent usage of Gaussian distributions. On one hand, the Gaussian is one of our most familiar distributions.

Its entire distribution is determined by the mean vector and the covariance matrix. Hence, it is relatively easy for us to trace the physical implication when we assume a variable is Gaussian distributed for a given problem. On the other hand, mathematically speaking, many density functions can be reduced to Gaussians under certain conditions, as shown in the fourth step of the above proof. This property often streamlines investigation of some otherwise complicated problems.

2.3 Other Positioning Methods

By now we have gone through the complete analysis of TOA positioning. In essence, we have introduced a general framework to examine various geolocation methods. However, to see how TDOA and signal strength based positioning methods are cast in the framework, we need to elaborate on some technical details, which is the theme of this section. We may reuse symbols such as $\mathbf{H}$ and $\mathbf{H}_L$ in favor of logic consistency, when no confusion is caused from the context.

2.3.1 TDOA Positioning

In a non-synchronous mobile system, there is an additional unknown parameter, i.e., the time offset between the clock at an MS and those at BSs. We denote this time offset as l_0/c, where c is the speed of light and l_0 is in the unit of length. The time offset may be represented by a negative value, in contrast that NLOS induced path lengths $\boldsymbol{l}$ have to be positive. The delay in the received signal of Eq. (2.2) is then modified to be

$$\tau_b = \frac{1}{c}\left\{\sqrt{(x_b - x)^2 + (y_b - y)^2} + l_0 + l_b\right\}, \quad \text{for } b \in \mathcal{B}, \tag{2.83}$$

to accommodate l_0 for the non-synchronous system. If we order the current unknown parameters as

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{p} \\ l_0 \\ \boldsymbol{l} \end{pmatrix}, \tag{2.84}$$

the only modification needed to compute the best accuracy is to replace the former $\mathbf{H}_{NL}$ and $\mathbf{H}_L$ in Eq. (2.10) with

$$\mathbf{H}_{NL} = \begin{pmatrix} \cos\phi_1 & \cos\phi_2 & \cdots & \cos\phi_M \\ \sin\phi_1 & \sin\phi_2 & \cdots & \sin\phi_M \\ 1 & 1 & \cdots & 1 \end{pmatrix} \tag{2.85}$$

and

$$\mathbf{H}_L = \begin{pmatrix} \cos\phi_{M+1} & \cos\phi_{M+2} & \cdots & \cos\phi_B \\ \sin\phi_{M+1} & \sin\phi_{M+2} & \cdots & \sin\phi_B \\ 1 & 1 & \cdots & 1 \end{pmatrix}, \tag{2.86}$$

respectively. By following the same analytical procedure we developed for TOA positioning, we reach the following conclusions regarding the optimum receiver for the non-synchronous system:

- In absence of the prior knowledge concerning the NLOS delay variables, the MLE of (x, y, l_0) based on $\{\hat{\tau}_b,\ b \in \mathcal{L}\}$ is an optimum receiver, and it achieves the CRLB asymptotically.

- When the prior p.d.f of the NLOS delays $\boldsymbol{l}$ is given, the MAP estimator for

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{p} \\ l_0 \\ \boldsymbol{l} \end{pmatrix}$$

extracted from $\{\hat{\tau}_b,\ b \in \mathcal{B}\}$ is an optimum receiver, and it asymptotically attains the G-CRLB.

Note that we do not use any concept of the time-difference-of-arrival (TDOA) method to draw the above conclusions. In other words, these conclusions pertain to the NLOS positioning with a non-synchronous system, regardless of which methods are adopted.

In practice, however, the TDOA method has long been accepted as a principal positioning approach for a non-synchronous system, since it can avoid estimation of the time offset l_0. A common argument lies in the triangulation principle that an MS can be located at the intersection of two hyperbolic curves determined by two TDOA data from three BSs. Thus, the following question will be inevitably prompted: what is the relation between the theoretically best accuracy and the conventional TDOA based estimate?

We now set forth to answer this question. Since the conventional TDOA approach is mainly designed for LOS scenarios, the best geolocation accuracy should be represented by the CRLB for the MS position estimate, i.e., $[\mathbf{J}_{\boldsymbol{\theta}}^{-1}]_{2\times 2}$, where

$$\mathbf{J}_{\boldsymbol{\theta}} = \frac{1}{c^2}\mathbf{H}_L \cdot \boldsymbol{\Lambda}_L \cdot \mathbf{H}_L^T, \tag{2.87}$$

$\mathbf{H}_L$ and $\boldsymbol{\Lambda}_L$ are given in Eqs. (2.86) and (2.13), respectively, and

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{p} \\ l_0 \end{pmatrix}. \tag{2.88}$$

Compared with $\boldsymbol{\theta}$ of Eq. (2.84) for the NLOS scenario, the above definition of $\boldsymbol{\theta}$ for the LOS situation does not include the NLOS delays $\boldsymbol{l}$. The LOS BSs are relabeled as $\mathcal{L} = \{1, 2, \cdots, L\}$ for simplicity. We first present an important result.

Proposition 2.5

In a LOS environment, the CRLB associated with TDOA data, denoted $\mathbf{J}_{TDOA}^{-1}$*, is equivalent to* $[\mathbf{J}_{\boldsymbol{\theta}}^{-1}]_{2\times 2}$*, where* $\mathbf{J}_{\boldsymbol{\theta}}$ *is given Eq. (2.87), i.e.,*

$$\mathbf{J}_{TDOA}^{-1} = \left[\mathbf{J}_{\boldsymbol{\theta}}^{-1}\right]_{2\times 2}. \tag{2.89}$$

Proof.

To facilitate further development, we separate $\mathbf{H}_L$ and $\mathbf{\Lambda}_L$ as

$$\begin{aligned}\mathbf{H}_L &= \left(\begin{array}{ccc|c} \cos\phi_1 & \cdots & \cos\phi_{L-1} & \cos\phi_L \\ \sin\phi_1 & \cdots & \sin\phi_{L-1} & \sin\phi_L \\ -- & -- & -- & -- \\ 1 & \cdots & 1 & 1 \end{array}\right) \\ &\stackrel{\text{def}}{=} \left(\begin{array}{cc} \mathbf{H}_{L-1} & \boldsymbol{h}_L \\ \mathbf{1}^T & 1 \end{array}\right),\end{aligned} \tag{2.90}$$

and

$$\begin{aligned}\mathbf{\Lambda}_L &= \left(\begin{array}{ccc|cc} \lambda_1 & & \mathbf{0} & | & 0 \\ & \ddots & & | & \vdots \\ \mathbf{0} & & \lambda_{L-1} & | & 0 \\ -- & -- & -- & -- & -- \\ 0 & \cdots & 0 & | & \lambda_L \end{array}\right) \\ &\stackrel{\text{def}}{=} \left(\begin{array}{cc} \mathbf{\Lambda}_{L-1} & \mathbf{0} \\ \mathbf{0}^T & \lambda_L \end{array}\right).\end{aligned} \tag{2.91}$$

The CRLB is then derived as

$$\begin{aligned}\mathbf{J}_{\boldsymbol{\theta}}^{-1} &= c^2\left(\mathbf{H}_L\mathbf{\Lambda}_L\mathbf{H}_L^T\right)^{-1} \\ &= c^2\left(\begin{array}{cc} \mathbf{H}_{L-1}\mathbf{\Lambda}_{L-1}\mathbf{H}_{L-1}^T + \lambda_L\boldsymbol{h}_L\boldsymbol{h}_L^T & \mathbf{H}_{L-1}\mathbf{\Lambda}_{L-1}\mathbf{1} + \lambda_L\boldsymbol{h}_L \\ \mathbf{1}^T\mathbf{\Lambda}_{L-1}\mathbf{H}_{L-1}^T + \lambda_L\boldsymbol{h}_L & \mathbf{1}^T\mathbf{\Lambda}_{L-1}\mathbf{1} + \lambda_L \end{array}\right)^{-1}\end{aligned}$$

$$\begin{aligned} &\stackrel{\text{def}}{=} \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & D \end{pmatrix}^{-1} \\ &= \begin{pmatrix} \mathbf{A}^{-1} + G^{-1}\mathbf{F}\mathbf{F}^T & -G^{-1}\mathbf{F} \\ -G^{-1}\mathbf{F}^T & G^{-1} \end{pmatrix}, \end{aligned} \tag{2.92}$$

where matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{D}$ are defined as in the above equation, and

$$\begin{aligned} D &= \sum_{b=1}^{L} \lambda_b, \\ G &= D - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B}, \\ \mathbf{F} &= \mathbf{A}^{-1}\mathbf{B}. \end{aligned}$$

Note that D and G are scalers. Therefore,

$$\left[\mathbf{J}_{\boldsymbol{\theta}}^{-1}\right]_{2\times 2} = \mathbf{A}^{-1} + \mathbf{F}\mathbf{W}^{-1}\mathbf{F}^T. \tag{2.93}$$

We are now ready to show the equivalence relationship of Eq. (2.89).

1. Selecting BS_L as the reference station, we construct TDOA data by taking the difference between the TOA $\hat{\tau}_b$ of Eq. (2.32) and $\hat{\tau}_L$, the TOA obtained at the reference station, as

$$\begin{aligned} \tilde{\tau}_b &= \hat{\tau}_b - \hat{\tau}_L \\ &= (\tau_b - \tau_L) + (\xi_b - \xi_L) \\ &= (\tau_b - \tau_L) + \psi_b, \quad \text{for } b = 1, \cdots, L-1, \end{aligned} \tag{2.94}$$

where

$$\psi_b \stackrel{\text{def}}{=} \xi_b - \xi_L.$$

We define

$$\boldsymbol{\psi} = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{L-1} \end{pmatrix}. \tag{2.95}$$

Then $\boldsymbol{\psi}$ is the multivariate Gaussian $\mathcal{N}(0, \boldsymbol{\Psi})$ with

$$\begin{aligned} \boldsymbol{\Psi} &= \begin{pmatrix} \lambda_1^{-1} + \lambda_L^{-1} & \lambda_L^{-1} & \cdots & \lambda_L^{-1} \\ \lambda_L^{-1} & \lambda_2^{-1} + \lambda_L^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \lambda_L^{-1} \\ \lambda_L^{-1} & \cdots & \lambda_L^{-1} & \lambda_{L-1}^{-1} + \lambda_L^{-1} \end{pmatrix} \\ &= \boldsymbol{\Lambda}_{L-1}^{-1} + \lambda_L^{-1} \mathbf{1} \cdot \mathbf{1}^T, \end{aligned} \tag{2.96}$$

where $\boldsymbol{\Lambda}_{L-1}$ is given in Eq. (2.91). Note that $\boldsymbol{\Psi}$ is not a diagonal matrix due to the correlation among ψ_b's introduced by the common reference BS.

2. The FIM associated with the TDOA data is derived in our familiar form:

$$\mathbf{J}_{TDOA} = \mathbf{H}_{TDOA} \cdot \boldsymbol{\Psi}^{-1} \cdot \mathbf{H}_{TDOA}^T, \tag{2.97}$$

where

$$\mathbf{H}_{TDOA} = \frac{1}{c} \left(\mathbf{H}_{L-1} - \boldsymbol{h}_L \cdot \mathbf{1}^T \right).$$

To evaluate $\boldsymbol{\Psi}^{-1}$, we employ the *Sherman-Morrison-Woodbury formula* (SMW) [36], i.e.,

$$\left(\mathbf{D} + \mathbf{U}\mathbf{V}^T \right)^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{U} \left(\mathbf{I} + \mathbf{V}^T\mathbf{D}^{-1}\mathbf{U} \right)^{-1} \mathbf{V}^T\mathbf{D}^{-1}, \tag{2.98}$$

where $\mathbf{I}$ is an identity matrix, and $\mathbf{D}$, $\mathbf{V}$ and $\mathbf{U}$ are as defined with appropriate dimensions. By applying $\mathbf{D} = \boldsymbol{\Lambda}_{L-1}^{-1}$, $\mathbf{U} = \lambda_L^{-1}\mathbf{1}$ and $\mathbf{V} = \mathbf{1}$ to the above

formula, we find

$$\mathbf{\Psi}^{-1} = \mathbf{\Lambda}_{L-1} - \frac{1}{\sum_{b=1}^{L}\lambda_b}\mathbf{\Lambda}_{L-1}\cdot\mathbf{1}\cdot\mathbf{1}^T\cdot\mathbf{\Lambda}_{L-1}. \tag{2.99}$$

3. In order to compute the inverse of $\mathbf{J}_{TDOA}$ with the SMW formula, we need to rearrange $\mathbf{J}_{TDOA}$ of Eq. (2.97) in the form of $\left(\mathbf{D}+\mathbf{U}\mathbf{V}^T\right)$, the left-hand side of Eq. (2.98), as

$$\begin{aligned}
\mathbf{J}_{TDOA} &= \frac{1}{c^2}\left(\mathbf{H}_{L-1}-\boldsymbol{h}_L\cdot\mathbf{1}^T\right)\cdot\left(\mathbf{\Lambda}_{L-1}-\frac{1}{\sum_{b=1}^{L}\lambda_b}\mathbf{\Lambda}_{L-1}\cdot\mathbf{1}\cdot\mathbf{1}^T\cdot\mathbf{\Lambda}_{L-1}\right)\cdot \\
&\quad \left(\mathbf{H}_{L-1}-\boldsymbol{h}_L\cdot\mathbf{1}^T\right)^T, \\
&= \frac{1}{c^2}\mathbf{H}_{L-1}\mathbf{\Lambda}_{L-1}\mathbf{H}_{L-1}^T+\frac{\lambda_L\sum_{b=1}^{L-1}\lambda_b}{c^2\sum_{b=1}^{L}\lambda_b}\boldsymbol{h}_L\boldsymbol{h}_L^T- \\
&\quad \frac{1}{c^2\sum_{b=1}^{L}\lambda_b}\mathbf{H}_{L-1}\mathbf{\Lambda}_{L-1}\cdot\mathbf{1}\cdot\mathbf{1}^T\cdot\mathbf{\Lambda}_{L-1}\mathbf{H}_{L-1}^T- \\
&\quad \frac{\lambda_L}{c^2\sum_{b=1}^{L}\lambda_b}\left(\boldsymbol{h}_L\cdot\mathbf{1}^T\mathbf{\Lambda}_{L-1}\mathbf{H}_{L-1}^T+\mathbf{H}_{L-1}\mathbf{\Lambda}_{L-1}\mathbf{1}\cdot\boldsymbol{h}_L^T\right), \\
&= \frac{1}{c^2}\left(\mathbf{H}_{L-1}\mathbf{\Lambda}_{L-1}\mathbf{H}_{L-1}^T+\lambda_L\boldsymbol{h}_L\boldsymbol{h}_L^T\right)- \\
&\quad \frac{1}{c^2\sum_{b=1}^{L}\lambda_b}\left(\mathbf{H}_{L-1}\mathbf{\Lambda}_{L-1}\mathbf{1}+\lambda_L\boldsymbol{h}_L\right)\left(\mathbf{H}_{L-1}\mathbf{\Lambda}_{L-1}\mathbf{1}+\lambda_L\boldsymbol{h}_L\right)^T.
\end{aligned} \tag{2.100}$$

Then, by setting

$$\begin{aligned}
\mathbf{D} &= \frac{1}{c^2}\left(\mathbf{H}_{L-1}\mathbf{\Lambda}_{L-1}\mathbf{H}_{L-1}^T+\lambda_L\boldsymbol{h}_L\boldsymbol{h}_L^T\right), \\
\mathbf{U} &= \frac{1}{c^2\sum_{b=1}^{L}\lambda_b}\left(\mathbf{H}_{L-1}\mathbf{\Lambda}_{L-1}\mathbf{1}+\lambda_L\boldsymbol{h}_L\right),
\end{aligned}$$

and

$$\mathbf{V} = \mathbf{H}_{L-1}\mathbf{\Lambda}_{L-1}\mathbf{1}+\lambda_L\boldsymbol{h}_L,$$

and applying the SMW of Eq. (2.98) to Eq. (2.100), we are able to show

$$\mathbf{J}_{TDOA}^{-1} = \left[\mathbf{J}_{\boldsymbol{\theta}}^{-1}\right]_{2\times 2}.$$

□

Proposition 2.5 ensures that the TDOA approach does not compromise the achievable accuracy while gaining the benefit of not estimating the time offset l_0. Moreover, we are able to show that the CRLB for TDOA, $\mathbf{J}_{TDOA}^{-1}$, is achievable by the TDOA based MLE, which can be proved by adopting the same technique as that for the achievability of the CRLB for TOA positioning in Proposition 2.1.

In summary, an optimum LOS positioning procedure for a non-synchronous system consists of three steps: (1) obtain $\hat{\boldsymbol{\tau}}$, the time delay estimates, at matched filter outputs; (2) create a set of TDOA data by selecting a reference BS; and (3) obtain the MLE of the MS position from the TDOAs.

Several relevant issues are worth mentioning before we close this section.

- Since it is confirmed that the optimum receiver is constructed by the MLE method, the commonly used Least Square based method is not optimal.

- The choice of the reference BS in producing TDOA data does not affect the final estimation accuracy. We can see this point clearly if we express $\mathbf{J}_{TDOA}$ of Eq. (2.100) in terms of outer product of vectors as shown later, in Eq. (3.3) in Chapter 3. That expression reveals that the quantities associated with the reference BS actually play the same role as those associated with other BSs.

- Optimum NLOS receivers are exactly similar as those for TOA positioning: if no statistics on delays are available, we should discard the NLOS TOA measurements, then utilize the optimum LOS receiver; alternatively, given some NLOS prior information, the MAP estimator based on TDOAs can be shown to be optimal.

2.3.2 Signal Strength based Positioning

Information regarding a mobile's position is contained in both the arrival times and the amplitudes of the received waveforms. The TOA and TDOA methods exploit the former type of information, whereas the signal-strength (SS) based method relies on the latter type of information. Compared with the other approaches, the SS method has two advantages. One is that some simple devices can be utilized to measure the energy of the received signal. The other is that the formulation of SS data, as will be discussed soon, takes into account some complex propagation effects that may be caused by a mobile environment or an indoor environment. Consequently, the SS based method has become a well-accepted method for indoor geolocation. However, as we need to locate the MS within a wide region, the poor positioning precision becomes a major limitation for the SS method. We shall postpone a full investigation of this shortcoming until the next chapter, where we explore the relation between the TOA and SS positioning methods.

In general, attenuation of signal strength through a mobile radio channel is caused by three nearly independent factors: (i) *path loss*, (ii) *multipath fading* and (iii) *shadowing*. Here we briefly review the three attenuation factors, which will be applied to the formulation of the SS data that follows.

- The path loss factor, ϵ, characterizes the rate at which the signal power decays as the distance d from the transmitter increases. In the free space, the signal power is proportional to d^{-2}, hence $\epsilon = 2$. A path loss factor of $\epsilon > 2$ is observed when signal propagation is subject to reflection and deflection from surrounding objects, such as floors, walls and foliage. Particularly, $\epsilon = 4$ is often used to characterize the path loss in urban areas [34].

- Multipath fading, also called fast fading, is the rapid fluctuation of the complex envelope of the received signal, caused by reception of multiple copies of a transmitted signal through multipath propagation. The amplitude distribution is often described by a Rayleigh or Rician distribution, depending on whether there exists a dominant component among the multiple copies. When there is no distinguishable component, the real and imaginary components of the complex envelope of the received signal can be viewed as a sum of numerous small random variables. Thus, both the real (R) and the imaginary (I) parts are modeled by Gaussian variables. Since the inphase and quadrature components of a bandpass signal are uncorrelated, R and I are also uncorrelated. Hence they are independent Gaussian variables. When R and I have zero means and common variance, the amplitude $\sqrt{R^2 + I^2}$ can be shown to be a Rayleigh variable. When R or I (or both) has non-zero mean but the common variance, it results in what is known as Rician distribution.

- Shadowing, often referred to as slow fading, represents a slow variation in the received signal strength, due to obstacles in the propagation paths. Experimental observations reveal that log-normal shadowing widely exists in nature. Several analytical models are proposed to explain the phenomenon [38, 39].

  Here is our simple derivation of the log-normal distribution model, which can be viewed as a special case of the general formulation presented in [39]. The propagation path between a transmitter and a receiver can be decomposed into a large number N of short "path elements" concatenated in tandem. The overall attenuation factor γ ($0 < \gamma \leq 1$) of the path should be given as the product of

the attenuation factor γ_i of these path elements:

$$\gamma = \prod_{i=1}^{N} \gamma_i.$$

Taking the logarithm on both sides of the above equation, we have

$$S \stackrel{\text{def}}{=} \ln \gamma = \sum_{i=1}^{N} \ln \gamma_i.$$

Let

$$\ln \gamma_i = \mu_i + \xi_i,$$

where

$$\mu_i = E(\ln \gamma_i),$$

and ξ_i is a random variable with zero mean. Then, by virtue of the central limit theorem [35], we can show that if ξ_i's are statistically independent, $S = \ln \gamma$ will attain to a normal variable as N goes to infinity, with mean

$$\mu = \sum_{i=1}^{N} \mu_i = \sum_{i=1}^{N} E(\ln \gamma_i) < +\infty \tag{2.101}$$

and variance

$$\sigma^2 = \sum_{i=1}^{N} \sigma_i^2 = \sum_{i=1}^{N} \text{var}(\ln \gamma_i), \tag{2.102}$$

where

$$\sigma_i^2 = E(\xi_i^2).$$

Thus, $S = \ln \gamma$ is asymptotically $\mathcal{N}(\mu, \sigma^2)$ when $\mu < +\infty$, and γ is a log-normal variable. Note that it is not necessary to assume that ξ_i's are identically distributed, as usually assumed in the central limit theorem, as long as the individual σ_i^2 are small as compared with their sum σ^2 (Lindberg's condition. See pp. 256–257 and pp. 451–494 in [37]).

We now return to our main discussion. The received signal power, P_r, consists of two components:

$$P_r = P_S + P_n, \tag{2.103}$$

where P_S is the signal power and P_n is the noise power. In many circumstances, the main noise component is thermal noise introduced at the receiver front. The thermal noise is largely determined by the receiver noise temperature and is nearly independent of the receiver's location. Even if effects of atmospheric and other noise are present, the overall noise level will be constant over a short period of time. Hence it is usually assumed that P_n can be detected ahead of time. Consequently, we are able to obtain the signal power P_S in Eq. (2.103).

The signal power P_S at the receiver is given by

$$P_S = k \cdot P_t \cdot \frac{g^2 \cdot \gamma}{d^\epsilon}, \tag{2.104}$$

where P_t is the (known) transmitted power from a BS tramsmitter, k is defined by

$$k = \frac{G_t G_r}{4\pi},$$

and is a known constant if the antenna gains G_t and G_r at the transmitter and the receiver are specified, d is the distance between the MS and the BS, g and γ are variables that conform to Rayleigh[1] (or Rician) and log-normal distributions, respectively.

In radio engineering, it is a common practice to represent signal power in dBw or dBm, and gain and loss factors in dB, because the "product" equation of the type of Eq. (2.104) will be converted into a "summation" equation. Therefore, from Eq. (2.104) we have

$$P_S \text{ [dBw]} = P_t \text{ [dBw]} + z + \kappa + w' + C', \tag{2.105}$$

[1] If g is Rayleigh distributed, g^2 is exponentially distributed.

where

$$z = -10 \cdot \epsilon \cdot \log_{10} d, \tag{2.106}$$

$$\kappa = 20 \log_{10} g, \tag{2.107}$$

$$w' = 10 \log_{10} \gamma, \tag{2.108}$$

and

$$C' = 10 \log_{10} k.$$

The transformed variable

$$w' = 10 \log_{10} \gamma = 10 \frac{\ln \gamma}{\ln 10} = \frac{10}{\ln 10} S$$

is normally distributed with mean $10\mu/\ln 10$ and variance $100\sigma^2/(\ln 10)^2$, where μ and σ^2 are the quantities defined in Eqs. (2.101) and (2.102). In Eq. (2.105), the term

$$\kappa = 20 \log_{10} g$$

will vary in time since g is the fast fading term. Then by averaging the received power over some internal, we have the time-averaged version of Eq. (2.105)

$$\overline{P}_S \text{ [dBw]} = P_t \text{ [dBw]} + z + \overline{\kappa} + w' + C', \tag{2.109}$$

In practice, the transmitted power P_t, the quantity C', the mean $10\mu/\ln 10$ of w' in Eq. (2.105), and the quantity $\overline{\kappa}$ in Eq. (2.109) are often known beforehand. Thus, we are able to simplify Eq. (2.105) as

$$\text{Model A:} \qquad e = z + \kappa + w, \tag{2.110}$$

where

$$e = P_S \text{ [dBW]} - P_t \text{ [dBW]} - C' - \frac{10}{\ln 10}\mu,$$

$$w = w' - \frac{10}{\ln 10}\mu,$$

and w is the Gaussian variable $\mathcal{N}(0, \eta^2)$, with

$$\eta^2 = \frac{100\sigma^2}{(\ln 10)^2}.$$

Similarly, Eq. (2.109) becomes

$$\text{Model B:} \qquad \overline{e} = z + w, \tag{2.111}$$

where

$$\overline{e} = \overline{P}_S \text{ [dBW]} - P_t \text{ [dBW]} - C' - \frac{10}{\ln 10}\mu - \overline{\kappa}.$$

Note that Model B is in a simple form of a useful "signal" plus Gaussian noise with the zero mean as the time delay estimate of Eq. (2.32), i.e.,

$$\hat{\tau_b} = \tau_b + \xi_b, \quad \text{for } b \in \mathcal{B}.$$

Here we present the CRLB for the SS positioning method based on Eq. (2.111) only, because of its practical interest. The SS data to be processed are

$$\overline{e}_b = z_b + w_b, \ b \in \mathcal{B},$$

where

$$z_b = -10 \cdot \epsilon \cdot \log_{10} d_b,$$

with

$$d_b = \sqrt{(x_b - x)^2 + (y_b - y)^2},$$

and "noises" w_b are independent Gaussian variables $\mathcal{N}(0, \eta_b^2)$. It takes some calculations to show that

$$\mathbf{J}_{\boldsymbol{p}}^{-1} = \left(\mathbf{H}_{SS} \cdot \mathbf{J}_{\boldsymbol{z}} \cdot \mathbf{H}_{SS}^T\right)^{-1} \tag{2.112}$$

where

$$\mathbf{H}_{SS} = \frac{10\epsilon}{c \ln 10} \cdot \begin{pmatrix} \cos\phi_1 & \cos\phi_2 & \cdots & \cos\phi_B \\ \sin\phi_1 & \sin\phi_2 & \cdots & \sin\phi_B \end{pmatrix} \cdot [\mathrm{diag}(d_1, d_2, \cdots, d_B)]^{-1}$$

and

$$\mathbf{J}_{\boldsymbol{z}} = \left[\mathrm{diag}(\eta_1^2, \eta_2^2, \cdots, \eta_B^2)\right]^{-1}.$$

We derive Eq. (2.112) by following the same steps as we developed for the CRLB regarding TOA positioning in Section 2.2.

2.4 Concluding Remarks

We have developed a unified analysis of NLOS geolocation which covers major geolocation methods. For the TOA and TDOA positioning methods, the relationship between the highest geolocation accuracy and the optimum receiver can be summarized as follows. When no prior information on NLOS delays is available, the best positioning accuracy can be derived in terms of the CRLB, which can be asymptotically achieved by the MLE based on the LOS time delay estimates. For a general case, the best accuracy is given by the G-CRLB, which is asymptotically attained by the MAP estimator using all time delay data. The common step for both the conventional TOA (or TDOA) method and the optimum geolocation schemes is that their first steps are to obtain time delay estimates. The analysis for the SS method can also be incorporated in the unified analysis as shown in Eq. (2.112).

The matrix separation technique plays an important role in the analysis. It produces the basic blocks which help us to organize the computation in a compact manner, and provides clear a physical interpretation of the final results. Especially, there

are two common types of matrices: one is $\mathbf{H}$ that contains the geometric relations among the MS and the BSs, and the other is $\mathbf{J}_{\boldsymbol{\tau}}$ that links to the relevant system parameters. Each type of matrices is then divided into the submatrices corresponding to LOS and NLOS signals.

Appendix 2.1: Derivation of Eq. (2.32)

We suppress subscript b for simplicity.

1. Estimating τ from the received signal $r(t)$ is to find the time $v = \hat{\tau}$ when the matched filter output

$$h(v) = \int r(t)s^*(t-v)\ dt \tag{2.113}$$

becomes maximum, where "*" means complex conjugate. The maximization can be done by setting the derivative of $y(v)$ to zero, i.e,

$$\frac{\partial}{\partial v}h(v) = 0, \tag{2.114}$$

and solving the equation for v. Substituting the explicit expression of $r(t)$ into the above equation, we write down

$$A\frac{\partial}{\partial v}g(v) + \gamma = 0, \tag{2.115}$$

where

$$g(v) = \int s(t-\tau)s^*(t-v)\ dt, \tag{2.116}$$

and

$$\gamma = \int n(t)\frac{\partial}{\partial v}s^*(t-v)\ dt. \tag{2.117}$$

2. Take the Taylor series expansion of $g(v)$ around $v = \tau$:

$$\begin{aligned} g(v) &= g(\tau) + \frac{dg(\tau)}{d\tau}\cdot(v-\tau) + \frac{1}{2!}\cdot\frac{d^2g(\tau)}{d\tau^2}\cdot(v-\tau)^2 + o\left[(v-\tau)^3\right] \\ &= 1 - 2\pi^2\beta^2(v-\tau)^2 + o\left[(v-\tau)^3\right], \end{aligned} \tag{2.118}$$

using

$$\begin{aligned}
g(\tau) &= \int |s(t)|^2 \, dt = 1, \quad \text{(the normalization condition)} \\
\frac{dg(\tau)}{d\tau} &= \frac{\partial}{\partial v} \int s(t-\tau) s^*(t-v) \, dt \Big|_{v=\tau} = 0, \quad \text{(maximization requirement)} \\
\frac{d^2 g(\tau)}{d\tau^2} &= \int s(t-\tau) \cdot \frac{\partial^2}{\partial v^2} s^*(t-v) \, dt \Big|_{v=\tau} \\
&= \int s(t+v-\tau) \cdot \frac{\partial^2}{\partial t^2} s^*(t) \, dt \Big|_{v=\tau} \quad \text{(with } t \longleftarrow t-v\text{)} \\
&= \int s(t) \cdot \frac{\partial^2}{\partial t^2} s^*(t) \, dt \\
&= s(t) \cdot \frac{\partial}{\partial t} s^*(t) \Big|_{-\infty}^{\infty} - \int \left| \frac{\partial}{\partial t} s(t) \right|^2 \, dt \quad \text{(Integration by parts)} \\
&= -4\pi^2 \beta^2. \qquad (2.119)
\end{aligned}$$

3. Since γ of Eq. (2.117) is defined by an integral of the Gaussian random process $n(t)$ multiplying a known function, γ is a Gaussian random variable, with mean

$$E[\gamma] = \int E[n(t)] \frac{\partial}{\partial v} s^*(t-v) \, dt = 0, \qquad (2.120)$$

and variance

$$\begin{aligned}
E[\gamma^2] &= \int \int E\left[n(t_1) n^*(t_2)\right] \frac{\partial}{\partial \tau} s(t_1 - v) \frac{\partial}{\partial \tau} s^*(t_2 - v) \, dt_1 dt_2 \\
&= N_0 \int \left| \frac{\partial}{\partial t} s(t) \right|^2 \, dt \\
&\qquad \text{(using } E\left[n(t_1) n^*(t_2)\right] = N_0 \delta(t_1 - t_2) \text{)} \\
&= N_0 \cdot 4\pi^2 \beta^2. \qquad (2.121)
\end{aligned}$$

4. With Eq. (2.118), it is straightforward to solve the maximization equation of Eq. (2.115):

$$\hat{\tau} = \arg\left\{\max_v \, h(v)\right\} = \tau + \frac{\gamma}{4\pi^2 \beta^2 A}. \qquad (2.122)$$

Hence the delay estimate at the matched filter output can be represented by the true delay, τ, plus a noise term. Since γ is a Gaussian variable $\mathcal{N}(0, 4\pi^2\beta^2 N_0)$ as discussed before, the noise term $\gamma/4\pi^2\beta^2 A$ is also a Gaussian random variable with $\mathcal{N}(0, N_0/4\pi^2\beta^2 A^2)$. For the sake of clarity, we rewrite Eq. (2.122) as

$$\hat{\tau} = \tau + \xi, \tag{2.123}$$

where

$$\xi = \frac{\gamma}{4\pi^2\beta^2 A}$$

conforms to $\mathcal{N}(0, \sigma^2)$ with

$$\sigma^2 = \frac{N_0}{4\pi^2\beta^2 A^2} = \frac{1}{4\pi^2\beta^2 R}.$$

5. Finally, we examine the region around $v = \tau$, where the approximation Eq. (2.118) of $g(v)$ is valid. Utilizing Eq. (2.122), we see that the region around $v = \tau$ means that the estimation error ξ is small, which in turn requires the variance $\sigma^2 = 1/4\pi^2\beta^2 R$ is small, or $R \cdot \beta^2$ is large.

□

Appendix 2.2: Derivation of Eq. (2.67) from Eq. (2.62)

With Eq. (2.62), we have

$$\begin{aligned} E\left(\frac{\partial}{\partial l}\log G(l|\,\alpha, 2)\right)^2 &= \alpha^2 + \alpha^2\int_0^{+\infty}\exp(-\alpha l)\cdot l^{-1}\,dl + 2\alpha^3\int_0^{+\infty}\exp(-\alpha l)\,dl \\ &\quad \text{(using } \Gamma(n) = (n-1)! \text{ for } n = 1, 2, \cdots.) \\ &= 3\alpha^2 + \alpha^2\int_0^{+\infty}\exp(-l)\cdot l^{-1}\,dl \\ &\geq \alpha^2\int_1^{+\infty}\exp(-l)\cdot l^{-1}\,dl \\ &= \alpha^2\left(\ln l + l + \frac{l^2}{2\cdot 2!} + \frac{l^3}{3\cdot 3!} + \cdots\right)\Bigg|_{l=1}^{+\infty}. \end{aligned} \tag{2.124}$$

Now we need to show the lower bound of the above expression is infinite. When $l = 1$,

$$\begin{aligned} & \left(\ln l + l + \frac{l^2}{2 \cdot 2!} + \frac{l^3}{3 \cdot 3!} + \cdots\right)\Big|_{l=1} \\ = & \left(1 + \frac{1}{2 \cdot 2!} + \frac{1}{3 \cdot 3!} + \cdots\right) \\ \leq & \left(1 + \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{n(n+1)} + \cdots\right) \\ = & \left(1 + \frac{n}{n+1}\Big|_{n\to\infty}\right) \\ = & 1. \end{aligned} \tag{2.125}$$

For $l \to \infty$, it is easy to see

$$\left(\ln l + l + \frac{l^2}{2 \cdot 2!} + \frac{l^3}{3 \cdot 3!}\right)\Big|_{l\to+\infty} \to +\infty. \tag{2.126}$$

We obtain Eq. (2.67) by substituting Eqs. (2.125) and (2.126) into Eq.(2.124). □

Chapter 3

Relationship among Distanced-based Methods

TOA, TDOA and signal strength (SS) positioning methods are the three principal distance-based methods. To the best of our knowledge, most studies reported in the literature investigate these techniques in isolation from one another, and few results on their relationships have been reported. However, a better appreciation of the connections among the three methods is of both theoretical and practical interest.

The link between the TOA and TDOA methods is first examined. When a set of BS locations and an MS position are given, we know in principle that the TOA method should achieve higher positioning precision than the TDOA counterpart, because in the latter there is an extra unknown parameter to be estimated, i.e., the time offset between the clock at the MS and those at the BSs. Here we provide an analytical explanation for this argument. We show that the two positioning methods may attain the same level of accuracy under certain conditions. We then pursue the tradeoff between the accuracy limits for TOA and SS positioning, which leads to a new hybrid geolocation scheme that combines TOA and SS data.

3.1 Relationship between TOA and TDOA Methods

We investigate the relationship between the TDOA and TOA methods by comparing their CRLBs. According to Proposition 2.1, the CRLB for the NLOS geolocation depends solely on LOS signals. Thus it suffices to consider a LOS scenario only.

To facilitate our development, we first prepare the FIMs associated with the two CRLBs in terms of outer product of vectors. For TOA positioning, we found in Proposition 2.1 that

$$\begin{aligned}\mathbf{J}_{TOA} &= \frac{1}{c^2}\left(\mathbf{H}_L\mathbf{\Lambda}_L\mathbf{H}_L^T\right)\\ &= \frac{1}{c^2}\begin{pmatrix}\cos\phi_1 & \cos\phi_2 & \cdots & \cos\phi_L\\ \sin\phi_1 & \sin\phi_2 & \cdots & \sin\phi_L\end{pmatrix}\operatorname{diag}\begin{pmatrix}\lambda_1 & \lambda_2 & \cdots & \lambda_L\end{pmatrix}\begin{pmatrix}\cos\phi_1 & \sin\phi_1\\ \cos\phi_2 & \sin\phi_2\\ \vdots & \vdots\\ \cos\phi_L & \sin\phi_L\end{pmatrix}.\end{aligned}$$

By defining a unit vector

$$\boldsymbol{h}_b = \begin{pmatrix}\cos\phi_b\\ \sin\phi_b\end{pmatrix},$$

we rewrite the TOA FIM as

$$\mathbf{J}_{TOA} = \frac{1}{c^2}\sum_{b\in\mathcal{L}}\lambda_b\boldsymbol{h}_b\boldsymbol{h}_b^T. \tag{3.1}$$

Since λ_b's are all positive, we define weight coefficient w_b as

$$w_b = \frac{\lambda_b}{\lambda},$$

where

$$\lambda = \sum_{b\in\mathcal{L}}\lambda_b.$$

Thus, Eq. (3.1) becomes

$$\mathbf{J}_{TOA} = \frac{\lambda}{c^2} \sum_{b \in \mathcal{L}} w_b \boldsymbol{h}_b \boldsymbol{h}_b^T. \tag{3.2}$$

As for the TDOA FIM, we have two equivalent expressions by Proposition 2.5. We invoke the one given in Eq. (2.100) to compute

$$\begin{aligned} \mathbf{J}_{TDOA} &= \frac{1}{c^2} \left(\mathbf{H}_{L-1} \mathbf{\Lambda}_{L-1} \mathbf{H}_{L-1}^T + \lambda_L \boldsymbol{h}_L \boldsymbol{h}_L^T \right) - \\ &\quad \frac{1}{c^2 \lambda} \left(\mathbf{H}_{L-1} \mathbf{\Lambda}_{L-1} \mathbf{1} + \lambda_L \boldsymbol{h}_L \right) \left(\mathbf{H}_{L-1} \mathbf{\Lambda}_{L-1} \mathbf{1} + \lambda_L \boldsymbol{h}_L \right)^T \\ &= \frac{1}{c^2} \sum_{b \in \mathcal{L}} \lambda_b \boldsymbol{h}_b \boldsymbol{h}_b^T - \frac{1}{c^2 \lambda} (\sum_{b \in \mathcal{L}} \lambda_b \boldsymbol{h}_b)(\sum_{b \in \mathcal{L}} \lambda_b \boldsymbol{h}_b)^T \\ &= \frac{\lambda}{c^2} \left(\sum_{b \in \mathcal{L}} w_b \boldsymbol{h}_b \boldsymbol{h}_b^T - (\sum_{b \in \mathcal{L}} w_b \boldsymbol{h}_b)(\sum_{b \in \mathcal{L}} w_b \boldsymbol{h}_b)^T \right). \end{aligned} \tag{3.3}$$

Define random vector $\boldsymbol{h}$ that takes values of

$$\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_L$$

with probabilities

$$w_1, w_2, \cdots, w_L,$$

and its weighted average

$$\bar{\boldsymbol{h}} \stackrel{\text{def}}{=} \sum_{b=1}^{L} w_b \boldsymbol{h}_b.$$

We can express $\mathbf{J}_{TOA}$ and $\mathbf{J}_{TDOA}$ in terms of the second moment and covariance of $\boldsymbol{h}$, respectively, i.e.,

$$\mathbf{J}_{TOA} = \frac{\lambda}{c^2} \cdot E \left[\boldsymbol{h} \cdot \boldsymbol{h}^T \right], \tag{3.4}$$

and

$$\mathbf{J}_{TDOA} = \frac{\lambda}{c^2} \cdot E \left[(\boldsymbol{h} - \bar{\boldsymbol{h}}) \cdot (\boldsymbol{h} - \bar{\boldsymbol{h}})^T \right]. \tag{3.5}$$

By utilizing Eqs. (3.2) and (3.3), we immediately see

$$\mathbf{J}_{TOA} - \mathbf{J}_{TDOA} = \frac{\lambda}{c^2} \cdot \bar{\boldsymbol{h}} \cdot \bar{\boldsymbol{h}}^T \geq 0. \tag{3.6}$$

Hence,

$$\mathbf{J}_{TDOA}^{-1} \geq \mathbf{J}_{TOA}^{-1}. \tag{3.7}$$

The above inequality confirms the long-held argument that as far as we use the same set of time delay estimates $\{\hat{\boldsymbol{\tau}}_b,\ b \in \mathcal{L}\}$, TDOA positioning cannot perform better than the TOA method because of the unknown time-offset l_0 assumed in the TDOA. The amount of degradation is given in Eq. (3.6). Thus, we see that weighted average $\bar{\boldsymbol{h}}$ is a crucial quantity in determining the degradation amount. Moreover, the sufficient and necessary condition for the equality to hold in Eq. (3.7) is

$$\mathbf{J}_{TDOA}^{-1} = \mathbf{J}_{TOA}^{-1} \quad \Leftrightarrow \quad \bar{\boldsymbol{h}} = 0. \tag{3.8}$$

That is, the TOA and TDOA positioning methods can attain the same accuracy if and only if when $\bar{\boldsymbol{h}}$ is exactly zero, which may be viewed as a kind of symmetry condition among the configuration among the BSs and the MS. This symmetry can annul the accuracy degradation of the TDOA method with respect to the TOA accuracy. Here is a simple example for $\bar{\boldsymbol{h}} = 0$. Consider that L BSs are distributed evenly around the circle with the center at the mobile's location. We then have

$$w_1 = w_2 = \cdots = w_L,$$

and

$$\sum_{b \in \mathcal{L}} \boldsymbol{h}_b = 0.$$

Thus, $\bar{\boldsymbol{h}} = 0$.

The relation of Eq. (3.7) implies that $\mathbf{J}_{TDOA}^{-1}$, the CRLB of the TDOA method, is lower bounded by $\mathbf{J}_{TOA}^{-1}$. Hence, a closely related question is raised: does there exist an upper bound for $\mathbf{J}_{TDOA}^{-1}$ determined by some TOA configuration? To be more specific, denote $\mathbf{J}_{TOA}(l)$ the TOA FIM associated with l BSs of (BS_1, BS_2, $\cdots$, BS_l),

which is the subset of $\mathcal{L}$. Can we find an $0 < l < L$ such that

$$\mathbf{J}_{TDOA}^{-1} \leq \mathbf{J}_{TOA}(l)^{-1} \text{ ?}$$

We pose this question, since it is conceivable that use of fewer BSs should reduce the positioning accuracy. Along with the lower bound of Eq. (3.7), the plausible relation

$$\mathbf{J}_{TOA}^{-1}(L) \leq \mathbf{J}_{TDOA}^{-1} \leq \mathbf{J}_{TOA}^{-1}(l)$$

would possibly allow to approximate the performance of a given TDOA scheme by two related TOA solutions. The conjectured upper bound is equivalent to requiring

$$\sum_{b=l+1}^{L} w_b \boldsymbol{h}_b \boldsymbol{h}_b^T \geq (\sum_{b\in\mathcal{L}} w_b \boldsymbol{h}_b)(\sum_{b\in\mathcal{L}} w_b \boldsymbol{h}_b)^T. \tag{3.9}$$

However, the answer depends on the specific configuration of $\{\boldsymbol{h}_b,\ b \in \mathcal{L}\}$ and $\{w_b,\ b \in \mathcal{L}\}$. We provide two examples in Appendix 3.1, where such an l exists in one example and $l = 0$ in the other.

3.2 Relationship between TOA and SS Methods

We pointed out in Section 2.3.2 that the main disadvantage for the SS positioning method is its poor precision when locating an MS within a wide region. In this section, we elaborate on this claim by examining the achievable accuracy of the SS method along with that of the TOA method.

Consider a one-dimensional case like in a radar ranging problem. Our task is to estimate the distance d between an MS and one BS based on the SS or TOA estimates.

For the SS method, we use the time averaged of SS data as in Eq. (2.111) to compute the corresponding CRLB. Recall

$$\overline{e} = z + w,$$

where

$$z = -10 \cdot \epsilon \cdot \log d,$$

w is a Gaussian variable $\mathcal{N}(0, \eta^2)$ representing log-normal shadowing, and ϵ is the path loss factor. Thus, the p.d.f. of $\overline{e}$ conditioned on d is

$$f_d(\overline{e}) \propto \exp\left\{-\frac{1}{2\eta^2}\left(\overline{e} + 10 \cdot \epsilon \cdot \log d\right)^2\right\}. \tag{3.10}$$

Substituting the above p.d.f to the one-dimensional FIM

$$(\mathbf{J}_d)_{SS} = E_d\left[\frac{\partial}{\partial d}\ln f_d(\overline{e}) \cdot \left(\frac{\partial}{\partial d}\ln f_d(\overline{e})\right)^T\right], \tag{3.11}$$

we can show that the CRLB is

$$(\mathbf{J}_d)_{SS}^{-1} = \left(\frac{\ln 10}{10}\right)^2 \cdot \frac{\eta^2}{\epsilon^2} \cdot d^2, \tag{3.12}$$

or

$$\sqrt{\mathrm{var}(\hat{d})} \geq \frac{\ln 10}{10} \cdot \frac{\eta}{\epsilon} \cdot d. \tag{3.13}$$

Note that the accuracy of the above expression is proportional to d. In other words, in order to maintain the estimation error of less than δd, the MS has to be within the range of

$$r_0 = \frac{10}{\ln 10} \cdot \frac{\epsilon}{\eta} \cdot \delta d \tag{3.14}$$

from the specific BS's location. For typical numbers $\epsilon = 4$ and $\eta^2 = 8$ pertaining to outdoor geolocation, the accuracy of Eq. (3.13) is roughly $0.2d$. Thus, to secure the accuracy of 100m, the maximum distance between the MS and the BS is 500m.

We should notice in Eq. (3.13) that ϵ and η^2 are completely determined by the characteristics of a communication channel. Hence there is little we can do to control or improve these factors and the resulting positioning accuracy.

For TOA positioning, the case is different. Recall the TOA model of Eq. (2.32) is

$$\hat{\tau} = \tau + \xi,$$

where ξ is a Gaussian random variable $\mathcal{N}(0, \sigma^2)$ with

$$\sigma^2 = \frac{1}{8\pi^2\beta^2 R},$$

and R is the SNR. The associated CRLB is derived as

$$(\mathbf{J}_d)^{-1}_{TOA} = \frac{c^2}{8\pi^2\beta^2 \cdot R}, \tag{3.15}$$

or equivalently,

$$\sqrt{\text{var}(\hat{d})} \geq \frac{c}{2\sqrt{2}\pi} \cdot \frac{1}{\beta} \cdot \frac{1}{\sqrt{R}}. \tag{3.16}$$

Consider a CDMA signal with chip rate W. By using the relation between W and the effective bandwidth β of Eq. (2.29)

$$\beta = \frac{W}{\sqrt{3}},$$

we obtain

$$\sqrt{\text{var}(\hat{d})} \geq \frac{\sqrt{3}c}{2\sqrt{2}\pi} \cdot \frac{1}{W} \cdot \frac{1}{\sqrt{R}}. \tag{3.17}$$

Evidently, we are able to control the system performance by adjusting the chip rate W and/or the SNR R. Therefore, the TOA based method can perform well for long-range positioning. Figure 3.1 plots the lower bound of $\sqrt{\text{var}(\hat{d})}$ in Eq. (3.17) vs. the SNR, for various chip rates ranging from 2Mcps (the top curve) to 8Mcps (the bottom curve).

By now we have seen that the utility of the SS based method is limited to short-range positioning, while the TOA based method can be used in a wider area. For a

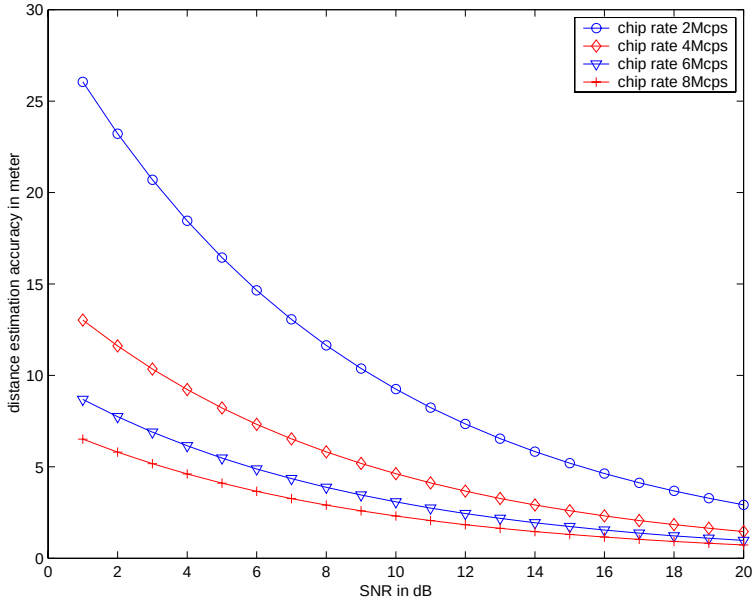


Figure 3.1: The distance estimation accuracy of Eq. (3.17) vs. the SNR with TOA data.

quantitative comparison of the "functioning ranges" of these two methods, we introduce the concept of "critical distance" by equating the lower bounds of Eqs. (3.16) and (3.13):

$$d_c = \frac{5\sqrt{3}c}{(\sqrt{2}\ln 10)\pi} \cdot \frac{\epsilon}{\eta} \cdot \frac{1}{W\sqrt{R}}. \tag{3.18}$$

Once the signal bandwidth W and the channel characteristics in terms of ϵ, η and SNR are specified, the critical distance d_c can be calculated. As a numerical example, we set $W = 5$Mcps, $\epsilon = 4$, $\eta = 8$, and $R = 0$dB. It follows that d_c is around 25m. Then we are able to predict that the TOA method should outperform the SS method in the range of $d > d_c$, and vice versa. On the other hand, we may use both TOA and SS data to improve a distance estimate. The corresponding CRLB is derived as

$$\begin{aligned} (\mathbf{J}_d)^{-1}_{TOA+SS} &= \frac{1}{(\mathbf{J}_d)_{TOA} + (\mathbf{J}_d)_{SS}} \\ &= \frac{1}{\frac{8\pi^2\beta^2 \cdot R}{c^2} + \left(\frac{10}{\ln 10}\right)^2 \cdot \frac{\epsilon^2}{\eta^2} \cdot \frac{1}{d^2}}, \end{aligned} \tag{3.19}$$

where we assume that the errors in the distance estimates from TOA and SS data are independent. It is straightforward to see

$$(\mathbf{J}_d)^{-1}_{TOA+SS} < \min\left\{(\mathbf{J}_d)^{-1}_{TOA}, (\mathbf{J}_d)^{-1}_{SS}\right\}, \tag{3.20}$$

where $\min\{a, b\}$ stands for the smaller value of a and b. That is to say, the distance estimation using both TOA and SS data can achieve higher accuracy than the estimation based on only one type of data. However, the improvement is not significant when $(\mathbf{J}_d)_{TOA} << (\mathbf{J}_d)_{SS}$ and $(\mathbf{J}_d)_{TOA} >> (\mathbf{J}_d)_{SS}$, which correspond to $d << d_c$ and $d >> d_c$, respectively.

These observations lead us to devise a hybrid distance estimation scheme, provided both TOA and SS data are available. Denote $\tilde{d}$ a rough estimate of d, e.g., based on some prior information. The scheme consists of three modes:

- The signal-strength mode. If any prior information suggests $\tilde{d} << d_c$, the SS measurements are the principal data to be employed, because the inclusion of TOA data will not make much improvement for the positioning accuracy.
- The hybrid mode. When $\tilde{d}$ is comparable with d_c, both TOA and SS data should be taken.
- The time-delay mode. The use of TOAs should be dominant for those remote BSs, i.e., when $\tilde{d} >> d_c$.

For the sake of clarity, we express the distance estimate[1] from each of the three modes in a unified formula:

$$\hat{d} = d + \zeta, \tag{3.21}$$

[1]In contrast to the rough distance estimation $\tilde{d}$, $\hat{d}$ denotes the estimate from the ML estimator.

where ζ is a estimation error, represented by a Gaussian variable $\mathcal{N}(0, \omega^2)$ with

$$\omega^2 = \begin{cases} \frac{c^2}{8\pi^2} \cdot \frac{1}{\beta^2 \cdot R}, & \text{for } \tilde{d} >> d_c, \\ \left(\frac{\ln 10}{10}\right)^2 \cdot \frac{\eta^2}{\epsilon^2} \cdot d^2, & \text{for } \tilde{d} << d_c, \\ \frac{1}{\frac{8\pi^2 \beta^2 \cdot R}{c^2} + \left(\frac{10}{\ln 10}\right)^2 \cdot \frac{\epsilon^2}{\eta^2} \cdot \frac{1}{d^2}}, & \text{for } \tilde{d} \sim d_c. \end{cases} \tag{3.22}$$

Switching among these three modes can be made automatically, depending on the mobile's location vis-a-vis a given BS.

We shall incorporate this hybrid scheme to a geolocation scheme and discuss several simulation results in Chapter 4.

3.3 Concluding Remarks

We have clarified the relationships among the three distance-based positioning methods: with Eq. (3.7), we confirm that TDOA positioning (with non-synchronous systems) cannot attain higher geolocation accuracy than TOA positioning (for synchronous systems); by defining the critical distance d_c in Eq. (3.18), we conclude that when an MS and a BS are separated more than d_c, the TOA based range estimation performs better than the SS based method, thus is more suitable for long-range geolocation, and vice versa. A hybrid distance estimation scheme is proposed by taking into account the tradeoff between the TOA and SS based schemes.

Appendix 3.1: Two Examples for $\mathbf{J}_{TDOA}^{-1} \leq \mathbf{J}_{TOA}^{-1}(l)$

We first provide an example where such an l exists. Consider L BSs, $L > 6$, include three pair of BSs. Each pair of BSs, say $(\text{BS}_i, \text{BS}_{i+3})$, for $i = 1, 2, 3$, is deployed in

such a way that the MS is located at the central point of the straight line connecting the two BS locations, which corresponds to

$$w_i = w_{i+3}, \quad \text{and} \quad \boldsymbol{h}_i = -\boldsymbol{h}_{i+3}.$$

By using the relation

$$\sum_{b=1}^{L} w_b \boldsymbol{h}_b \boldsymbol{h}_b^T \geq (\sum_{b=1}^{L} w_b \boldsymbol{h}_b)(\sum_{b=1}^{L} w_b \boldsymbol{h}_b)^T,$$

which is an immediate result from

$$E\left[(\boldsymbol{h} - \bar{\boldsymbol{h}}) \cdot (\boldsymbol{h} - \bar{\boldsymbol{h}})^T\right] \geq 0,$$

we have

$$\sum_{b=7}^{L} w_b \boldsymbol{h}_b \boldsymbol{h}_b^T \geq (\sum_{b=7}^{L} w_b \boldsymbol{h}_b)(\sum_{b=7}^{L} w_b \boldsymbol{h}_b)^T.$$

Combining the above equation and

$$\sum_{b=4}^{6} w_b \boldsymbol{h}_b \boldsymbol{h}_b^T \geq (\sum_{b=1}^{6} w_b \boldsymbol{h}_b)(\sum_{b=1}^{6} w_b \boldsymbol{h}_b)^T = \mathbf{0},$$

we obtain

$$\sum_{b=4}^{L} w_b \boldsymbol{h}_b \boldsymbol{h}_b^T \geq (\sum_{b=1}^{L} w_b \boldsymbol{h}_b)(\sum_{b=1}^{L} w_b \boldsymbol{h}_b)^T.$$

With Eq. (3.9), it is clear that

$$\mathbf{J}_{TOA}^{-1}(L) \leq \mathbf{J}_{TDOA}^{-1} \leq \mathbf{J}_{TOA}^{-1}(3),$$

i.e., $l = 3$ is a choice for this case.

For the second example, consider

$$\sum_{b=1}^{L} w_b \boldsymbol{h}_b \boldsymbol{h}_b^T = (\sum_{b \in \mathcal{L}} w_b \boldsymbol{h}_b)(\sum_{b \in \mathcal{L}} w_b \boldsymbol{h}_b)^T,$$

where

$$\boldsymbol{h}_i = \boldsymbol{h}_j, \quad \text{for all } i, j \in \mathcal{L}.$$

Hence, $l = 0$. It corresponds to the layout where all the BSs and the MS are lined up. Both the TOA and TDOA positioning system collapse in this circumstance, because of the infinite estimation errors as discussed in Section 2.2.

Chapter 4

A New Geolocation Approach

4.1 Introduction

The previous two chapters represent the theoretical contributions of the dissertation. We shall now turn to some practical issues regarding the geolocation problem. In this chapter, we propose a new geolocation approach by incorporating the so-called *sequential simplex method* (SSM) and some analytical results obtained earlier.

The least square (LS) method is adopted as a standard technique when more than three base stations are involved in geolocation [2]. It provides an acceptable accuracy in a LOS propagation environment. However, the performance will be considerably degraded when an NLOS propagation exists, which is often the case in a cellular system. Therefore, mitigation of the NLOS effects has been an important issue in wireless geolocation. Several methods, e.g. [2, 14, 15, 25], have been proposed in this field. In [14], the authors suggest that NLOS BSs can be distinguished from LOS ones based on deviations of TOA measurements from their mean, since the deviations tend to be much larger in NLOS data than in LOS data. If prior information of the TOA error statistics is available, the errors contained in the TOA measurements due

to NLOS propagation may be mitigated. However, this scheme requires prior information about the TOA error statistics, which may not always be available. In [2], a penalty function is proposed to modify the conventional LS cost function, by considering the fact that the NLOS induced delays $\boldsymbol{l}$ are always positive. The modified cost function is then minimized by a gradient search. According to our simulation results, this algorithm exhibits a slow convergence in many situations, and even no convergence in certain cases. In [25], we discuss our devised geolocation algorithm, where the NLOS geolocation is formulated as a constrained optimization problem. The objective function is based on the LS criterion, and is subject to two types of constraint. In addition to the aforementioned constraint $\boldsymbol{l} > \mathbf{0}$, we take into account the other constraint that an MS is either stationary or moving at a speed below some upper limit. The SSM serves as an optimization tool, which can handle the boundary conditions in a simple manner. Simulation results show that the accuracy of less than 100m can be obtained in typical cases. However, a major limitation of the algorithm is that it is not based on the ML or MAP estimator which can achieve better positioning accuracy than the LS estimator.

In order to achieve a higher positioning accuracy, combination of an optimum receiver and the SSM seems to be a promising solution. In this chapter, we propose a geolocation approach that incorporates both aspects. In addition, we shall implement the hybrid distance estimation scheme which takes advantage of the tradeoff between SS and TOA data as discussed in Section 3.2. Besides the two types of constraint described above and adopted in [25], the geographic conditions such as street and highway layouts will be considered to further limit the mobile's possible position.

The rest of the chapter is organized as follows. In Section 4.2, we introduce the fundamentals of the SSM. Our new geolocation approach is then presented in

Section 4.3. Section 4.4 provides the associated accuracy analysis. Simulation results are discussed in Section 4.5. Conclusions are given in the last section.

4.2 Fundamentals of Sequential Simplex Method (SSM)

The SSM [7] is an optimization technique that involves an iterative evaluation of a given objective function itself in contrast to the evaluation of the derivative of the objective function performed in typical gradient methods. We first introduce three basic rules the SSM must comply with, and then discuss advantages and limitations of this technique with respect to the gradient schemes.

The SSM takes a regular geometric figure (known as a simplex) as a basis. In a two-dimensional case, the simplex is an equilateral triangle. For three dimensions, we extend the equilateral pattern by adding one point such that there are four points lying at the vertices of a regular tetrahedron. The generalization to an n-dimensional case is straightforward. For the sake of simplicity, we shall illustrate the three basic rules of this technique in examples of two-dimensional case.

The minimization of an unconstrained objective function, say $f(x, y)$, will be considered. Three points A, B and C are initially chosen so that they form an equilateral triangle as shown in Figure 4.1. This figure depicts a contour map of $f(x, y)$. The value of the objective function is then calculated for the coordinate (x, y) corresponding to each of the three vertices. The value of $f(x, y)$ at A is the largest, thus the worst, among the three. Next, we remove the worst point (i.e., the point A) from the triangle and to replace it with point D, which is the mirror image of A with respect to the line connecting the other two vertices (B and C). We thus

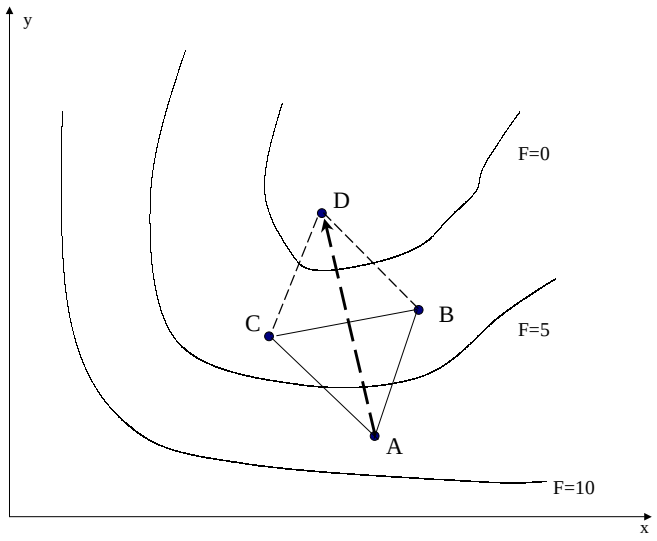


Figure 4.1: Triangle ABC is the starting simplex, and the new point D is found in the first search.

get a new equilateral triangle CBD, which can be used as the basis for the next move. It then follows that the value of $f(x, y)$ is evaluated at D, and a comparison is made among the values of $f(x, y)$ at the points B, C, and D. As this procedure repeats itself, the direction of a new round movement is always away from the worst point. Therefore, our search for an optimum point proceeds in a desired direction. This is *Rule 1* of the SSM.

However, a closed cycle of operations may be entered if this rule is applied to the circumstance where the triangle at an intermediate step straddles a ridge formed by a given objective function, such as the example shown in Figure 4.2. Since the value $f(x, y)$ at A is the worst of the three, Rule 1 replaces A by its mirror image A_1. Thus the new triangle A_1BC is formed. However, A_1 is the worst among A_1, B and C. The same rule rejects A_1 and puts A back, recreating the original triangle ABC. In this way, the search simply oscillates between A and A_1, and no further moves can be

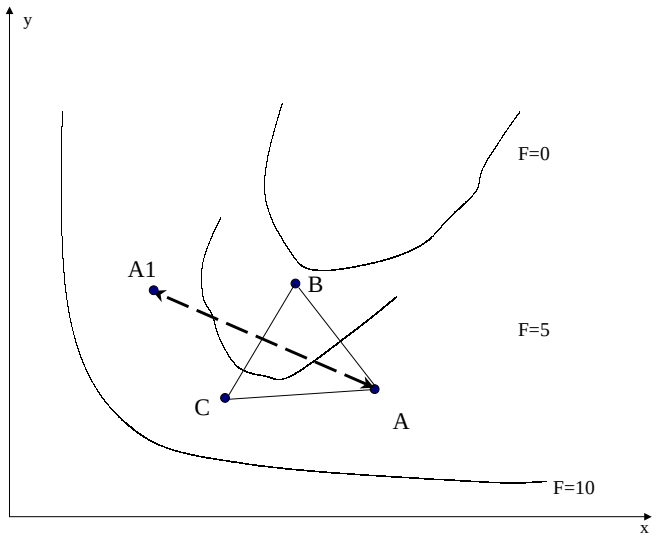


Figure 4.2: Failure of *Rule 1* on a ridge. The closed cycle of operations $A \leftrightarrow A_1$ is entered.

made. This obstacle can be easily overcome, however, if we introduce *Rule 2* that no return is allowed to the point that has just been left. In the example of Figure 4.2, it means that instead of rejecting the worst point A in the original triangle, we remove the point at which $f(x, y)$ has the second-worst value, i.e., the point C.

Obeying the above two rules, the search process can be carried out until the region containing the optimum point is attained. The convergence of the SSM to an optimal point is guaranteed by the following rule, i.e., *Rule 3*. When the number of iterations with one vertex fixed at the same point exceeds some number, we reduce the size of the working triangle, e.g., by one half, to improve the accuracy of locating the optimum point. The same procedure as before is resumed with this reduced triangle. The search can finally be terminated when the triangle size is small enough. The smallest size can be set to be comparable with some required accuracy, or the estimation accuracy of the optimum point based on certain error analysis.

As we see, the SSM is significantly different from those widely-practiced gradient-based optimization techniques in that it focuses on an iterative evaluation of a given objective function itself. Accordingly, the SSM has the distinctive advantage that it can accommodate the discontinuities of an objective function and complicated boundary conditions, where the derivative of the function does not exist or is difficult to compute. However, there is one major limitation in this technique: in each searching move, the SSM usually cannot determine the best direction. In contrast, many gradient methods have such capabilities. The direction that the SSM takes in a given step depends on the orientation of the simplex (or the triangle in our examples) obtained in the previous step. Yet the overall direction of several consecutive movements should follow closely the direction determined by the conventional gradient methods. Therefore, for regular unconstrained optimization problems (which are typically solved by a gradient method), the SSM will require more iterative steps to reach a convergence than the gradient method.

In order to take advantage of its merit and limit the shortcoming, we shall adopt the SSM to perform a "refining" estimation so that the initial point of the SSM search is not very far from the optimum point. To be more specific, in our following geolocation algorithm, the start point to search for the MS position will be its estimate obtained in the previous measurement, which is available in the case of mobile tracking, or a coarse estimate determined by some less accurate but fast-convergence approach.

Another point worth mentioning is that as an iterative method the SSM may terminate its searching process in a local optimum region, instead of the global optimum as is desired. However, for the geolocation problem, especially for the case of mobile tracking, the convergence to global optimum points are suggested by most of our

simulation results.

4.3 Algorithm Description

Before we present our geolocation approach, it is helpful to review briefly the three relevant results we obtained in the previous chapters:

- The ML or MAP estimator is an optimum receiver for the TOA positioning method, depending on whether the prior information of NLOS delays is available or not. (See Section 2.2.3 for the ML estimator and Section 2.2.5 for the MAP estimator.)

- Once the signal bandwidth W and the channel characteristics in terms of the path loss factor ϵ, the log-normal shadowing index η and SNR are specified, the critical distance d_c can be calculated. Then the distance estimation based on TOA should outperform the one based on SS in the range of $d > d_c$, and vice versa (Recall d is the total propagation path length between the MS and a given BS). The hybrid distance estimation scheme is to take advantage of both SS and TOA data:

  - When $\tilde{d} << d_c$, the SS measurements are the principal data to employ (Recall $\tilde{d}$ is a rough estimate of d).
  - When $\tilde{d}$ is compatible with d_c, both TOA and SS data should be taken.
  - The use of TOAs should be dominant for those remote BSs, i.e., when $\tilde{d} >> d_c$.

  (Refer to Section 3.2.)

- As an optimization tool, the SSM can handle a non-differentiable objective function and irregular boundaries in a simple manner.

The new NLOS geolocation algorithm is designed to incorporate the above results. We discuss the algorithm in the following three steps.

Step 1. Data Preparation

We first determine the critical distance (see its definition of Eq. (3.18) for details)

$$d_c = \frac{5\sqrt{3}c}{(\sqrt{2}\ln 10)\pi} \cdot \frac{\epsilon}{\eta} \cdot \frac{1}{W\sqrt{R}}$$

for each BS, and obtain a rough estimate $\tilde{d}$ of the distance between the MS and the BS, e.g., by using the MS position estimate obtained in the previous measurement.

We then extract TOA and/or SS data by processing the received signals at each BS. According to the hybrid distance estimation scheme, we define the following two subsets of $\mathcal{B}$ (recall $\mathcal{B}$ is the set of all the BSs): set $\mathcal{S}$ of the BSs that produce SS measurements, and set $\mathcal{T}$ of the BSs that provide TOA data. Here is another way to interpret the two subsets: the subset $\mathcal{S}$ should contain the BSs within the range $\tilde{d} << d_c$ and $\tilde{d} \sim d_c$, while $\mathcal{T}$ includes the BSs at distances $\tilde{d} >> d_c$ and $\tilde{d} \sim d_c$. It is understood that the set $\mathcal{S} \cap \mathcal{T}$ corresponds to those BSs with $\tilde{d} \sim d_c$, where both TOA and SS data are adopted to perform a distance estimation. Note $\mathcal{B}$ can also be divided into two disjoint subsets, i.e., $\mathcal{L}$ for the LOS BSs and $\mathcal{NL}$ for the NLOS stations. When NLOS induced error is present in some TOA estimate, we shall include the corresponding SS data for the distance estimation in order to improve the estimation accuracy.

Next, we estimate the distance between the MS and a BS by using TOA or SS data. Recall the distance estimate extracted from TOAs is

$$\hat{d_b} = d_b + l_b + \zeta_b, \quad b \in \mathcal{T}, \tag{4.1}$$

where

$$d_b = \sqrt{(x - x_b)^2 + (y - y_b)^2},$$

$l_b = 0$ for $b \in \mathcal{L}$, and ζ_b is a Gaussian variable $\mathcal{N}(0, \omega_b^2)$ with

$$\omega_b^2 = \frac{c^2}{8\pi^2} \cdot \frac{1}{\beta^2 \cdot R_b}.$$

The distance estimate from SS data is

$$\hat{d}_b = d_b + \zeta_b, \quad b \in \mathcal{S}, \tag{4.2}$$

where ζ_b is a Gaussian variable $\mathcal{N}(0, \omega_b^2)$ with

$$\omega_b^2 = \left(\frac{\ln 10}{10}\right)^2 \cdot \frac{\eta^2}{\epsilon^2} \cdot d_b^2. \tag{4.3}$$

Step 2. Formulation of the constrained optimization problem

We first formulate an objective function by adopting the ML or MAP estimator based on the distance estimates obtained in the previous step. For a LOS environment or situation where no NLOS statistics are available beforehand, the objective function is the negative of logarithm of the likelihood function associated with the MLE, i.e.,

$$\begin{aligned} O_L(\boldsymbol{p}) \;=\; & \sum_{b \in \mathcal{T} \cap \mathcal{L}} \frac{1}{\omega_b^2} \left(\hat{d}_b - \sqrt{(x - x_b)^2 + (y - y_b)^2}\right)^2 \\ & + \sum_{b \in \mathcal{S}} \frac{1}{\omega_b^2} \left(\hat{d}_b - \sqrt{(x - x_b)^2 + (y - y_b)^2}\right)^2. \end{aligned} \tag{4.4}$$

Note the above formulation excludes the NLOS TOA estimates, due to Proposition 2.1. From Eq. (4.3), we see that ω_b^2, $b \in \mathcal{S}$, is a function of the distance d_b which is impossible to obtain without first locating the MS position $\boldsymbol{p}$. Our solution is to use the position estimate $\hat{\boldsymbol{p}}$ instead of the true $\boldsymbol{p}$ in evaluating these ω_b^2's. In addition, $\hat{\boldsymbol{p}}$ is updated together with these ω_b^2's at each iteration in the optimization procedure

as will be discussed in the third step. On the other hand, when the prior p.d.f of the NLOS delays, $p_{\boldsymbol{l}}(\boldsymbol{l})$, is available, the objective function is expressed in terms of the logarithm of the posterior probability function involving all the distance estimates:

$$\begin{aligned} O_{NL}(\boldsymbol{p},\boldsymbol{l}) &= O_L(\boldsymbol{p}) - \ln p_{\boldsymbol{l}}(\boldsymbol{l}) \\ &\quad + \sum_{b\in\mathcal{T}\cap\mathcal{NL}} \frac{1}{\omega_b^2}\left(\hat{d}_b - \sqrt{(x-x_b)^2+(y-y_b)^2} - l_b\right)^2 \\ &= \sum_{b\in\mathcal{T}} \frac{1}{\omega_b^2}\left(\hat{d}_b - \sqrt{(x-x_b)^2+(y-y_b)^2}\right)^2 \\ &\quad + \sum_{b\in\mathcal{S}} \frac{1}{\omega_b^2}\left(\hat{d}_b - \sqrt{(x-x_b)^2+(y-y_b)^2}\right)^2 - \ln p_{\boldsymbol{l}}(\boldsymbol{l}). \end{aligned} \tag{4.5}$$

We now establish three types of constraint functions. In practice, the movement of an MS is usually confined within certain regions, such as roads and parking areas. Hence, the first type of constraint is to address such limitation:

$$\textit{Constraint 1:} \quad (x,y) \in C, \tag{4.6}$$

where C represents a possible region for the mobile's position. Since the MS will be either stationary or moving with a speed below some finite value, typically, 30m/s, the incremental change of the MS position from its position at the previous measurement instant is also bounded by some finite number, denoted by $D(t,\delta t)$. We thus formulate the second type of constraint as

$$\textit{Constraint 2:} \quad |\boldsymbol{p}(t) - \boldsymbol{p}(t-\delta t)| \le D(t,\delta t). \tag{4.7}$$

Denote $v_m(t,\delta t)$ the maximum speed of MS during time interval $[t-\delta t, t]$ and $\mathcal{P}_m(t)$ the maximum positioning error at time t. The explicit expression of $D(t,\delta t)$ is

$$D(t,\delta t) = v_m(t,\delta t)\cdot\delta t + \mathcal{P}_m(t-\delta t). \tag{4.8}$$

The third constraint takes into account that an NLOS induced path length is always positive:

$$\textit{Constraint 3:} \quad \boldsymbol{l} > \mathbf{0}. \tag{4.9}$$

If any of the above three constraints is violated at a certain location, we shall set the corresponding value of the objective function to be infinite. The first two constraints can be interpreted as "hard limits" on the possible region of the mobile's position. In case some prior p.d.f of the MS position is also available, we should include the logarithm of such p.d.f in the objective function of Eq. (4.5) as a "soft limit".

Step 3. Optimization of the constrained objective function with the SSM

Estimate the MS position by

$$\min O_L(\boldsymbol{p}), \quad \text{subject to} \quad \textit{Constraint 1, 2,} \tag{4.10}$$

when no information of the NLOS delays is available, or otherwise

$$\min O_{NL}(\boldsymbol{p}, \boldsymbol{l}), \quad \text{subject to} \quad \textit{Constraint 1,2,3.} \tag{4.11}$$

The SSM serves as an optimization tool. For simplicity, we assume a mobile tracking scenario here so that the position estimate obtained in the previous measurement can be used as the initial point to search for the current MS position. In a general "cold start" case, we may obtain the initial point by a coarse estimation with some fast-converging optimization technique, e.g., a gradient method. The optimization of $O_L(\boldsymbol{p})$ in Eq. (4.10) is a two-dimensional problem. Thus, the simplex for the SSM is an equilateral triangle. The smallest size of the triangle (see *Rule 3* of the SSM) is selected to be comparable with the CRLB based on $\hat{\boldsymbol{p}}(t - \delta t)$. When the prior p.d.f

of the NLOS delays, $p_{\boldsymbol{l}}(\boldsymbol{l})$, is available, the optimization of $O_{NL}(\boldsymbol{p}, \boldsymbol{l})$ in Eq. (4.11) is more complicated. In addition to $\boldsymbol{p}$, the NLOS delays $\boldsymbol{l}$ need to be estimated as well. Since the estimation of $\boldsymbol{p}$ and $\boldsymbol{l}$ may have different precision requirements, we use two types of simplex, an equilateral triangle for $\boldsymbol{p}$ and an M-dimensional simplex for the M-dimensional vector $\boldsymbol{l}$. Accordingly, there are two searching processes: the one for $\boldsymbol{p}$ is made in a two dimensional plane, and the other for $\boldsymbol{l}$ is in an M-dimensional space. The two processes are carried out in turn. Optimum estimates are obtained when both processes indicate convergence. The initial M-dimensional simplex is constructed around the center of an arbitrary feasible point.

4.4 Accuracy Analysis

The above geolocation scheme involves both TOA and SS data. An easy way to obtain its corresponding CRLB and G-CRLB is to modify the CRLB of Eq. (2.17) and the G-CRLB of Eq. (2.50) derived for the TOA positioning method, by accommodating the contribution from the SS data. The only quantities we need to modify are $\boldsymbol{\Lambda}_{NL}$ and $\boldsymbol{\Lambda}_{L}$ in Eq. (2.10). We replace the diagonal terms of

$$\boldsymbol{\Lambda}_{NL} = \mathrm{diag}\left(\lambda_1,\ \lambda_2,\ \cdots, \lambda_M\right),$$

and

$$\boldsymbol{\Lambda}_{L} = \mathrm{diag}\left(\lambda_{M+1},\ \lambda_{M+2},\ \cdots, \lambda_B\right),$$

by

$$\lambda_b = \frac{c^2}{w_b^2},\ b = 1, 2, \cdots, B. \tag{4.12}$$

where

$$\omega_b^2 = \begin{cases} \frac{c^2}{8\pi^2} \cdot \frac{1}{\beta^2 \cdot R_b}, & \text{for } b \in \mathcal{B}/\mathcal{S}, \\ \left(\frac{\ln 10}{10}\right)^2 \cdot \frac{\eta^2}{\epsilon^2} \cdot d_b^2, & \text{for } b \in \mathcal{B}/\mathcal{T}, \\ \frac{1}{\frac{8\pi^2\beta^2 \cdot R_b}{c^2} + \left(\frac{10}{\ln 10}\right)^2 \cdot \frac{\epsilon^2}{\eta^2} \cdot \frac{1}{d_b^2}}, & \text{for } b \in \mathcal{S} \cap \mathcal{T}, \end{cases} \tag{4.13}$$

which was given in Eq. (4.13).

Therefore, the CRLB and the G-CRLB for the new approach can be expressed still in the forms of Eq. (2.17) and Eq. (2.50), respectively.

4.5 Simulation Results

The cellular CDMA system described in Section 2.2.2 will be used in the simulation experiments in this section. We first introduce an illustrative picture to see how the SSM works in our geolocation scheme. Then some simulation results are investigated. The standard deviation of the position estimates, i.e.,

$$\mathcal{E} = \sqrt{E \parallel \hat{\boldsymbol{p}} - \boldsymbol{p} \parallel^2},$$

is adopted as the measure of the estimation performance. One hundred simulation runs are executed to produce the average in the above definition.

Simulation 4.1.

Figure 4.3 helps to describe the optimization steps of the SSM in our proposed scheme. The unshaded area represents a road which confines a likely mobile's position, corresponding to Constraint 1 in Eq. (4.6). Symbol "*" denotes the initial point, which is at the center of the initial triangle (or simplex). The curved line is determined by

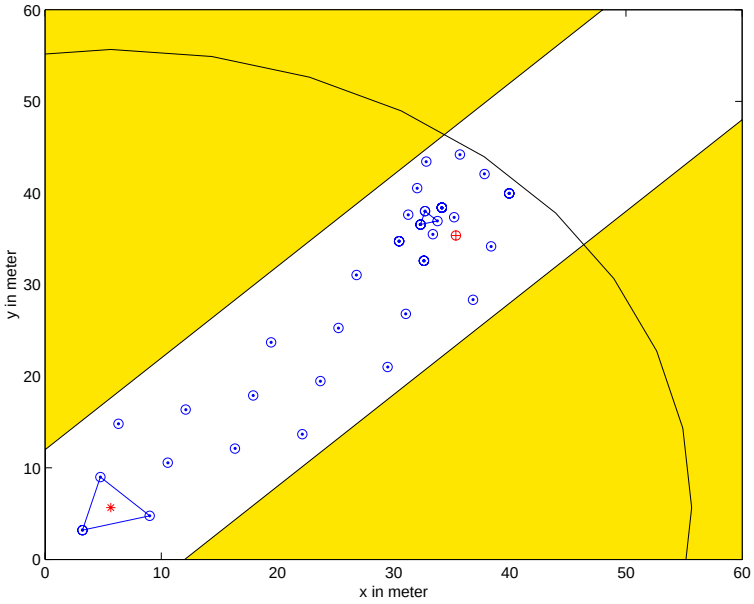


Figure 4.3: An illustrative picture of the application of the SSM in the proposed geolocation method.

the maximum possible separation between the initial point and an MS estimate, i.e. Constraint 2 of Eq. (4.7), The true MS position is marked by symbol "$\oplus$", and the nearby reduced triangle is created at the final searching step when the SSM achieves a convergence. The marks "$\odot$" trace the intermediate searching points during the optimization, and are well confined within the feasible region. It is indicated that the size of the moving triangle is reduced as the search approaches to the neighborhood region of the optimum point, i.e., the center of the smaller triangle (see *Rule 3* of the SSM).

Simulation 4.2.

Here we consider an example in which combined use of SS and TOA may outperform a scheme that uses TOA or SS data only. Possible application scenarios may include indoor or semi-indoor geolocation, such as locating an individual in an office building,

hospital, nursing home, airport lobby and railroad station. It may be also applicable to locate a car or pedstrian in urban environments (e.g., New York City) where LOS signals are absent or weak. In such applications the distance between BSs may be on the order of tens of meters instead of thousand meters. In some applications, the position information of z-coordinate (i.e., height) as well as the (x, y) coordinates may be required. Our formulation based on the two-dimensional model is easily extended to the three-dimensional case. No prior information of NLOS delays is assumed. Four LOS BSs, i.e., BS_1, BS_2, BS_3 and BS_4 are involved in the geolocation. The distance between two adjacent BSs is 30m. The chip rate of CDMA signals is 2Mcps. The path loss factor ϵ and the log-normal index η are 4 and 6, respectively. The MS is located at $20(\cos(\pi/4), \sin(\pi/4))$. We use $5(\cos(\pi/4), \sin(\pi/4))$ as a start point. A straight 10m-wide road serves as Constraint 1, similar to that shown in Figure 4.3. The center line of the road is along $y = \tan(\pi/4)x$. The maximum separation between the initial point and an MS estimate is set to be 25m (Constraint 2). Figure 4.4 shows the performance measure $\mathcal{E}$ of the scheme vs. the SNR, using TOA, SS and both types of data, respectively. The SNR is defined at a position 30m from the MS. It is confirmed that adopting both types of data (i.e., the hybrid scheme) can achieve better positioning accuracy than use of TOA or SS measurements only. However, such advantage over the TOA based scheme is not significant when the difference between the accuracy of TOA data based estimate and that of SS based estimate is sufficiently large, e.g., when the SNR is higher than 16dB in this case, as shown in Figure 4.4. It is also observed the performance of the SS based method dose not improve as the SNR increases, because "noise" in the SS data (see Eq. (2.111)) comes from the log-normal fading effects and the actual noise at receivers does not count.

Simulation 4.3.

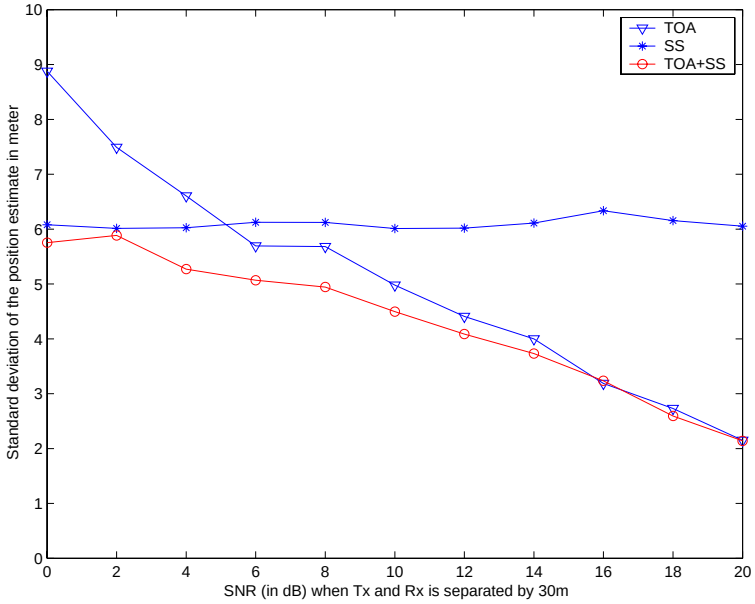


Figure 4.4: Comparison of the performance of the proposed scheme using SS, TOA and combination of SS and TOA data.

By using the same set of model parameters as the previous simulation, we show in Figure 4.5 the performance of the hybrid scheme with or without Constraints 1 and 2. The square root of the trace of the CRLB, $\mathcal{P}_{CR}$, is drawn (the second curve from top), which is a theoretical limit for the positioning accuracy when no prior information of MS position is available. As expected, the scheme without the constraints (corresponding to the top curve) cannot perform better than the CRLB limit, but can attain the CRLB as the SNR grows sufficiently high. When the constraints are applied, the scheme yields higher positioning precision (the lower two curves) even than the CRLB, and the lower two curves converge to the CRLB when the SNR is above 14dB. It is seen that the amount of improvement decreases as the SNR increases, which is determined by how tight the constraints can confine an MS estimated with respect to the CRLB. We shall explore the relationship between the "tightness" of Constraint 1 and the performance enhancement in the next simulation.

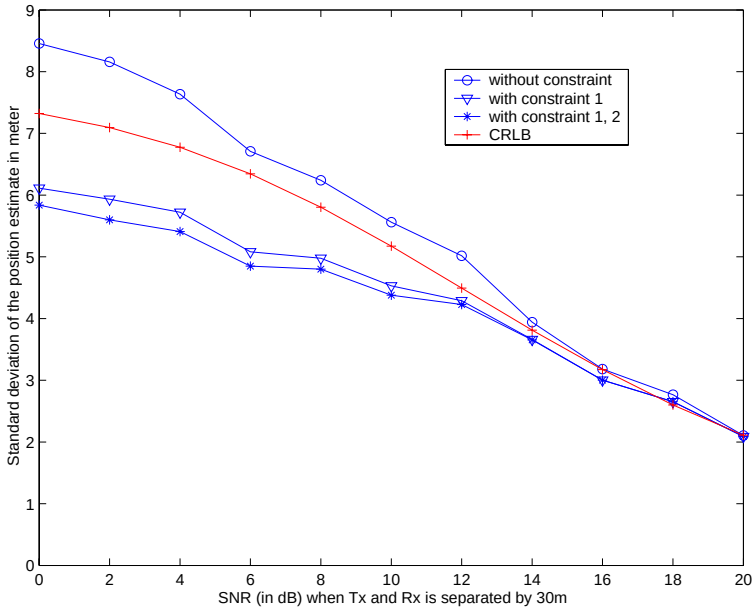


Figure 4.5: The performance of the hybrid scheme with and without constraints on the MS position.

Simulation 4.4.

In the remaining two examples, we consider outdoor geolocation in a wide region. From the discussion given in Section 3.2, geolocation now should rely mainly on TOA data. The distance between two adjacent BSs is set to be 2000m. The SNR is -3dB when the MS and a receiver is separated by 2000m. The chip rate of CDMA signals is still 2Mcps. The MS is located at $600(\cos(\pi/4), \sin(\pi/4))$. We use $500(\cos(\pi/4), \sin(\pi/4))$ as an initial point. The road extends along the line of $y = \tan(\pi/4)x$ as before. The maximum separation between the initial point and an MS estimate is chosen to be 140m (Constraint 2). Since $\mathcal{P}_{CR}$ is approximately 31m as shown in Figure 4.6, Constraint 2 is a rather loose boundary for an MS estimate. Hence, the performance enhancement due to the constraints, if there is any, should result from Constraint 1, or the width of the road denoted by W_{rd}. For convenience,

we define the normalized size of Constraint 1 as

$$sz \stackrel{\text{def}}{=} \frac{W_{\text{rd}}/2}{\mathcal{P}_{CR}}. \tag{4.14}$$

Figure 4.6 shows $\mathcal{E}$ vs. sz in dB. The associated CRLB data are plotted as a comparison. It shows that the the performance enhancement is significant when $W_{\text{rd}}/2$ is

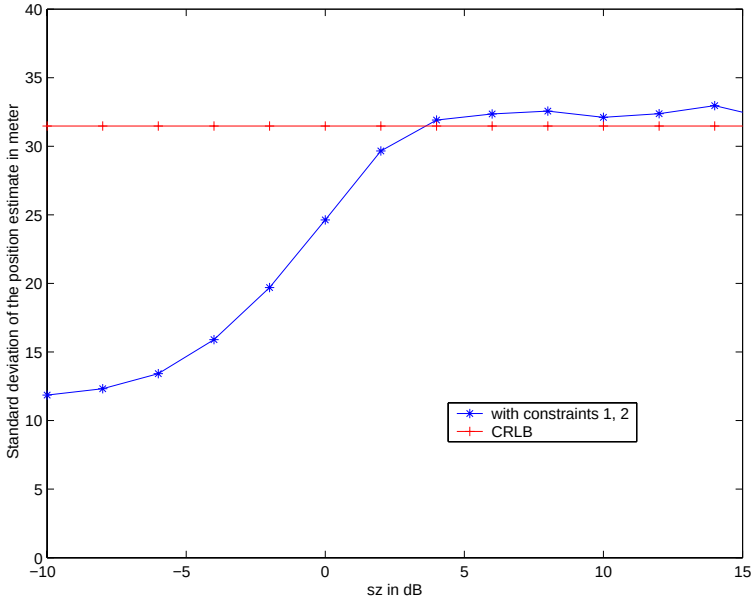


Figure 4.6: The performance of the proposed geolocation scheme with various road width.

less than $\mathcal{P}_{CR}$ (corresponding to $sz \leq 0$dB in the figure). However, further improvement is negligible once sz becomes less than -5dB, or $W_{\text{rd}} = 10$m. This is because Constrain 2 is fixed, and is not so tight as Constraint 1.

Simulation 4.5.

In this last simulation, we assume that prior information of NLOS delays is available. There are three LOS BSs, BS_1, BS_2, BS_3 and two NLOS stations, BS_4 BS_5. The path loss ϵ is 4 for NLOS signals. The NLOS delays are assumed to be Gaussian distributed with $\mathcal{N}(90, 15^2)$. Constraint 3, or that NLOS delays are always positive,

is adopted in the simulation. The road width is 20m. Other simulation parameters are same as those of the previous experiment. Figure 4.7 shows the deviation $\mathcal{E}$ vs. the SNR when the distance between the MS and a receiver is 2000m, in three situations: neither NLOS information or the constraints are available (the top curve), only NLOS information is known (the second curve from the top), and both types of knowledge can be obtained (the bottom line). As expected, the third situation

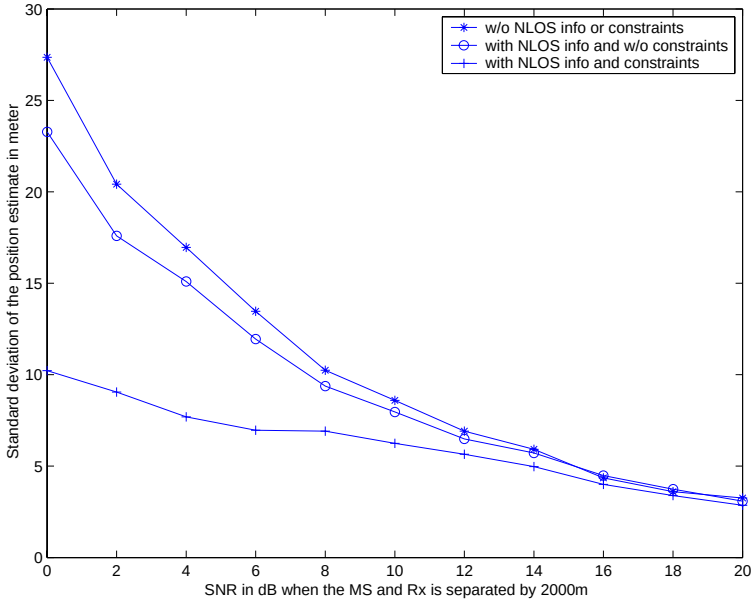


Figure 4.7: The performance of the proposed scheme with and without NLOS delay information and constraints.

exhibits the best performance (the bottom curve). However, its improvement over the other two approaches decreases as the SNR grows higher. It is observed that the enhancement due to NLOS information (the gap between the top two curves) is smaller than that resulting from the limitation on an MS estimate (the difference between the lower two curves). This is because the SNR of NLOS signals is greatly reduced by the extra NLOS propagation path and $\epsilon = 4$, and the weights for these NLOS signals in the objective function of Eq. (4.5) are much smaller than those for

LOS signals.

4.6 Concluding Remarks

We present a new geolocation approach using the SSM, where the location estimation is formulated as a constrained optimization problem. The objective function is constructed based on the ML or MAP estimator, subjected to the three types of constraint. The simulation results show that the SSM is an efficient optimization tool which can accommodate the boundary conditions, and introducing appropriate constraints can greatly improve the geolocation accuracy.

Chapter 5

A Wavelet Approach to Channel Estimation

Up to now, we have been concerned with the geolocation in a single (LOS or NLOS) path propagation environment. In many practical situations, however, a received signal between a given MS and BS is inevitably subject to multipath propagation. It has been well understood that the best reception scheme is, as far as recovery of an information sequence is concerned, to make use of all available multipath signal components, often referred to as *fingers* in the so-called *rake receiver* [44]. For the geolocation problem, however, the conventional wisdom is to extract the first arriving signal component, or a dominant component, and estimate its associated time delay, followed by use of a single-path based algorithm. Yet it is intuitively clear that the best possible geolocation scheme should take into account all the available signal components, since not only the primary signal component but also the secondary, tertiary and other components carry the information regarding the position of the MS. Therefore, channel estimation is an important issue in geolocation.

In this chapter, we develop a simple wavelet-based approach to channel estimation. We should note that this approach is applicable to identification of a linear system,

not just estimation of a radio channel for geolocation.

5.1 Introduction

Channel estimation has been a critical problem in many communication systems, especially when the characteristics of a channel are time-varying. Besides the multipath geolocation problem as has been discussed above, a large class of interference cancellation algorithms and decoding algorithms require complete or partial knowledge of channel characteristics. A generic strategy is to transmit (or input) a signal sequence, which may or may not carry data information, and estimate the corresponding channel impulse response, or its related parameters, by processing the received signal. Channel estimation schemes can be classified into (i) training-based, (ii) blind and (iii) semi-blind approaches depending on the extent of use of a training signal [43]. In a training-based scheme, a receiver uses a known input signal (i.e., a training sequence) as a reference to process the output of a channel in the presence of additive noise, and estimate the channel response. Adaptive equalization techniques [44] are often employed in such schemes. More recently, semi-blind and blind algorithms have been extensively studied (see [45, 46, 47] and references therein), where the observation corresponding to unknown data are utilized. Compared with those training-based methods, the semi-blind and blind schemes may achieve higher data transmission rate by using less or no training signals, but at the cost of lower estimation accuracy and increased computation complexity. Although it can be viewed as a training-based scheme, the approach in [48] is distinguished in that it emphasizes the design of an input signal waveform particularly in favor of a simple channel reconstruction procedure. A Gaussian-noise-like signal is proposed to be used as an input signal. Then

channel estimation is conducted by a correlation receiver.

In this chapter, we propose a wavelet-based approach to channel estimation, which is motivated by a consideration similar to that of [48]. We design a training signal using the *scaling function* associated with a *compactly supported orthonormal wavelet*, which takes advantage of two properties of the scaling function: the bandwidth efficiency and orthogonality between the function and its time-shift replica. By sampling the received signal in a proper manner, we can obtain the projection of the channel impulse response onto the subspace spanned by the scaling functions, which allows a simple yet accurate channel reconstruction.

We first in Section 5.2 review fundamentals of compactly supported orthonormal wavelets and their applications to some communication problems. The basic idea of our wavelet-based approach to channel estimation is described in Section 5.3. An error analysis is presented in Section 5.4. In Section 5.5 we apply this approach to the estimation of a time-invariant channel with additive noise and to the problem of tracking a time-varying channel. Some simulation results are presented in Section 5.6. Section 5.7 gives the conclusion of this study.

5.2 Fundamentals of Compactly Supported Orthonormal Wavelets

We limit our discussion to compactly supported orthonormal wavelets in the real domain, which will facilitate the development of our wavelet approach in the remaining sections.

Compactly supported orthonormal wavelets have been extensively explored in the

past fifteen years [50]. In general, these wavelets and their associated scaling functions possess the following three important properties[1]:

- These waveforms have an approximately compact support in the frequency domain, often referred to as the bandwidth efficiency.
- Orthogonality is guaranteed between a wavelet (or a scaling function) and its time-shift replica.
- Their dilation and time-shift replicas constitute an orthonormal basis for $L^2(\mathcal{R})$, which is the foundation of the well-known *multi-resolution analysis*.

Recall that $L^2(\mathcal{R})$ is a functional space that contains all "signals" with finite energy, or

$$L^2(\mathcal{R}) = \left\{ f(t) \;\middle|\; \int_{-\infty}^{+\infty} f^2(t)\, dt < \infty \right\}.$$

In what follows, we shall expand the above three properties and emphasize their engineering interpretations (see Chapter 6 in [8] for a rigorous mathematical discussion).

Given a compactly supported orthonormal wavelet, denoted by $\psi(t)$, there exists a unique scaling function $\phi(t)$ associated with it, satisfying

$$\phi(t) = \sqrt{2} \sum_n c_n \cdot \phi(2t - n), \tag{5.1}$$

and

$$\psi(t) = \sqrt{2} \sum_n (-1)^n \cdot c_{-n+1} \cdot \phi(2t - n), \tag{5.2}$$

where coefficients c_n are determined through the construction procedure of the wavelet, and the number of these coefficients is finite. Here is an illustrative example. The Daubechies wavelet-8[1], traditionally denoted by $\psi^8(t)$, and its Fourier transform $\Psi^8(t)$

[1] These properties are more or less shared by other types of wavelets.

[1] All the Daubechies wavelets, $\psi^1(t), \psi^2(t), \cdots, \psi^n(t), \cdots$, are compactly supported orthonormal wavelets. The construction of these wavelets follows a unified procedure.

are shown in Figure 5.1 (a) and (b), respectively, together with the associated scaling function $\phi^8(t)$ and the Fourier transform $\Phi^8(f)$. The solid curves correspond to $\psi^8(t)$ and $\Psi^8(f)$, and the dotted ones are $\phi^8(t)$ and $\Phi^8(f)$. By referring to Figure 5.1, we can understand easily the following claims regarding a compactly supported orthonormal wavelet and the Daubechies wavelets:

- As the terminology "compactly supported" suggests, a compactly supported orthonormal wavelet $\psi(t)$, together with its associated scaling function $\phi(t)$, has a finite duration in the time domain. In addition, their Fourier transforms $\Psi(f)$ and $\Phi(f)$ have an approximately compact support in the frequency domain.

- The wavelet $\psi(t)$ can be viewed as a band-pass signal. In contrast, $\phi(t)$ is a base-band signal, whose bandwidth is roughly the same as that of $\psi(t)$.

- The larger the index n of the Daubechies scaling function $\phi^n(t)$ is, the more localized its Fourier transform $\Phi^n(f)$ is in the frequency domain. However, a further improvement is not significant once $n > 10$. (Hence we shall select $\phi^8(t)$ in our simulation.)

Define the dilation and time-shift versions of $\psi(t)$ and $\phi(t)$ as

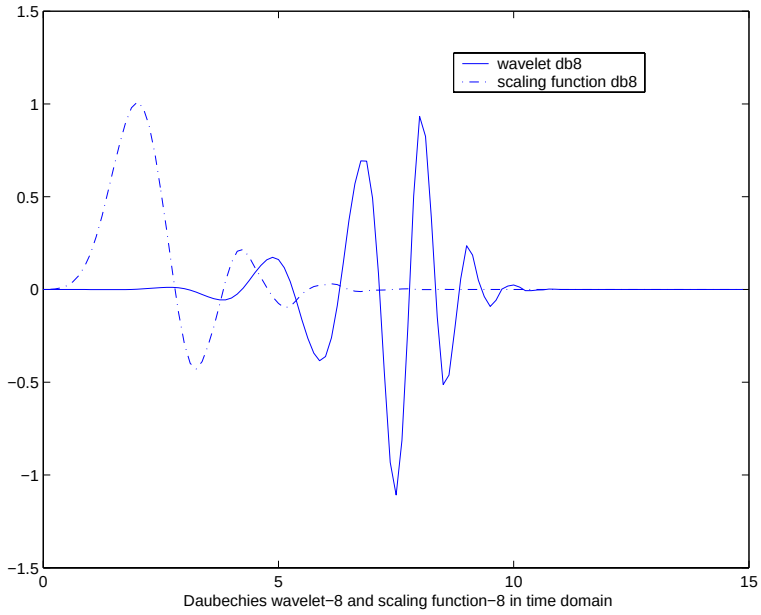
$$\psi_{j,k}(t) \stackrel{\text{def}}{=} 2^{j/2} \cdot \psi(2^j t - k), \tag{5.3}$$

and

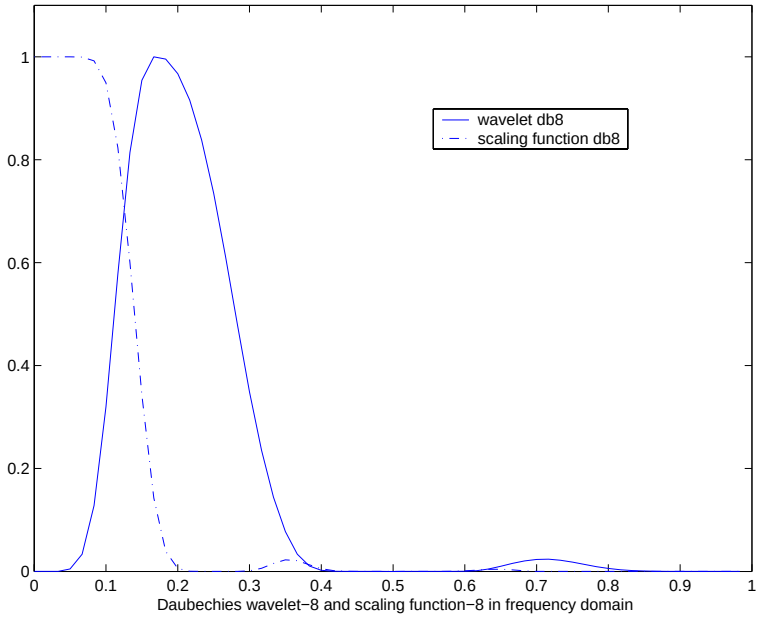
$$\phi_{j,k}(t) \stackrel{\text{def}}{=} 2^{j/2} \cdot \phi(2^j t - k), \tag{5.4}$$

respectively, where j and k are integers, i.e., $j, k \in \mathcal{Z}$. For fixed j, there exist orthonormal relations:

$$\langle \psi_{j,k}, \psi_{j,l} \rangle = \delta_{k,l}, \tag{5.5}$$



(a)



(b)

Figure 5.1: The Daubechies wavelet-8 $\psi^8(t)$ and scaling function-8 $\phi^8(t)$ (a) in the time domain and (b) in the frequency domain.

and

$$\langle \phi_{j,k}, \phi_{j,l} \rangle = \delta_{k,l}, \tag{5.6}$$

where $\delta_{k,l}$ is the Kronecker delta, i.e.,

$$\delta_{k,l} = \begin{cases} 1, & \text{for } k = l, \\ 0, & \text{for } k \neq l. \end{cases}$$

Furthermore, these $\psi_{j,k}(t)$'s and $\phi_{j,k}(t)$'s provide orthonormal bases for two types of subspaces:

$$W_j \stackrel{\text{def}}{=} \text{span}\left\{\psi_{j,k}(t), k \in \mathcal{Z}\right\}, \tag{5.7}$$

and

$$V_j \stackrel{\text{def}}{=} \text{span}\left\{\phi_{j,k}(t), k \in \mathcal{Z}\right\}. \tag{5.8}$$

Roughly speaking, subspace V_j consists of the base-band signals within the bandwidth comparable with that of $\phi_{j,0}(t)$, while W_j is for the band-pass signals within the same bandwidth. The higher the layer index j is, the wider the bandwidth of V_j and W_j is. The scaling function subspaces V_j have the following inclusion relation:

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots, \tag{5.9}$$

with

$$\lim_{j \to \infty} V_j = L^2(\mathcal{R}), \tag{5.10}$$

and

$$\lim_{j \to -\infty} V_j = \emptyset, \tag{5.11}$$

where $\emptyset$ denotes the empty set. In contrast, different W_j's are disjoint to each other:

$$W_{j+1} \bigcap W_j = \emptyset, \tag{5.12}$$

and

$$\bigcup_{j\in\mathcal{Z}} W_j = L^2(\mathcal{R}). \tag{5.13}$$

Furthermore, V_j and W_j are connected by the relation

$$V_j \bigcup W_j = V_{j+1}, \tag{5.14}$$

and

$$V_j \bigcap W_j = \emptyset. \tag{5.15}$$

The last two equations imply that the bandwidth of the base-band subspace V_{j+1} can be "divided evenly" in the frequency domain: V_j occupies the lower half of the bandwidth, and the upper half is left for W_j.

The well-known multi-resolution formula [8] is given by

$$V_j \bigcup \left(\bigcup_{l\geq j} W_l \right) = L^2(\mathcal{R}). \tag{5.16}$$

That is, any signal in $L^2(\mathcal{R})$ can be decomposed into a base-band component in the subspace V_j for some j, and numerous disjoint higher frequency band-pass components. In other words, the "information" of a signal is separated into the "basic" information (carried by the base-band component) and different levels of "detailed" information (corresponding to those higher frequency band-pass components). The multi-resolution formula has wide applications in signal processing and filtering techniques.

We are now in a position to briefly review some prior work concerning application of wavelets to channel estimation and signal design. Wavelets and their associated scaling functions are good candidates for pulse shaping in the signal design, because of their bandwidth efficiency and the orthogonality among their time-shift replicas.

In [51, 52, 60], use of wavelets is proposed to shape pulses in such conventional signaling schemes as PAM (pulse amplitude modulation) and QAM (quadrature amplitude modulation). Significant improvements have been reported on data transmission rates and the performance of timing recovery. The authors in [54, 55, 56] adopt wavelets to construct signature waveforms of a CDMA system and investigate related issues such as multiuser interference, low complexity detection and timing error effects.

Wavelet-based channel identification algorithms are investigated in [57, 58], where wavelets serve as basis functions in expanding the response function of a time-varying channel, because they have a good time-frequency localization property as well as providing an orthonormal basis for $L^2(\mathcal{R})$. The coefficients associated with these wavelets (i.e., the projections of a channel impulse response on these wavelets) are then estimated by applying some conventional filtering techniques based on the least mean square (LMS) or recursive least square (RLS) criterion.

Our wavelet-based approach to channel estimation is different from the previously reported schemes in the following aspects. We use sufficiently separated scaling-function-shaped pulses as a probing signal to avoid inter-symbol interference (ISI) in the received signal. Then by appropriately sampling the received signal, we immediately obtain the projections of the channel impulse response onto the subspace spanned by the scaling functions. To our best knowledge, there is no similar scheme reported in the literature.

5.3 Description of the Basic Scheme

To emphasize our principal idea, we focus on the estimation of a time-invariant channel in a noise-free situation in this section. We shall postpone discussions on how

to accommodate additive noise at the receiver end and a time-varying channel until Section 5.5.

Let $h(t)$ be a channel impulse response with bandwidth W. We assume

$$h(t) \in L^2(\mathcal{R}).$$

We further assume that the bandwidth of a chosen scaling function $\phi(t)$ is slightly larger than W, the bandwidth of $h(t)$. If not, we can rescale the scaling function by an appropriate number and use

$$\frac{1}{\sqrt{a}} \cdot \phi(at).$$

The projection of $h(t)$ onto subspace V_0 can be expressed as

$$h_0(t) = \sum_k \langle h, \phi_{0,k} \rangle \cdot \phi(t-k), \tag{5.17}$$

where

$$\langle h, \phi_{0,k} \rangle \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} h(t) \cdot \phi(t-k)\ dt. \tag{5.18}$$

Recall the scaling function $\phi(t)$ and its time shift replicas $\phi(t-k)$ constitute an orthonormal basis for V_0. In addition, V_0 roughly consists of the base-band functions within the bandwidth of $\phi(t)$, which makes it appropriate to approximate $h(t)$ by $h_0(t)$.

Consider a communication system that transmits the time-reversed version of the scaling function waveform, i.e., $\phi(-t)$, as a probing signal for channel estimation. The received signal in a noise-free situation is

$$r(t) = \int_{-\infty}^{\infty} h(\tau)\phi(\tau - t)d\tau. \tag{5.19}$$

By sampling $r(t)$ at time instants $t = 0, 1, 2, \cdots, k, \cdots$, we have

$$r(k) = \int_{-\infty}^{\infty} h(\tau)\phi(\tau - k)d\tau, \quad k = 0, 1, 2, \cdots. \tag{5.20}$$

One implicit assumption here is that the time offset between the transmitter and the receiver is known. Comparing Eqs. (5.20) and (5.18), we immediately see that a sampled received signals is exactly the projection of $h(t)$ on a basis vector of V_0, or specifically

$$r(k) = \langle h, \phi_{0,k} \rangle.$$

Therefore, the reconstruction of $h(t)$ in the subspace V_0 is simply given by

$$h_0(t) = \sum_k r(k)\phi(t-k), \tag{5.21}$$

which is equivalent to Eq. (5.17). Since $h(t)$ usually has a finite duration, the number of these coefficients $r(k)$ should be finite.

5.4 Error Analysis

Here we shall provide a general result on the error incurred by approximating $h(t)$ with subspace V_j, for $j \geq 0$, where $h_0(t)$, the approximation in V_0, in the previous section is a special case.

The projection of $h(t)$ onto subspace V_j, for $j \geq 0$, denoted by $h_j(t)$, is expressed as

$$h_j(t) = \sum_k \langle h, \phi_{j,k} \rangle \cdot \phi_{j,k}(t), \tag{5.22}$$

where

$$\langle h, \phi_{j,k} \rangle = \int_{-\infty}^{\infty} h(t) \cdot \phi_{j,k}(t)\ dt. \tag{5.23}$$

See Eq. (5.8) for the definition of V_j.

The approximation error of $h_j(t)$ is measured in terms of the square of the distance square between $h(t)$ and $h_j(t)$ in $L^2(\mathcal{R})$, i.e.,

$$\|h - h_j\|^2 \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} |h(t) - h_j(t)|^2\ dt. \tag{5.24}$$

We can show that

$$\|h - h_j\|^2 \leq \kappa D 2^{-2\alpha j}, \tag{5.25}$$

where D is the smallest integer that is larger than the channel delay spread, and κ and α are positive constants that depend on the scaling function we select. We provide a proof of Eq. (5.25) in Appendix 5.1, which is mainly based on the multi-resolution formula of Eq. (5.16).

From Eq. (5.14), we see that the subspace V_j with higher index j can include more "detailed" information of a given signal. Therefore, an approximation of $h(t)$ in subspace V_j with larger j should achieve higher accuracy, which is consistent with the quantitative conclusion of Eq. (5.25).

5.5 Two Applications

5.5.1 Estimation of a Time-invariant Channel in the Presence of Additive Noise

When additive noise is present, we should modify the sampled received signal of Eq. (5.20) as

$$\begin{aligned} y(k) &= \int_{-\infty}^{\infty} h(\tau)\phi(\tau - k)d\tau + n(k) \\ &= r(k) + n(k), \ k = 0, 1, 2, \cdots, \end{aligned} \tag{5.26}$$

where $n(k)$ is a Gaussian random variable $\mathcal{N}(0, \sigma^2)$. If the noise level σ^2 is too high to allow estimation of $h(t)$ with sufficient accuracy, we should adopt a sequence of $\phi(-t)$'s, instead of the single pulse, as a probing signal. Let L be the number of the pulses in the probing signal and I be the smallest integer such that

$$\phi(t) = 0, \text{ for any } t \notin [0, I].$$

We then design a probing signal by

$$s(t) = \sum_{l=0}^{L} (-1)^l \cdot \phi((I+D) \cdot l - t), \tag{5.27}$$

where the term $(-1)^l$ is to ensure

$$\int_{-\infty}^{\infty} s(t)dt = 0,$$

or the D.C. component of $s(t)$ is zero. Recall D is the integer corresponding to the channel delay spread. Separation of the pulses by the interval $(I+D)$ guarantees that there is no inter-symbol interference (ISI) in the received signal, which leads to the following simple channel estimation scheme.

The sampled received signal is now given by

$$\begin{aligned} y(m) &= \int_{-\infty}^{\infty} h(\tau)s(m-\tau)d\tau + n(m) \\ &= \sum_{l=0}^{L} (-1)^l \int_{-\infty}^{\infty} h(\tau)\phi(\tau - m + (I+D)l)\, d\tau + n(m), \end{aligned} \tag{5.28}$$

for $m = 0, 1, \cdots, (I+D)L$. We divide these $y(m)$'s into $(I+D)$ groups, in such a manner that the k-th group contains L noisy observations of $r(k)$ (see Eq. (5.20) for the definition of $r(k)$). The sample mean of the k-th group is then taken as an estimate of $r(k)$, i.e.,

$$\hat{r}(k) = \frac{1}{L} \sum_{l=0}^{L} (-1)^l \cdot y(k + (I+D) \cdot l), \text{ for } k = 0, 1, \cdots, I+D. \tag{5.29}$$

Compared with the variance of $n(k)$, σ^2, in Eq. (5.28), the variance of noise in $\hat{r}(k)$ is reduced to σ^2/L after the averaging operation. Finally, we can estimate the channel response $h(t)$ by

$$\hat{h}_0(t) = \sum_{k=0}^{I+D} \hat{r}(k) \cdot \phi(t-k), \tag{5.30}$$

which is similar to the noise-free case of Eq. (5.21).

Let $\hat{h}_j$ be the estimate of $h(t)$ in the presence of additive noise in subspace V_j and $\hat{r}_j(k)$ be the noise reduced estimate of

$$r_j(k) = \langle h, \phi_{j,k} \rangle.$$

By combining Eq. (5.25) and

$$\begin{aligned} E\|\hat{h}_j - h_j\|^2 &= \sum_{k=0}^{2^j D+I} E(\hat{r}_j(k) - r_j(k))^2, \\ &= (2^j D + I + 1) \cdot \frac{\sigma^2}{L}, \end{aligned} \tag{5.31}$$

we find the following upper bound for the estimation error:

$$\begin{aligned} E\|h - \hat{h}_j\|^2 &\leq \|h - h_j\|^2 + E\|\hat{h}_j - h_j\|^2 \\ &\leq \kappa D 2^{-2\alpha j} + \frac{2^j D + I + 1}{L} \cdot \sigma^2. \end{aligned} \tag{5.32}$$

We see from Eq. (5.32) that as j grows larger, the approximation error $\|h - h_j\|^2$ exponentially decreases; on the other hand, the increase in the error incurred by noise, $E\|\hat{h}_j - h_j\|^2$ is almost proportional to 2^j. The latter is because the pulse $\phi_{j,k}(t)$ becomes narrower in time for larger j and more coefficients $r_j(k)$ are needed to present $h(t)$. Equation (5.32) also implies that once j is fixed, increasing L (i.e. using a longer probing sequence) will reduce the noise incurring error, as expected.

5.5.2 Tracking of a Time-varying Channel

In order to track a time-varying channel, we only need to modify Eq. (5.29) in the previous scheme by introducing a "forgetting factor" γ $(0 < \gamma < 1)$:

$$\hat{r}(k; i) = (1 - \gamma) \cdot \sum_{l=0}^{i} \gamma^{i-l} \cdot (-1)^l \cdot y(k + l(I + D)), \tag{5.33}$$

for $k = 0, 1, 2, \cdots, (D + I)$, which is the time-vary version of $\hat{r}(k)$ in Eq. (5.29). We set $(1 - \gamma)$ as the normalization constant by noting

$$1 + \gamma + \gamma^2 + \cdots = \frac{1}{1 - \gamma}, \text{ for } 0 < \gamma < 1. \tag{5.34}$$

The factor γ controls how fast the scheme can follow a variation in the channel: for a fast varying channel, we use a smaller value of γ, and vice versa. For real-time processing, we rewrite Eq. (5.33) in the following recursive form:

$$\hat{r}(k; i + 1) = \gamma \cdot \hat{r}(k; i) + (-1)^i \cdot y(k + i(I + D)). \tag{5.35}$$

The channel reconstruction step is essentially same as Eq. (5.30):

$$\hat{h}_0(t; i) = \sum_{k=0}^{I+D} \hat{r}(k; i) \cdot \phi(t - k). \tag{5.36}$$

5.6 Simulation Results

In our simulation experiments, we estimate a complex-valued channel impulse response $h(t)$ with bandwidth W. The wavelet-based schemes and error analysis presented in the previous sections are still valid except that the sampled received signals $y(m)$'s in Eq. (5.28) have complex-values. Since the real and the imaginary parts of $h(t)$ are two independent bandlimited functions in the real domain, their convolutions with the real-valued probing signal of Eq. (5.27), thus their reconstructions, are also independent from each other. We generate the channel response $h(t)$ by lowpass filtering[1] the waveform

$$\sum_{m=1}^{M} A_m \cdot \exp\{-q_m \cdot (t - \tau_m)\}, \tag{5.37}$$

[1]The lowpass filter adopted is the FIR (finite impulse response) filter with linear-phase, designed with least-squares error minimization (see function routine FIRLS in Matlab [61] for details).

where M is the number of multipaths, A_m is the complex amplitude for the m-th path, q_m and τ_m are the corresponding exponential decay rate and time delay, respectively. The Daubechies scaling function-8, $\phi^8(t)$, is used for pulse shaping. Its bandwidth is set to be slightly larger than that of $h(t)$. The estimation of $h(t)$ with subspace V_0 is conducted in each simulation.

The normalized mean-square error (MSE), i.e.,

$$\mathcal{E} = \frac{E \parallel \hat{h}_0(t) - h(t) \parallel^2}{\parallel h(t) \parallel^2}, \tag{5.38}$$

is employed as the measure of the estimation performance. The SNR of the sampled received signal is defined by

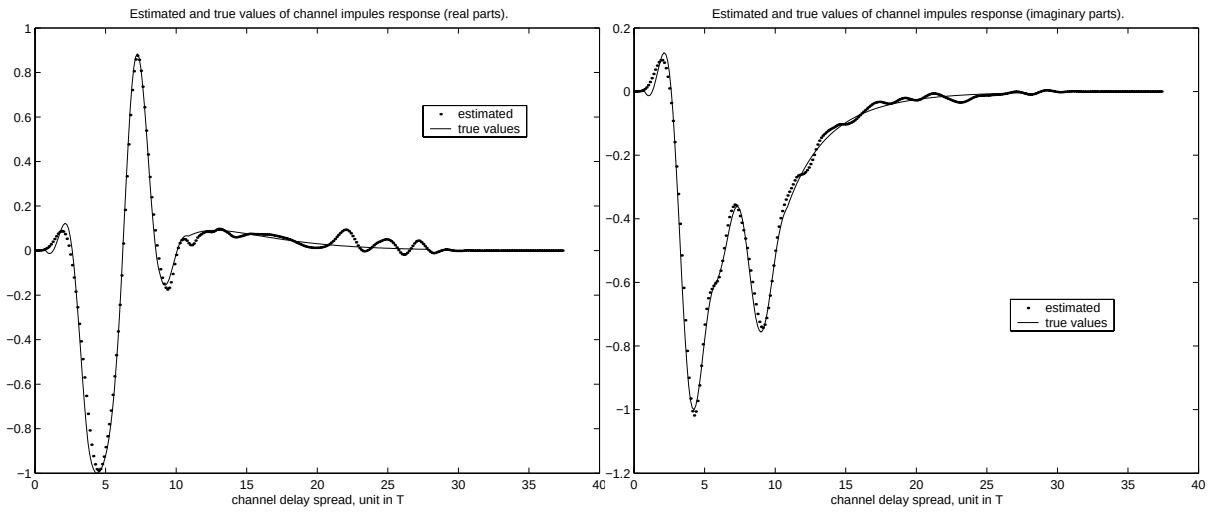
$$R = \frac{\sum_{k=1}^{I+D} |r(k)|^2/(I+D)}{\sigma^2}, \tag{5.39}$$

where $r(k)$ is the noise-free sampled signal given in Eq. (5.20), and σ^2 is the variance of the sampled noise $n(k)$ in Eq. (5.26).

Simulation 5.1.

As an illustrative example, we show in Figure 5.2 the estimation results of the real and imaginary parts of a time-invariant $h(t)$ by using the noise reduced scheme in Section 5.5.1. The estimated results are given by the dotted curves; for comparison, their true values are plotted in solid curves. Two paths are used in this simulation. The SNR of the sampled received signal R is 3dB and the number of scaling-function-shaped pulses in the probing sequence L is 100. We see that the two set of curves are fairly close, which suggests a good channel estimation.

Simulation 5.2.



(a) (b)

Figure 5.2: Estimation of (a) real and (b) imaginary parts of a time-invariant channel impulse response with the noise reduced scheme.

We now examine the performance of the noise reduced scheme with various SNR and L. Three paths are assumed with time delays

$$\{\tau_m,\ m = 1, 2, 3\} = \{1, 4, 6\}.$$

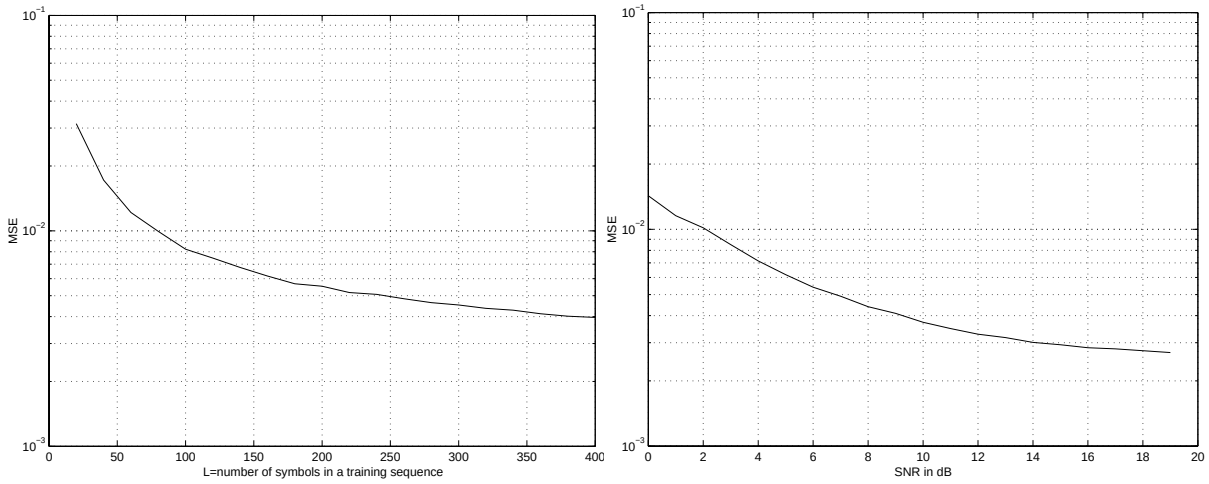
The exponential decay rate q_m is determined in terms of the delay spread for the m-th path, denoted by T_m, as

$$\exp\left(-q_m \cdot T_m\right) = 10^{-2}.$$

Here we set

$$\{T_m,\ m = 1, 2, 3\} = \{20, 20, 10\}.$$

Figure 5.3 plots (a) the MSE vs. L with fixed SNR=0dB, and (b) the MSE vs. SNR with $L = 50$. Each point is averaged over 100 simulation runs. It is verified that increasing L or SNR can improve the estimation precision as suggested by Eq. (5.32).



(a) (b)

Figure 5.3: The performance of the noise reduced scheme: (a) the MSE vs. L, the length of the training sequence, with fixed SNR=0dB; (b) the MSE vs. SNR with $L = 50$.

Simulation 5.3.

We now consider the case where a channel is time-varying. The channel model we use in this simulation is summarized in Table 5.1. The entire time period is divided into three sub-intervals, each of which contains 300 probing pulses. The channel is assumed stationary within a sub-interval. As we see, during the first interval, the channel has three paths with SNR=0dB, then it becomes a two-path channel with 3dB SNR and

	Sub-interval 1	Sub-interval 2	Sub-interval 3
i (pulse)	$1 \sim 300$	$301 \sim 600$	$601 \sim 900$
SNR (dB)	0	3	10
M	3	2	4
$\lvert A_m \rvert$	[1 .8 .6]	[1 .5]	[1 .8 .5 .3]
T_m	[20 20 10]	[30 10]	[20 30 10 10]
τ_m	[1 4 6]	[1 5]	[1 5 8 11]

Table 5.1: A summary of time-varying channel parameters used in Simulation 5.3.

four-path channel with 10dB SNR in the second and third sub-intervals, respectively. The forgetting factor $\gamma = 0.93$ is applied. Figure 5.4 shows the performance of the scheme for tracking a time-varying channel described in Section 5.5.2. One hundred simulation runs are averaged to produce each point of the curve. We see the method can efficiently trace the time-varying channel impulse response.

From the simulation experiment, we notice that the value of γ needs to be carefully chosen. Generally speaking, the smaller the γ is, the faster the algorithm converges. However, if γ is too small, e.g. less than 0.90, the algorithm will be unstable in the sense that the MSE curve will oscillate even after convergence. This is similar to the time-invariant case with a small L.

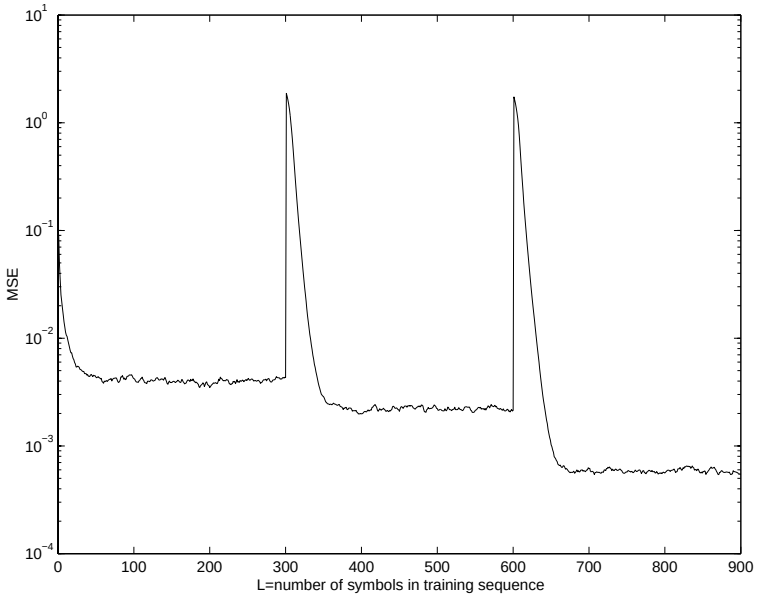


Figure 5.4: The performance of the channel tracking scheme in Section 5.5.2.

5.7 Concluding Remarks

In this chapter, we developed a simple wavelet-based approach to estimate a band-limited channel impulse response. This approach takes advantage of two properties of the scaling function associated with an orthornormal wavelet: the bandwidth efficiency and the orthogonality between the scaling function and its time-shift replica. The applications to time-invariant and time-varying channels were investigated separately. Computer simulations confirmed a good performance.

Appendix 5.1: Derivation of Eq. (5.25)

The proof can be done in the following four steps:

Step 1.

By using the multi-resolution formula of Eq. (5.16), we can express $h(t)$ in terms of the projections of $h(t)$ onto one scaling-function-spanned subspace V_j and multiple wavelet-spanned subspaces W_l with $l \geq j$:

$$h(t) = \sum_k \langle h, \phi_{j,k} \rangle \phi_{j,k}(t) + \sum_{l \geq j} \sum_k \langle h, \psi_{l,k} \rangle \psi_{l,k}(t). \tag{5.40}$$

Since

$$h_j(t) = \sum_k \langle h, \phi_{j,k} \rangle \phi_{j,k}(t),$$

it is straightforward to obtain

$$\|h - h_j\|^2 = \sum_{l \geq j} \sum_k |\langle h, \psi_{l,k} \rangle|^2. \tag{5.41}$$

Step 2.

In order to estimate the coefficients $\langle h, \psi_{j,k} \rangle$ in the above expression, we first consider the following approximation for

$$\langle h, \psi_{0,0} \rangle = \int_{-\infty}^{\infty} h(t)\psi(t)\ dt.$$

Since the spectrum of $h(t)$ is confined in the interval $[-W, W]$ in the frequency domain, we approximate $\Psi(f)$, the Fourier transform of the wavelet $\psi(t)$, within $[-W, W]$ with an easy-integrated function $\kappa_0 f^\alpha$ such that

$$\kappa_0 f^\alpha \geq |\Psi(f)|, \text{ for } f \in [-W, W]. \tag{5.42}$$

where κ_0 and α are some suitable positive numbers that can be determined by simulation trials.

Step 3.

Using the absolute value of the Fourier transform of $\psi_{l,k}(t) = 2^{l/2}\psi(2^l t - k)$, i.e.,

$$|\Psi_{l,k}(f)| = 2^{-l/2}\left|\Psi(2^{-l}f)\right|,$$

we derive the upper bound of the inner product $\langle h, \psi_{l,k}\rangle$ in Eq. (5.41) as

$$\begin{aligned}
|\langle h, \psi_{l,k}\rangle| &= |\langle H, \Psi_{l,k}\rangle| \\
&\leq H_{max}\int_{-W}^{W} 2^{-l/2}\left|\Psi(2^{-l}f)\right| \, df \\
&\leq 2^{1-\frac{l}{2}}\kappa_0 H_{max}\int_0^W (2^{-l}f)^\alpha \, df \\
&= \frac{2\kappa_0 H_{max} W^{\alpha+1}}{\alpha+1}\cdot 2^{-(\frac{1}{2}+\alpha)l} \\
&= \kappa_1 2^{-(\frac{1}{2}+\alpha)l},
\end{aligned} \tag{5.43}$$

where

$$H_{max} = \max_f |H(f)|,$$

and

$$\kappa_1 = \frac{2\kappa_0 H_{max} W^{\alpha+1}}{\alpha+1}.$$

Note the upper bound of Eq. (5.43) is independent of the time-shift index k of $\psi_{l,k}(t)$.

Step 4.

Since the channel delay spread is less than D, the maximum number of non-zero coefficients $\langle h, \psi_{l,k} \rangle$ at level l is $2^l D$. Substituting Eq. (5.43) into Eq. (5.41), we have

$$\begin{aligned} \|h - h_j\|^2 &\leq \sum_{l \geq j} 2^l D \cdot \left| \kappa_1 2^{-(\frac{1}{2}+\alpha)l} \right|^2 \\ &= \kappa_1^2 D \sum_{l \geq j} 2^{-2\alpha l} \\ &= \kappa_1^2 D \cdot \frac{2^{-2\alpha j}}{1 - 2^{-2\alpha}} \\ &= \kappa D 2^{-2\alpha j}, \end{aligned} \tag{5.44}$$

where

$$\kappa = \frac{\kappa_1^2}{1 - 2^{-2\alpha}}.$$

This concludes our proof.

Chapter 6

Conclusions

In this dissertation, we have presented four topics on wireless geolocation. The contributions of this dissertation involve analysis of major NLOS geolocation approaches, development of a geolocation algorithm for the mitigation of NLOS effects and a wavelet-based approach to channel estimation. We summarize the contributions for each of the four focused areas as follows:

- **Chapter 2:** In the unified analysis of the NLOS geolocation, major estimation approaches such as the time of arrival (TOA), time difference of arrival (TDOA) and signal strength based (SS) schemes are investigated in a coherent framework. The theoretically achievable geolocation accuracy is derived in terms of the Cramer-Rao Lower Bound (CRLB) or its generalized version depending on whether prior information of NLOS delays is available. It is then shown that the maximum likelihood (ML) and maximum *a posteriori* probability (MAP) estimators based on measurements pertaining to the position information can asymptotically achieve these lower bounds. In addition, the connection between theoretically optimum receivers and the conventional geometry-based methods is clarified.

- **Chapter 3:** We have explored the relationship among the three distance-based positioning schemes, i.e., the TOA, TDOA and SS based methods. We provided an analytical explanation for the claim that given a set of BS locations and an MS position, the TOA method should achieve a higher positioning precision than its TDOA counterpart. However, the two positioning methods may attain the same level of accuracy under certain conditions. We investigated the tradeoff between the accuracy limits of the TOA and SS based methods, based on which we devised a hybrid distance estimation scheme that combines TOA and SS data.

- **Chapter 4:** Combining an optimum receiver and the sequential simplex method (SSM), we have developed a new geolocation algorithm. This algorithm incorporated the aforementioned hybrid distance estimation scheme and the geographic constraint, such as street and highway layout, as boundary conditions for a mobile's likely position. We adopted SSM as an optimization tool, because it can handle a non-differentiable objective function and irregular boundaries in a simple manner.

- **Chapter 5:** We have proposed a wavelet-based approach to channel estimation. This approach takes advantage of two properties of the scaling function associated with an orthonormal wavelet: its bandwidth efficiency and the orthogonality between the scaling function and its time-shift replica. A training sequence is constructed using the scaling function. By appropriately sampling the received signal, we are able to obtain the projection of the channel impulse response onto the subspace spanned by the scaling functions, which then allows a simple yet accurate reconstruction of the channel response.

Some directions of our future work are in order:

- The geolocation algorithm developed in Chapter 4 is suitable for a mobile tracking scenario, or a "refining" estimation, where the start point of an optimization search is not very far from the optimum point. For a general "cold start" case, it is desirable to extend this algorithm to a two-step estimation scheme, i.e., an initial estimation using some less accurate but fast-convergent optimization technique, followed by the proposed "refining" estimation with the SSM.

  Another useful modification of this algorithm is to incorporate some scheme that adapts parameters depending on whether the MS of being tracked represents a pedestrian, a vehicle on a urban street or on a highway, etc. For example, when the MS is a pedestrian, local streets and shopping malls may provide an appropriate guidance to be considered in inferring a likely position of the MS, and geolocation update need not be done as frequently as that for a highway vehicle.

- Current techniques usually adopt special "probing" signals for the geolocation purpose. It will be of practical and theoretical interest if geolocation can be performed in conjunction with a data extraction procedure. We plan to extend our current results to the setting of joint estimation of a mobile's position and information sequences. The EM (Expectation and Maximization) algorithm is expected to be employed.

- Reconsider the geolocation problem from a network perspective. First, a network may play an active role in providing flexible and efficient geolocation solutions. For example, when poor channel quality and an unfavorable MS and BS configuration prohibit the positioning accuracy from attaining a desired level, a

network may enhance the performance by implementing an extra BS at a specific "trouble" spot, or requesting assistance from an overlay positioning system such as GPS. Second, it is conceivable that several types of networks, which may adopt different geolocation methods depending on their terrestrial conditions, need to work jointly to offer ubiquitous geolocation services in a near future. How to coordinate these systems and handle "handoffs" in a boundary region of two networks also prompts some interesting problems. Third, a network is desired to provide different levels of geolocation accuracy depending on characteristics of its services. Thus, some adaptive geolocation algorithms should be investigated.

- Field testing is crucial to both pre-deployment selection of a geolocation scheme and post-deployment demonstration of its compliance with certain accuracy requirement [62]. However, what is lacking is a fair and standardized test procedure. An appropriate answer may be found if we have a better understanding of interactions among channel conditions, network parameters and geolocation schemes. Our recent study of unified analysis of NLOS geolocation may serve as a guideline to provide predictions to be compared with field test results. More works are called for in this area.

Bibliography

[1] J. Reed, K. Krizman, B. Woerner and T. Rappaport, "An overview of the challenges and progress in meeting the E-911 requirement for location services," *IEEE Communications Magazine*, vol. 36, pp. 30–37, April 1998.

[2] J. Caffery, *Wireless location in CDMA cellular radio system*, Kluwer Academic Publisher, 1999.

[3] B. Hofmann-Wellenhof, H. Lichtenegger and J. Collins, *Global Positioning System: theory and practice*, Springer-Verlag, New York, 4th edition, 1997.

[4] C. Drane, M. Macnaughtan and C. Scott, "Positioning GSM Telephones," *IEEE Communications Magazine*, vol. 36, pp. 46-59, April 1998.

[5] H. L. Van Trees, *Detection, Estimation and Modulation Theory, Part I*, John Wiley & Sons, Inc., 1968.

[6] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer, 2nd edition, 1998.

[7] G. Beveridge and R. Schechter, *Optimization: Theory and Practice*, McGraw-Hill Book Co., 1970.

[8] I. Daubechies, *Ten Lectures on Wavelets,* CBMS Lecture Notes, No. 61 SIAM, Philadelphia, 1992.

[9] C. E. Cook and M. Bernfeld, *Radar Signals: an introduction to theory and applications*, Academic Press, 1970.

[10] C. R. Rao, *Linear Statistical Inference and Its Applications*, John Wiley & Sons, Inc., 1965.

[11] H. Koorapaty, H. Grubeck and M. Cedervall, "Effect of biased measurement errors on accuracy of position location methods," *Proc. IEEE Global Telecommunications Conference* (GLOBECOMM 1998), Sydney, Australia, November 8-12, 1998, vol. 3, pp. 1497–1502.

[12] C. Botteron, M. Fattouche and A. Hϕst-Madsen, "Statistical theory of the effects of radio location system design parameters on the positioning performance," *Proc. IEEE Vehicle Technology Conference* (VTC 2002 Fall), Vancouver, Canada, September 2002.

[13] D. Torrieri, "Statistical theory of passive location systems," *IEEE Trans. on Aerospace and Electronic Systems*, AES-20(2): pp. 183–198, 1984.

[14] M. Wylie and J. Holtzman, "The non-line of sight problem in mobile location estimation," *Proc. IEEE International Conference on Universal Personal Communications* (IEEE ICUPC 1996), Cambridge, MA, September 29-October 2, 1996, pp. 827–831.

[15] L. Xiong, "A selective model to suppress NLOS signals in angle-of-arrival (AOA) location estimation," *IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications 1998*, Boston, MA, 1998, pp. 461–465.

[16] J. Borras, P. Hatrack and N. B. Mandayam, "Decision theoretic framework for NLOS identification," *Proc. IEEE Vehicular Technology Conference*, Ottawa, Canada, Spring 1998, vol. 2, pp. 1583–1587.

[17] S. Venkatraman and J. Caffery, "A statistical approach to non-line-of-sight BS identification," *Proc. the 25th International Symposium on Wireless Personal Multimedia Communications* (IEEE WPMC 2002), Honolulu, Hawaii, USA, October 27-30, 2002.

[18] S. Gezici, H. Kobayashi and H. V. Poor, "Non-parametric non-line-of-sight identification," to appear in *Proc. IEEE Semiannual Vehicle Technology Conference* (VTC-2003 Fall), Orlando, Florida, USA, October 6-9, 2003.

[19] C. K. Knapp and G. C. Carter,"The generalized correlation method for estimation of time delay," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-24, no. 4, August 1976, pp. 320–327.

[20] G. C. Carter,"Coherence and time delay estimation," *Proc. IEEE*, vol. 75, February 1987, pp. 236–255.

[21] A. Jakobsson, A. L. Swindlehurst and P. Stoica, "Subspace-base estimation of time delays and doppler shifts," *IEEE Trans. Signal Processing*, vol. 46, no. 9, September 1998, pp. 2472–2483.

[22] A. L. Swindlehurst, "Time delay and spatial signature estimation using known asynchronous signals," *IEEE Trans. Signal Processing*, vol. 46, no. 2, February 1998, pp. 449–462.

[23] P. Chen, "A cellular based mobile location tracking system," *Proc. IEEE Vehicle Technology Conference* (VTC 1999), 1999, pp. 1979–1983.

[24] H. Kim, W. Yoon, D. Kim and Y. Kim, "Mobile positioning using improved least squares algorithm in cellular systems," *IEICE Trans. Communications*, vol. E84-B, no. 1, January 2001, pp. 183–190.

[25] Y. Qi and H. Kobayashi, "Mitigation of non-line-of-sight effects in time-of-arrival positioning," *Proc. 35th Annual Conference on Information Sciences and Systems* (CISS 2001), The Johns Hopkins University, March 2001, pp. 590–592.

[26] Y. Qi and H. Kobayashi, "A unified analysis for Cramer-Rao Lower Bound for geolocation," *Proc. 36th Annual Conference on Information Sciences and Systems* (CISS 2002), Princeton University, March 2002.

[27] Y. Qi and H. Kobayashi, "Cramer-Rao Lower Bound for geolocation in a non-line-of-sight environment," *Proc. 2002 IEEE International Conference on Acoustic Speech and Signal Processing* (ICASSP 2002), Orlando, Florida, May 2002, pp. 2473–2476.

[28] Y. Qi and H. Kobayashi, "A wavelet-based approach to channel estimation," *Proc. IASTED International Conference on Wireless and Optical Communications* (WOC 2002), Banff, Canada, July 2002, pp. 255–260.

[29] Y. Qi and H. Kobayashi, "On achievable bounds for non-line-of-sight geolocation," *Proc. IASTED International Conference on Wireless and Optical Communications* (WOC 2002), Banff, Canada, July 2002, pp. 475–479.

[30] Y. Qi and H. Kobayashi, "On geolocation accuracy with prior information in a non-line-of-sight environment," *Proc. IEEE Semiannual Vehicle Technology Conference* (VTC 2002 Fall), Vancouver, Canada, September 2002.

[31] Y. Qi and H. Kobayashi, "On time delay distribution of multipaths in mobile environment," *Proc. IEEE International Symposium on Advances in Wireless Communications* (ISWC'02), Victoria, Canada, September 23-24, 2002.

[32] Y. Qi and H. Kobayashi, "An optimum positioning receiver for non-synchronized mobile systems," *Proc. 37th Annual Conference on Information Sciences and Systems* (CISS 2003), The Johns Hopkins University, March 2003.

[33] K. Hoffman and R. Kunze, *Linear Algebra*, 2nd edition, Prentice-Hall, Englewood Cliffs, N.J., 1971.

[34] W. C. Jakes, *Microwave Mobile Communications*, IEEE Press, 1994.

[35] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*, 2nd edition, Clarendon Press, 1992.

[36] G. H. Golub and C. F. Van Loan, *Matrix Computation*, 3rd edition, The Johns Hopkins University Press, 1996.

[37] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. II, John Wiley & Sons, Inc., 1966.

[38] A. J. Coulson, A. G. Williamson and R. G. Vaughan, "A statistical basis for lognormal fading effects in multipath fading channel," *IEEE Trans. on Communications*, vol. 46, no. 4, April, 1998, pp. 494–502.

[39] C. D. Charalambous and N. Menemenlis, " A state space approach in modeling multipath fading channels: stochastic differential equations and Ornstein-Uhlenbeck processes," *Proc. IEEE International Conference on Communications* (IEEE ICC 2001), June 11-15, 2001, Helsinki, Finland.

[40] Y. Chen and H. Kobayashi, "Signal strength based indoor geolocation," *Proc. IEEE International Conference on Communications* (IEEE ICC 2002), April 28-May 2, 2002, New York City, NY.

[41] R. B. Ertel and J. H. Reed, "Angle and time of arrival statistics for circular and elliptical scattering models," *IEEE Journal on Selected Areas in Communications*, November 1999, vol. 17, no. 11, pp. 1829–1840.

[42] R. Janaswamy, "Angle and time of arrival statistics for the Gaussian scatter density model," *IEEE Trans. Wireless Communications*, July 2002, vol. 1, no. 3, pp. 488–497.

[43] J. K. Tugnait, L. Tong and Z. Ding, "Single-user channel estimation and equalization," *IEEE Signal Processing Magazine*, May 2000, vol. 17, no. 3, pp. 17–28.

[44] J. G. Proakis, *Digital communications,* New York, McGraw Hill 1989.

[45] L. Tong and S. Perreau, "Multichannel blind identification: from subspace to maximum likelihood methods," *Proc. IEEE*, October 1998, pp. 1951–1966.

[46] G. B. Giannakis and C. Tepedelenlioğlu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proc. IEEE*, October 1998, pp. 1969–1986.

[47] P. Chen, "Signal detection and channel estimation in multipath channels," PhD dissertaion, August 2001, Princeton University, NJ.

[48] T. Kailath, "Correlation Detection of Signals Perturbed by a Random Channel," *IEEE Trans. Information Theory*, June 1960, vol. IT-6, pp. 361–366.

[49] G. K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Trans. Communications*, July 1994, vol. 42, no. 7, pp. 2406–2413.

[50] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, 1988, vol. 41, pp. 909–996.

[51] J. N. Livingston and C. Tung, "Bandwidth efficient PAM signaling using wavelets," *IEEE Trans. Communications*, December 1996, vol. 44, no. 12, pp. 1629–1631.

[52] F. Danechgaran and M. Mondin, "Wavelet-based signal design for reduced jitter timing recovery," *IEEE Trans. Communications*, December 1997, vol. 45, no. 12, pp. 1523–1526.

[53] A. R. Lindsey, "Wavelet packet modulation for orthogonally multiplexed communications," *IEEE Trans. Signal Processing*, May 1997, vol. 45, no. 5, pp. 1336–1339.

[54] R. E. Learned, A. S. Willsky and D. M. Boroson, "Low complexity optimal joint detection for oversaturated multiple access communications," *IEEE Trans. Signal Processing*, January 1997, vol. 45, no. 1, pp. 113–123.

[55] D. M. Ionescu and M. A. Wickert, "On the performance of a CDMA system with user signatures based on packet wavelets in multipath channels," *Proc. IEEE Vehicular Technology Conference*, 1997, vol. 1, pp. 392–396.

[56] K. M. Wong, J. Wu, T. N. Davidson and Q. Jin, "Wavelet packet division multiplexing and wavelet packet design under timing error effects," *IEEE Trans. Signal Processing*, December 1997, vol. 45, no. 12, pp. 2877–2890.

[57] M. I. Doroslovacki and H. Fan, "Wavelet-based linear system modeling and adaptive filtering," *IEEE Trans. Signal Processing*, May 199, vol. 44, no. 5, pp. 1156–1167.

[58] H. Zhang, H. H. Fan and A. Linsey, "A wavelet packet based model for time-varying wireless communication channels," *Proc. of Third IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, Taiwan, China, March 2001, pp. 50–53.

[59] A. R. Lindsey, "Wavelet packet modulation for orthogonally multiplexed communications," *IEEE Trans. Signal Processing*, May 1997, vol. 45, no. 5, pp. 1336–1339.

[60] A. R. Lindsey, "Wavelet packet modulation for orthogonally multiplexed communications," *IEEE Trans. Signal Processing*, May 1997, vol. 45, no. 5, pp. 1336–1339.

[61] Matlab 6.0, Copyright 1984-2000, the MathWorks, inc.

[62] D. N. Hatfield, "A report on technical and operational issues impacting the provision of wireless enhanced 911 services," prepared for Federal Communications Commission, October 2002. (http://www.fcc.gov/911/enhanced/reports/)

Glossary of Principal Symbols

Symbol	Definition	Page or Equation Number
A_b	signal amplitude	(2.2)
BS	base station	1
$\mathcal{B}$	set of all BSs	9
subscript b	b-th base station	(2.2)
CRLB	Cramer-Rao Lower Bound	(2.5)
c	the speed of light	(2.1)
d	distance	(2.104)
$\hat{d}$	distance estimate	(3.21)
d_c	critical distance	(3.18)
E	expectation	(2.4)
$\mathcal{E}$	mean square error of channel estimation	(5.38)
e	instant SS estimate	(2.110)
$\overline{e}$	time averaged SS estimate	(2.111)
FIM	Fisher information matrix	(2.4)
$f_{\boldsymbol{\theta}}(\boldsymbol{r})$	probability density function conditioned on $\boldsymbol{\theta}$	(2.3)
G-CRLB	generalized CRLB	(2.41)
$\mathbf{H}$	matrix containing geometric relation among the MS and BSs for the TOA method	(2.8)
$\mathbf{H}_L$	submatrix related to LOS BSs in $\mathbf{H}$	(2.10)
$\mathbf{H}_{NL}$	submatrix related to NLOS BSs in $\mathbf{H}$	(2.10)

Symbol	Definition	Page or Equation Number
$\mathbf{H}_{SS}$	matrix $\mathbf{H}$ for the SS method	56
$\mathbf{H}_{TDOA}$	matrix $\mathbf{H}$ for the TDOA method	45
$h(t)$	channel impulse response	102
$\mathbf{J}$	information matrix	(2.42)
$\mathbf{J}_D$	information matrix due to observation	(2.43)
$\mathbf{J}_P$	information matrix due to prior knowledge	(2.44)
$\mathbf{J}_{\boldsymbol{\theta}}$	Fisher information matrix for estimation of $\boldsymbol{\theta}$	(2.4)
$\mathbf{J}_{\boldsymbol{\tau}}$	Fisher information matrix for estimation of $\boldsymbol{\tau}$	(2.9)
LOS	line of sight	2
subscript L	LOS base stations	(2.10)
$\mathcal{L}$	set of LOS BSs	9
l_b	NLOS induced path length	(2.1)
$\boldsymbol{l}$	NLOS induced path vector	9
MAP	maximum *a posteriori* probability	38
MLE	maximum likelihood estimation	22
MS	mobile station	1
NLOS	non-line of sight	1
subscript NL	NLOS base stations	(2.10)
$N_0/2$	spectrum density of white noise process	10
$\mathcal{NL}$	set of NLOS BSs	9
$n_b(t)$	noise in received signal $r_b(t)$	(2.2)
$\mathcal{P}_{CR}$	positioning accuracy of the CRLB	(2.30)
$\mathcal{P}_{G-CR}$	positioning accuracy of the G-CRLB	(2.71)

Symbol	Definition	Page or Equation Number
$\boldsymbol{p} = (x, y)^T$	MS position	9
R,R_b	SNR	(2.14)
$r_b(t)$	received signal	(2.2)
$\boldsymbol{r}$	vector of received signals	(2.3)
SNR	signal to noise ratio	2
SS	signal strength	3
SSM	sequential simplex method	4
$S(f)$	Fourier transform of $s(t)$	13
$\mathcal{S}$	set of BSs providing SS data	80
$s(t)$	signal waveform	(2.2)
TDOA	time difference of arrival	1
TOA	time of arrival	1
$\mathcal{T}$	set of BSs providing TOA data	80
V_j	j-th scaling function subspace	(5.8)
W	chip rate of a CDMA signal, bandwidth	18,102
W_j	j-th wavelet subspace	(5.7)
W_{rd}	road width	90
w	log-normal fading factor in dB	(2.110)
z,z_b	signal strength	(2.110)
β	effective bandwidth	13
ϵ	path loss factor	50
η_b^2,σ_b^2, ω_b^2	variances of Gaussian variables	55,33,46
$\boldsymbol{\Lambda}_L$	submatrix related to LOS BSs in $\mathbf{J}_{\boldsymbol{\tau}}$	(2.13)
$\boldsymbol{\Lambda}_{NL}$	submatrix related to NLOS BSs in $\mathbf{J}_{\boldsymbol{\tau}}$	(2.12)
λ_b	diagonal term of $\mathbf{J}_{\boldsymbol{\tau}}$	(2.14)

Symbol	Definition	Page or Equation Number
ϕ_b	geometric angle determined by the positions of an MS and BS$_b$	12
$\phi(t)$	scaling function	(5.1)
$\phi^n(t)$	Daubechies scaling function-n	97
$\mathbf{\Pi}$	covariance matrix of $\boldsymbol{l}$ with Gamma distribution	(2.70)
$\mathbf{\Psi}$	covariance matrix of TDOA data	(2.96)
$\Psi(f)$	Fourier transform of $\psi(t)$	113
$\psi(t)$	wavelet function	(5.2)
$\psi^n(t)$	Daubechies wavelet function-n	97
$\mathbf{\Omega}$	covariance matrix of $\boldsymbol{l}$ with Gaussian distribution	(2.46)
τ,τ_b	time delay	(2.1)
$\boldsymbol{\tau}$	vector of time delays	12
$\boldsymbol{\theta}$	$\boldsymbol{\theta} = (\boldsymbol{p}^T, \boldsymbol{l}^T)^T$	9
ξ,ξ_b	error in time delay estimate	(2.32)
ζ,ζ_b	error in distance estimate	(3.21)
$\bigcap$	intersection	(5.12)
$\bigcup$	union	(5.13)
$\emptyset$	a empty set	(5.11)
$\mathbf{1}$	a column vector of all 1's	(2.90)