

Copyright © 2013 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Review of Range-Based Positioning Algorithms

Junlin Yan^{*}, Christian C. J. M. Tiberius[†], Peter J. G. Teunissen^{†‡},
Giovanni Bellusci[§] and Gerard J. M. Janssen^{††}

Abstract

This article reviews algorithms which turn measured ranges into position solutions. From their basic mathematical principles, we relate and compare relevant aspects of these algorithms. Special attention is given to the direct (non-iterative) algorithms, which are frequently applied in indoor positioning. Most of them are shown to be essentially the same, as they can be related through applying different weighting schemes. This article is intended as a useful guide to help researchers and system designers evaluate and select appropriate range-based positioning algorithms for their applications at hand.

Index Terms

Least-squares, positioning algorithm, direct methods.

I. INTRODUCTION

Positioning and navigation have been of long-standing interest. Besides the value in the transport and military context, positioning services find their applications in a variety of areas, such as industry (tracking of objects in ware-houses), science (geodesy), sports (hiking), health care (monitoring of patients), logistics (package tracking) security applications (localizing authorized persons in high-security areas),

^{*}Junlin Yan was with Delft University of Technology and is now with Intel Mobile Communications GmbH, Dusseldorfer Landstr. 401, D-47259 Duisburg, Germany, junlin.yan@intel.com, +49 203 7298 511.

[†]Christian C. J. M. Tiberius and Peter J. G. Teunissen are with the Fac. of AE, Delft University of Technology, Kluyverweg 1, 2629 HS Delft, The Netherlands, C.C.J.M.Tiberius@tudelft.nl, +31 (0)15 27 82573, P.J.G.Teunissen@tudelft.nl.

[‡]Peter J. G. Teunissen is, meanwhile, leader of the GNSS Research Lab of Curtin University of Technology, Building 207, Spatial Sciences, Room 204, Bentley Campus, Perth, Australia, P.Teunissen@curtin.edu.au, (+618) 9266 7676

[§]Giovanni Bellusci was with Delft University of Technology and is now with Xsens in Enschede, The Netherlands, Giovanni.Bellusci@xsens.com.

^{††} Gerard J. M. Janssen is with the Fac. of EEMCS, Delft University of Technology, Mekelweg 4, 2628CD Delft, The Netherlands, G.J.M.Janssen@tudelft.nl, +31 (0)15 27 86736.

family supervision of children, and search and rescue (locating fire fighters, or avalanche/earthquake victims) [1]. These applications are being developed at an astonishing rate. Given the societal interest in the topic of positioning and the fact that there are a lot of related publications scattered widely across many disciplines, a structured and detailed review of existing positioning algorithms is of interest to the research community. This paper is different from earlier tutorial type of articles (outlined at the end of this section) since it focuses on the very basic principles of the algorithms and intends to cast similar types of methods in a uniform structure. The outcome of this work is applicable to various positioning applications and systems, no matter whether the ranging signal is optical, radio or acoustic.

The majority of positioning systems estimate the unknown position based on range measurements between transmitters and a receiver (or vice versa), which are obtained by extracting the information contained in the received signal, such as Received Signal Strength (RSS), Time of Arrival (ToA) and Time Difference of Arrival (TDoA). In TDoA systems, only *pseudorange*s can be obtained which contain a common offset (or bias) b_c , caused by the clock misalignment between the user receiver clock and the synchronized transmitter clocks. The offset b_c needs to be treated as an unknown parameter, in addition to the unknown user position coordinates. It can be either estimated using the pseudorange, or eliminated by taking differences among the pseudorange (hence TDoA). For fully synchronized (ToA) or non-timing based systems (like RSS), this type of offset does not exist.

Numerical estimation of the unknown parameters of interest based on range measurements is usually done by maximizing or minimizing an objective function. Typical objective functions are sums of squared residuals, likelihood functions, posterior density functions, risk functions, and robust loss functions. The two most common criteria are least-squares (LS) and maximum likelihood (ML). Applying the inverse of the measurement covariance matrix as the weight matrix, the LS estimator and the ML estimator are equivalent, and its variance achieves the Cramer-Rao Lower Bound (CRLB), when the ranging errors are distributed with an *elliptically* contoured joint probability density function (PDF). For Line-of-Sight (LoS) ranging, such an assumption is realistic, since the ranging errors can be modelled as Normally distributed [2]–[4]. Furthermore, compared to ML estimation, un-weighted LS estimation does not require any knowledge of the measurement PDF, and only first and second order moments of the measurement errors may be required in the weighted case. In this paper, we will focus on algorithms based on least-squares.

The unknown Cartesian position coordinates are involved in the range equations in a non-linear way. The coordinates are usually estimated using Iterative Descent (ID) techniques, such as the Steepest-Descent method, the Newton method, the Trust-Region method, the Gauss-Newton method and the

Levenberg Marquardt method. A detailed analysis of these methods can be found in [5]. The ID methods solve the non-linear positioning problem *iteratively* by means of iterative linear approximation, i.e. they start with an initial estimate x_0 of the unknown vector x , which is then updated, from iteration to iteration, based on the rule defined by the specific method. The final estimate \hat{x} results upon convergence. Traditional ID methods, when successfully converged, provide estimators that closely approximate strict least-squares solutions. One drawback is that they require an initial estimate to start. Additionally, the heavy computations involved in the iterations can be problematic for low cost and low power indoor applications.

In general, strict direct (non-iterative) least-squares solutions to non-linear problems do not exist. However, with some assumptions or simplifications, direct least-squares positioning algorithms can be developed, which are extensively documented in the literature, for both TDoA [6]–[18] and ToA/RSS [19]–[24] systems. All these methods apply a squaring operation on the used range equations to cancel the square root. After squaring, both linear and non-linear terms (functions) of the unknown parameters are obtained, and the non-linear term (function), denoted as χ , is treated as an unknown parameter in addition to the unknown vector x . As we will discuss later, these methods can be clearly classified based on the objective functions, and it can be shown that many different variants of the same method do exist, with the differences arising only through the use of different measurement weighting schemes. This fact has not been documented in the literature so far. The direct methods provide simple-to-compute estimates, but they are not strict LS solutions to the used equations, since the relation between χ and x is not rigorously exploited.

There also exists a third group of methods that solve the positioning problem iteratively but with only 1-dimensional calculations in each iteration. In [25], the estimation accuracy is improved over the direct methods, by introducing a Lagrange Multiplier in the objective function to fully account for the constraint between the non-linear term and the original unknowns after squaring. Applying eigenvalue decomposition, the final estimate is obtained by iteratively searching the roots of a high order polynomial with a single unknown. This method was further investigated in [26]. Another method [27], [28] first transforms the multi-dimensional positioning problem to 1-dimension, based on direct methods, then finds the estimate for the remaining single unknown iteratively using the original un-squared range equations. These methods perform better than the direct methods in terms of Root Mean Squared Error (RMSE), but they are again iterative and require an initial estimate to start.

Several tutorial type of articles exist in the literature [1], [29]–[38], regarding the topic of positioning techniques. These papers cover a wide range of aspects of the topic on a general level, and devote

most of the attention to the signal processing techniques at the stage where RSS, ToA and TDoA are estimated. Some of the papers do discuss the positioning algorithms, but the level is quite general and a structural comparison is missing, which compose the central subject of our paper. The purpose of this work is to give a *structured* and *detailed* review of the range-based least-squares *positioning methods* mentioned above, and give special attention to the *direct methods*. The basic principles, assumptions and approximations of these methods are discussed, their *relations* are pointed out, and the *properties* of different types of methods are summarized. Positioning results obtained with actual Ultra-Wideband (UWB) range measurements are used to illustrate and support the review. With the underlying principles of the methods revealed, and thereby their inherent strong and weak points, the paper can be used to assist researchers and developers in the field to find the right choice for their applications.

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

The paper is organized as follows. Section II introduces the positioning models and Section III gives an overview of the principle of Non-Linear Least Squares. Then, range-based positioning methods are reviewed in Section IV (ID methods), Section V (direct methods) and Section VI (1-dimensional iterative methods). The review of the ID methods is treated first, because these methods had been widely used for positioning long before the other two groups of methods were introduced. The comparison regarding the computational costs of the methods are given in Section VII. To support the review, the validation setup and the obtained results are shown in Section VIII. At the end of the paper, a summary is given in Section IX.

II. MEASUREMENT AND POSITIONING MODELS

In this section, we briefly describe the models that are generally used to characterize RSS, ToA and TDoA measurements. In the end, the positioning model is given. For any vector a , we will use the notation $\underline{\hat{a}}$ to represent its *estimator*, and the notation \hat{a} for its estimate. The estimator is a random vector (denoted with the underscore) while the estimate is a deterministic vector, which is a (numerical) realization of the estimator.

A. RSS

RSS can be modeled as a linearly decreasing variable on a logarithmic scale of the distance [39], [40]. The RSS from the i -th transmitter can be formulated as

$$\hat{P}_i(d_i) = P_0 - 10n_p \log_{10} \frac{d_i}{d_0} + \underline{S}_i \quad (1)$$

where $\hat{P}_i(d_i)$ is the estimator of the received power (in dB) at distance d_i , d_0 is the reference distance, P_0 is the known power (in dB) received at the reference distance, n_p is the pathloss exponent, and \underline{S}_i is a zero mean Gaussian random variable in dB. One should note that the pathloss exponent n_p can either be obtained by conducting a measurement campaign at the initialization stage, or be treated as an additional unknown parameter. In this paper, we assume that n_p is known from a campaign. Moreover, in the case where the user is sending the ranging signal to several stations, the transmitted signals strength can be unknown and it has to be considered as an additional parameter to estimate in the positioning algorithms. Then the situation is similar as in a TDoA system, which will be addressed in Section II-C. In this paper, we assume the transmitted signal strength is known.

With the estimated received power, model (1) is used to estimate the range between the i -th transmitter and the receiver as:

$$\hat{d}_i = d_0 10^{\frac{P_0 - \hat{P}_i(d_i)}{10n_p}} = d_i + \underline{e}_i, \quad (2)$$

which is a log-normal random variable with $\underline{e}_i = d_i e^{\frac{-\underline{S}_i \ln 10}{10n_p}} - d_i$.

For long distance applications, such as satellite navigation, it is practically not feasible to estimate ranges with RSS. The flat tail of the power-distance curve makes the ranging accuracy very poor. Therefore, RSS is usually used for ranging in short range applications. Under Line of Sight (LoS) condition for distances up to 15 m, the standard deviation (STD) of the estimated range using UWB signals, is about 1 m based on total received power [41]–[44] and is about 0.5 m if only first path power is considered [44].

B. ToA

ToA is the measured time at which a signal first arrives at a receiving end. It is widely used for range estimation. A review of the ToA estimation technologies can be found in [1]. When the clocks of all units are synchronized, the estimator of range between the i -th transmitter and the receiver can be obtained by multiplying the difference between the ToA estimator and the Time of Transmission (ToT) with the propagation speed of the medium c :

$$\hat{d}_i = c(\hat{T}_{ToA,i} - \tau_{ToT,i}) = d_i + \underline{e}_i. \quad (3)$$

Please note that the term $\hat{\tau}_{ToA,i} - \tau_{ToT,i}$ is the signal travel-time, which is directly measured in a system, for instance like GPS, by cross-correlating the received signal with a locally generated replica. Of course, as a GPS receiver clock is in general not synchronized to GPS satellite clocks, we can only obtain a pseudo-travel-time. How to model clock misalignments will be presented in the next section.

The error term \underline{e}_i in (3) accounts for additive noise, multipath effects and Non-line of Sight (NLoS) effects (if applicable). A more detailed description of the main sources of errors can be found in [34]. Under LoS conditions, \underline{e}_i is usually modeled as a zero mean Gaussian variable [2], [3], while under NLoS conditions, \underline{e}_i is modeled as a positive bias superimposed on a zero mean Gaussian variable. The bias is the extra distance due to the blockage of the direct path, which can be modeled as an exponentially distributed random variable [3], [45], [46]. Ranging using ToA can have much higher accuracy than using RSS, e.g. centimeter level accuracy can be achieved with an UWB signal in LoS cases [1].

C. TDoA

When the receiver clock is asynchronous to the transmitter clocks, the ToA estimator seen by the receiver should be written as:

$$\hat{\tau}_{ToA,i} = \hat{\tau}'_{ToA,i} + \tau_c, \quad (4)$$

where $\hat{\tau}'_{ToA,i}$ is the ToA estimator according to the i -th transmitter clock, τ_c is the clock misalignment between the receiver clock and the transmitter clock. In this case, one can only get a pseudorange by multiplying (4) with c :

$$\hat{\rho}_i = c(\hat{\tau}_{ToA,i} - \tau_{ToT,i}) = d_i + b_c + \underline{e}_i. \quad (5)$$

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

$$\hat{\Delta}_i = \hat{\rho}_i - \hat{\rho}_r = d_i - d_r + \underline{e}_i - \underline{e}_r. \quad (6)$$

After the differencing operation with a reference measurement (indicated with subscript 'r'), the number of available measurements is reduced by 1, but the unknown b_c is also gone. One can either use the pseudoranges to estimate the position and the offset b_c , or use the pseudorange differences to estimate the position only. Note that in a fully synchronized system, there is no need to form TDoA measurements, since no clock offset exists.

TABLE I
UNKNOWNNS FOR DIFFERENT SYSTEMS.

unknowns	system
$x = [x_u^T, b_c]$	In case of TDoA with m pseudo range equations
$x = x_u$	In case of ToA/RSS with m range equations ($b_c = 0$), or TDoA with $m - 1$ pseudorange differences (b_c is canceled)

From now on, the term TDoA will be used for the cases where the user receiver clock is asynchronous to the synchronized transmitter clocks, and the term ToA/RSS for the cases where all the clocks are synchronized or no timing information is required for ranging.

D. Positioning Model

Here, the notation \underline{y} is used to represent range or pseudorange measurements in (2), (3) and (5), and the general model for the positioning problem can be given as

$$\underline{y} = A(x) + \underline{e}, \quad (7)$$

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

Although the value of n is not fixed in this paper, please note that it is usually equal to 3 (ToA/RSS) or 4 (TDoA) in a practical system.

For one single pseudorange measurement, the equation reads

$$\underline{y}_i = \underbrace{\|x_u - x_i\|}_{d_i} + b_c + \underline{e}_i, \quad i = 1 \dots m, \quad (8)$$

where x_i is the position vector of the i -th transmitter, and $\|\cdot\| = \sqrt{(\cdot)^T(\cdot)}$.

Since b_c is common for all equations, instead of (8), range difference equations are often used, where b_c is canceled and x_u is estimated with $m - 1$ difference equations. Taking the last measurement as the reference $r = m$, denoted as \underline{y}_r , the range difference equations can be formed as:

$$\underline{y}_i - \underline{y}_r = \|x_u - x_i\| - \|x_u - x_r\| + \underline{e}_i - \underline{e}_r, \quad i = 1 \dots m - 1. \quad (9)$$

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

$$\underline{y} = A(\hat{x}) + \partial_{x^T} A(\hat{x})(x - \hat{x}) + o(x - \hat{x}) + \underline{e}, \quad (10)$$

where $\partial_{x^T} A(x)$ denotes the Jacobian matrix of $A(x)$, and the Landau order term $o(x - \hat{x})$ indicates the remainder in the Taylor series, which can be ignored when \hat{x} is close to x . The linear part of (10) can be used as the approximation of the original equations in (7). With the linear approximation, the solvability of the inverse problem can be described as follows:

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

An exception to the above description would be the case with $\underline{e} = 0$, i.e. error free measurements. The problem is then actually consistent with a unique solution even when $m > \text{rank}(\partial_{x^T} A(\hat{x}))$. However, this case is not considered in this paper, as the assumption of $\underline{e} = 0$ is impractical.

In Figure 1, the geometric interpretations of consistent ToA/RSS systems and TDoA systems are given with $x_u \in \mathbb{R}^2$ for a better demonstration, which can be described as:

- *ToA/RSS*: for each transmitter, the unknown user is located on the circle centered at the transmitter with radius equal to the corresponding range measurement. Localization is done by determining the intersection of all the circles.
- *TDoA*: in this case, it is not possible to use a spherical interpretation since each pseudorange contains an additional unknown parameter b_c . As b_c is canceled in the pseudorange difference, the user is located on the hyperbola with two of the transmitters as the foci, and the difference of distances to the two foci is equal to the pseudorange difference between the same pair of transmitters. Hence, the user position can be found as the intersection of all the hyperbolas.

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

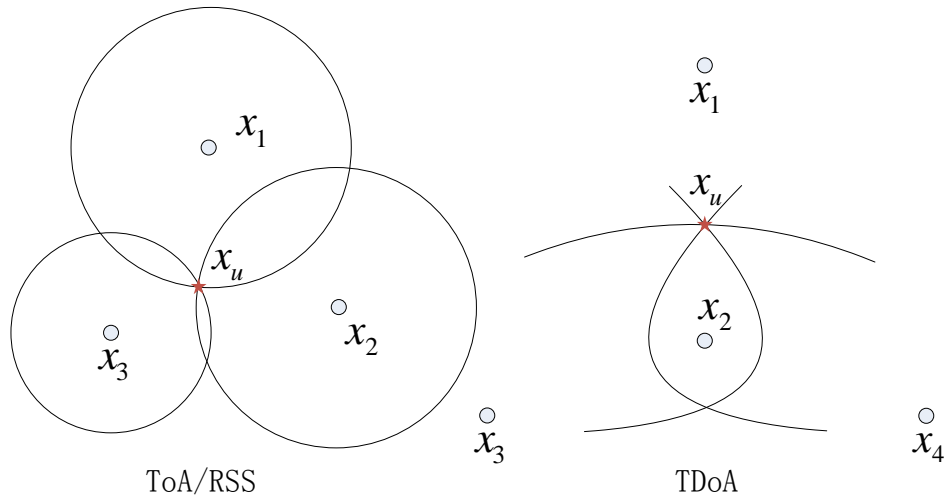


Fig. 1. Geometric interpretation of a consistent positioning problem. Transmitters are at known positions x_1 , x_2 , x_3 , and x_4 , and the user position x_u is unknown. Here we assume $x_u \in \mathbb{R}^2$ for easy illustration, i.e. $n = 2$ in case of ToA/RSS and $n = 3$ in case of TDoA.

III. NON-LINEAR LEAST-SQUARES ESTIMATION

A. Principle

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

$$F(x) = [y - A(x)]^T W [y - A(x)] = \|y - A(x)\|_W^2, \quad (11)$$

where W is a weight matrix. This is the weighted least-squares principle. For all values of $x \in \mathbb{R}^n$, $A(x)$ describes an n -dimensional manifold, embedded in \mathbb{R}^m . If the metric of \mathbb{R}^m is described by the positive definite matrix W , the scalar $\|y - A(x)\|_W$ equals the distance from y to the point $A(x)$ on the manifold. Hence, the problem of minimizing $F(x)$ corresponds to the problem of finding the point on the manifold $A(\hat{x})$, which has the least distance to y [5]. The geometry of the non-linear least-squares problem [5] is shown in Fig. 2.

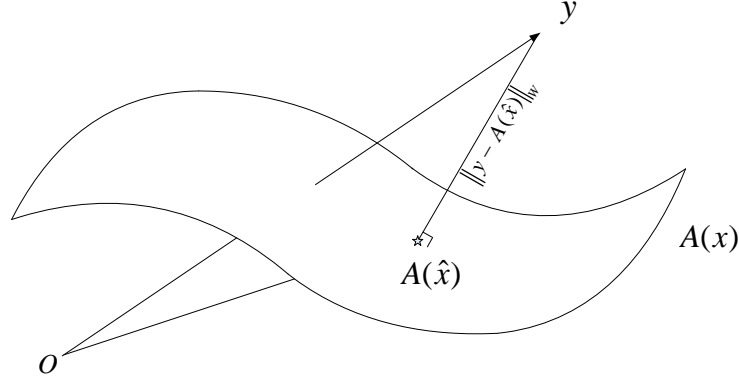


Fig. 2. Geometry of non-linear least-squares. The vector of collected observations \underline{y} is expected to lie on the manifold $A(x)$, but in practice it does not due to measurement errors.

B. Optimality Conditions

The sufficient and necessary conditions for a point \hat{x} to be a minimizer of $F(x)$ in (11) are shown in [47]:

$$\begin{aligned} a) \quad & \partial_x F(\hat{x}) = 0, \\ b) \quad & \partial_{xx^T}^2 F(\hat{x}) > 0. \end{aligned} \tag{12}$$

In a non-linear system, $F(x)$ could have multiple minima. This can be seen based on the first condition (12a). With $A(x)$ being non-linear, $\partial_x F(x)$ could contain non-linear terms of x , and $\partial_x F(x) = 0$ has multiple solutions. These solutions are usually categorized as local minimizers and a global minimizer. A point \hat{x} is said to be a local minimizer if $F(\hat{x}) < F(x)$ for all x near \hat{x} , e.g. for all x within a sphere (assuming $\dim(x) = 3$) centered at \hat{x} with a certain radius ϵ . A point is said to be a global minimizer if $F(\hat{x}) < F(x)$ for all $x \in R^n$.

Multiple minima can appear in both consistent and overdetermined systems. In a consistent 3-D ToA/RSS systems with $m = n = 3$, the 3 transmitters always define a plane. No matter which side of this plane the user is located, its position and its image position w.r.t the plane both satisfy $\partial_x F(\hat{x}) = 0$. Without additional information, it is not possible to pick the correct solution.

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

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

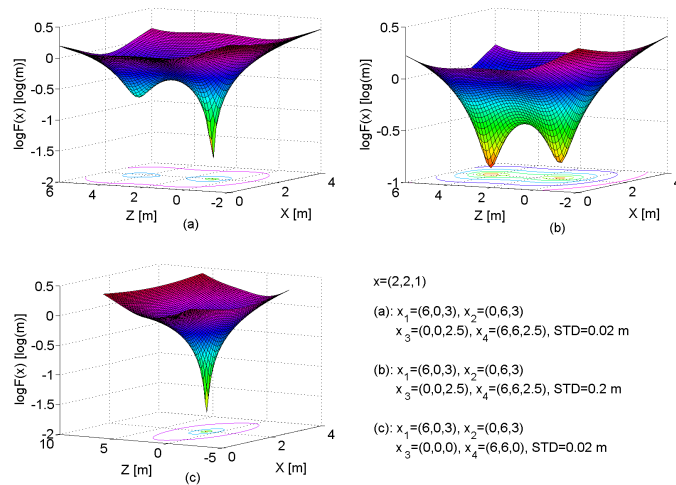


Fig. 3. x - z portion of the least squares objective function of three 3-D (x - y - z) positioning systems. The range measurement errors are assumed to be zero-mean Gaussian, independent and identically-distributed. The values of $F(x)$ are represented by different colors.

C. Least-Squares, Maximum Likelihood and Cramer-Rao Lower Bound

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

$$\hat{x} = \arg \max_x f_{\underline{y}}(y|x). \quad (13)$$

The Cramer-Rao Lower Bound (CRLB) indicates the smallest variance an unbiased estimator can possibly achieve [1], [34], [48]. Without considering the estimation method, the CRLB on the estimator of the model (7) is calculated as:

$$\text{CRLB} = \left\{ -\text{E} \left[\partial_{xx^T}^2 \ln f_{\underline{y}}(y|x) \right] \right\}^{-1} \quad (14)$$

When the joint PDF $f_{\underline{y}}(y|x)$ of the range measurements is *elliptically* contoured, the properly weighted least-squares estimate and the ML estimate are identical, and the variance achieves the CRLB, if the errors in \underline{y} have zero-mean. For example, $f_{\underline{y}}(y|x) = (\det(2\pi Q_{yy}^{-1}))^{-1/2} \exp\{-\frac{1}{2}\|y - A(x)\|_{Q_{yy}^{-1}}^2\}$, in the case of normally distributed measurements with zero-mean errors and Q_{yy}^{-1} the covariance matrix of \underline{y} . The ML estimate can be expressed as $\arg \max_x f_{\underline{y}}(y|x) = \arg \min_x \|y - A(x)\|_{Q_{yy}^{-1}}^2$, which is the same as the least-squares estimate when the weight matrix W is chosen as Q_{yy}^{-1} . The CRLB is then calculated as $\left\{ -\text{E} \left[-\frac{1}{2} \partial_{xx^T}^2 \|y - A(x)\|_{Q_{yy}^{-1}}^2 \right] \right\}^{-1} = \left[(\partial_{x^T} A(x))^T Q_{yy}^{-1} \partial_{x^T} A(x) \right]^{-1}$, which is identical to the expression of the least-squares estimator variance derived in [5], when $W = Q_{yy}^{-1}$.

D. Error Analysis

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

Here we review the general way [49] to derive the first and second moments of a (non-linear) least-squares estimator. Suppose the true position x can be obtained from $y = A(x)$ via a mapping $M(\cdot)$, i.e. $x = M(y)$, and its estimator \hat{x} can be obtained as $M(\underline{y})$. With \underline{y} containing error \underline{e} , i.e. $\underline{y} = y + \underline{e}$, a Taylor series expansion of \hat{x} at y gives an expansion in \underline{e} :

$$\hat{x} = M(y) + \partial_{y^T} M(y) \underline{e} + \frac{1}{2} \underline{e}^T \partial_{yy^T}^2 M(y) \underline{e} + \dots \quad (15)$$

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

$$\mu_{\hat{x}} = \mathbb{E}\{\underline{\mu}_{\hat{x}}\} = \frac{1}{2}\text{trace}\{\partial_{yy^T}^2 M(y)Q_{ee}\}, \quad (16)$$

$$Q_{\hat{x}\hat{x}} = \partial_{y^T} M(y)Q_{ee}(\partial_{y^T} M(y))^T, \quad (17)$$

$$\text{RMSE} = \sqrt{\text{trace}\{Q_{\hat{x}\hat{x}}\} + \mu_{\hat{x}}^T \mu_{\hat{x}}}, \quad (18)$$

where we assume $\mathbb{E}\{\underline{e}\} = 0$ and the terms with orders higher than 2 are ignored.

The rest of the task is to derive the expression of $M(\cdot)$. This is in general non-trivial in non-linear least-squares problems. However, in (16) and (17), only the expressions of the $n \times m$ Jacobian $\partial_{y^T} M(y)$ and the $n \times m \times m$ Hessian $\partial_{yy^T}^2 M(y)$ are required to evaluate the first and second moments of \hat{x} , which can be obtained as follows. The first necessary and sufficient condition (12a) can be further written as $(\partial_{x^T} A(\hat{x}))^T W[y - A(\hat{x})] = 0$. Expanding the left hand side into a Taylor series at x gives an expansion in $\hat{x} - x$. Using (15), an expansion in \underline{e} can be obtained with unknown coefficients $\partial_y M(y)$ and $\partial_{yy^T}^2 M(y)$. If \hat{x} is a least-squares solution, the condition (12a) holds for all \underline{e} and one can collect the terms of the same order and set them to zero. In this way, the expressions of $\partial_{y^T} M(y)$ and $\partial_{yy^T}^2 M(y)$ can be determined. More details of derivations and the full expressions for (16) and (17) can be found in [49].

In case one is interested in the higher order moments or the complete PDF of \hat{x} , one would need to propagate the PDF of \underline{y} through the mapping $M(\cdot)$. This can be done with Monte-Carlo or statistical bootstrapping techniques [50], which is however outside the scope of the current review.

IV. ITERATIVE DESCENT ALGORITHMS

In this section, we will review the ID methods based on [5], including the Steepest Descent method, the Newton method, the Trust Region method, the Gauss-Newton method, and the Levenberg Marquardt method. These methods are widely used to solve non-linear least-squares problems, including positioning. The mechanics of the ID methods can be described as:

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

According to [5], the estimates of the ID methods in each iteration are updated with a general form:

$$x_{k+1} = x_k - t_k Q(x_k) \partial_x F(x_k), \quad (19)$$

where t_k is a positive scalar and $Q(x_k)$ is an arbitrary but positive definite matrix, and with $F(x)$ as in (11).

An important performance measure of an iterative method is its rate of convergence, which can be derived by expanding (19) into a Taylor series at the final estimate \hat{x} :

$$x_{k+1} - \hat{x} = [I - t_k Q(\hat{x}) \partial_{xx^T}^2 F(\hat{x})](x_k - \hat{x}) + 2\text{nd and higher order terms}, \quad (20)$$

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

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

A. Steepest Descent Method

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

$$Q(x_k) = I, \quad (21)$$

and (19) can be further written as

$$x_{k+1} = x_k - t_k \partial_x F(x_k). \quad (22)$$

The motivation of this choice is that the opposite direction of the gradient vector, $-\partial_x F(x_k)$, represents a direction of maximum rate of decrease for the function $F(x)$ at x_k . Plugging (21) into (20) gives the rate of convergence for the steepest descent method, which is *linear* for points close to the solution \hat{x} .

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

$$\partial_{t_k} F(x_{k+1}) = -\partial_x F(x_{k+1})^T \partial_x F(x_k) = 0. \quad (23)$$

With the simple choice of $Q(x_k) = I$, the advantage of the steepest descent method is clearly its low-complexity (in terms of computational load). The involved highest order partial derivative of $F(x)$ is only one and no matrix needs to be inverted. However, the steepest descent method has the following two drawbacks [5]: 1) With the exact line search strategy, $\partial_x F(x_{k+1})$ is orthogonal to $\partial_x F(x_k)$, which means the steepest descent method has the tendency to zigzag when the contours of $F(x)$ are very elongated. 2) the performance of the steepest descent method is *variant* under reparameterization.

B. Newton Method

The Newton method is characterized by the choice of

$$t_k Q(x_k) = [\partial_{xx^T}^2 F(x_k)]^{-1}, \quad (24)$$

and the update equation then becomes

$$x_{k+1} = x_k - [\partial_{xx^T}^2 F(x_k)]^{-1} \partial_x F(x_k) \quad (25)$$

The motivation of the Newton method is that the root searching of the non-linear function $\partial_x F(x) = 0$ can be approximated by iteratively searching the root of its tangent, evaluated at the intermediate points x_k . Plugging (24) into (20) shows that the Newton method has a *quadratic* rate of convergence for points close to the solution \hat{x} . This makes it very effective in dealing with general non-linear minimization problems. However, one drawback of the Newton method is the heavy computational load required for the evaluation and inversion of the Hessian matrix $\partial_{xx^T}^2 F(x)$. The related calculations of $\partial_{xx^T}^2 A(x)$, a $m \times n \times n$ tensor, are usually quite involved. Another drawback of the Newton method is that it fails to work when $\partial_{xx^T}^2 F(x)$ is close to singular or fails to be positive definite. The Trust Region method is then developed to cope with these situations, and it is reviewed in the next subsection.

C. Trust Region Method

In case of a poorly conditioned $\partial_{xx^T}^2 F(x_k)$, the problem is highly non-linear and the contours of $F(x_k)$ can be very elongated.

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

$$t_k Q(x_k) = [\partial_{xx^T}^2 F(x_k) + \alpha_k R]^{-1}, \quad (26)$$

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

$$x_{k+1} = x_k - [\partial_{xx^T}^2 F(x_k) + \alpha_k R]^{-1} \partial_x F(x_k). \quad (27)$$

In this way, increasing the value of α_k results in a smaller step size $\|x_{k+1} - x_k\|$, and the trust region is contracted. Actually, the trust region method is capable of alternating between the Newton method (25) and the steepest descent method (22). That is, by adjusting the scalar α_k for $R = I$, the trust region

method switches between the quadratic approximation of $F(x)$ used by the Newton method (small α_k), and the linear approximation used by the steepest descent method (large α_k).

D. Gauss-Newton Method

Unlike the steepest descent method and the Newton method that use direct approximations of $F(x)$, the Gauss-Newton method completes the linearization *within* the $\|\cdot\|$ operator of $F(x) = \|y - A(x)\|_W^2$, which is done by replacing $A(x)$ with its linearized version [5]: $A(x) \approx A(x_k) + \partial_{x^T} A(x_k)(x - x_k)$.

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

$$t_k Q(x_k) = \frac{1}{2} \left((\partial_{x^T} A(x_k))^T W \partial_{x^T} A(x_k) \right)^{-1}, \quad (28)$$

by setting $\partial_{x^T} \|y - A(x_k) - \partial_{x^T} A(x_k)(x - x_k)\|_W^2 = 0$. The update equation then becomes

$$x_{k+1} = x_k - \frac{1}{2} \left((\partial_{x^T} A(x_k))^T W \partial_{x^T} A(x_k) \right)^{-1} \partial_x F(x_k) \quad (29)$$

The Gauss-Newton method does not have the zigzag problem of the steepest descent method, and only the first order derivative needs to be calculated. But the convergence rate is *linear*, which can be seen by plugging (28) into (20).

The Gauss-Newton method also struggles when the matrix $\partial_x A(x_k)^T W \partial_x A(x_k)$ is non-invertible. These difficulties can be overcome by an adaptive technique called the Levenberg-Marquardt method, which is reviewed next.

E. Levenberg-Marquardt Method

The Levenberg-Marquardt method [51] can be seen as a trust region modification of the Gauss-Newton method, since it takes the form

$$t_k Q(x_k) = \left[(\partial_{x^T} A(x_k))^T W \partial_{x^T} A(x_k) + \alpha_k R \right]^{-1}, \quad (30)$$

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

The salient properties of the ID methods are summarized in Table II.

TABLE II
SUMMARY OF REVIEWED ID ALGORITHMS.

Methods	Complexity	Convergence Rate	Comment
Steepest Descent	Simple $Q(x_k) = I$	Linear	May Zigzag
Newton	Heavy computations for $\partial_{xx^T}^2 F(x_k)$	Quadratic	
Trust Region	Similar to Newton	Linear with large α_k Quadratic with small α_k	Several α_k may be tested in one iteration
Gauss-Newton	Medium	Linear	
Levenberg-Marquardt	Similar to Gauss-Newton	Linear	Several α_k may be tested in one iteration

F. Initial Estimate

All iterative methods require an initial estimate x_0 as the starting point to gradually converge to the final solution. Intuitively, the closer the initial estimate is to the final solution, the smaller is the required number of iterations. Moreover, in case the objective function has multiple minima, the method will converge to the downward peak of the valley (see Fig. 3) where the $F(x_0)$ is located. Hence, obtaining a good initial estimate is one of the crucial problems that should be addressed when applying iterative techniques.

In long range maritime navigation and on-Earth satellite navigation, a good initial estimate is easy to obtain, e.g. even the center of the Earth works. It gets tricky however for indoor applications, where the scale of the system is much smaller than a satellite system and the prior information on whereabouts of the target is more limited. An additional challenge is that systems should be built with preferably (very) low-cost and low-power equipment. One promising way to obtain a good initial estimate is to use so called direct methods, which complete position estimation non-iteratively. These non-iterative methods are discussed in the next section.

V. DIRECT METHODS

There is a large number of direct methods documented in the literature, scattered across the fields of radar, aerospace engineering, oceanic engineering, (acoustic) signal processing and wireless communications. In this section, these direct methods are classified based on the least-squares objective function

to minimize, which is the soul of the least-squares estimation and determines the accuracy (in terms of RMSE) of the corresponding estimator.

Recall from (8), the non-linear part of the range equation takes the form $\sqrt{(\cdot)^T(\cdot)}$. To solve the non-linear problem without iterations, one needs to first obtain a linear term of x by applying a squaring operation to cancel the square-root. The new set of equations after squaring contains both linear and non-linear terms that involve the unknown vector x . The general form of the equations used by all direct methods can be written as:

$$\underline{y}_D = A_D x_u + B_D b_c + C_D \chi + \underline{e}_D, \text{ with } f(x_u, b_c, \chi) = 0, \quad (31)$$

where the subscript D stands for ‘direct’, \underline{y}_D is the new measurement vector obtained from \underline{y} , different for different direct methods, A_D , B_D , C_D are known matrices, χ is an unknown non-linear term (function) of x , and \underline{e}_D is the error term. Explicit examples of expressions for χ will be given in V-A. Note that the above equation is *linear* in the unknown parameters, when $f(x_u, b_c, \chi) = 0$ is ignored.

During least-squares estimation, the constraint $f(x_u, b_c, \chi) = 0$ should be taken into consideration since it provides information in addition to the model $\underline{y}_D = A_D x_u + B_D b_c + C_D \chi + \underline{e}_D$. One of the possible ways is to use a Lagrange Multiplier λ to account for $f(x_u, b_c, \chi) = 0$. The *rigorous* least-squares solution of (31) then can be found as:

$$\begin{aligned} [\hat{x}_u^T, \hat{b}_c, \hat{\chi}]^T &= \arg \min_{x_u, b_c, \chi} \left[\|\underline{y}_D - A_D x_u - B_D b_c - C_D \chi\|_{W_D}^2 \right], \\ \text{subject to: } & f(x_u, b_c, \chi) = 0, \end{aligned} \quad (32)$$

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

$$\|\underline{y}_D - A_D x_u - B_D b_c - C_D \chi\|_{W_D}^2 + \lambda f(x_u, b_c, \chi). \quad (33)$$

This solution satisfies both (12a) and (12b), where $F(x)$ is now the Lagrange function and $x = [x_u^T, b_c, \chi, \lambda]^T$. In the rest of this article, it is assumed by default that (12b) is satisfied when referring to a stationary point of a Lagrange function.

In (33), the weight matrix W_D is often chosen as the inverse of the covariance matrix of \underline{y}_D , i.e. $W_D = Q_{y_D y_D}^{-1}$. Please note that here we refer to (32) as a *rigorous*, least-squares solution to (31), but not a strict least-squares solution to the original positioning problem, since the objective function to minimize is different from the one in (11).

Furthermore, it should be noted that although with the direct methods some of the design matrices in (31) become measurement-dependent, that we do not use an errors-in-variables or Total Least-Squares

approach [52] for solving (31). Such an approach would increase the computational complexity, thereby making the direct methods less attractive.

In general, the *rigorous* least-squares solution to model (31) can only be approximated *iteratively*, due to the non-linear constraint $f(x_u, b_c, \chi) = 0$. Simplifications are therefore required, in order to obtain a solution directly with no iterations. In the literature, two major types of simplifications exist:

- 1) With the first type of simplification, the constraint is completely ignored. In this case, (31) becomes just a linear model

$$\underline{y}_D = A_D x_u + B_D b_c + C_D \chi + \underline{e}_D, \quad (34)$$

and the corresponding least-squares solution reads:

$$[\hat{x}_u^T, \hat{b}_c, \hat{\chi}]^T = \arg \min_{x_u, b_c, \chi} \|\underline{y}_D - A_D x_u - B_D b_c - C_D \chi\|_{W_D}^2 \quad (35)$$

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

$$\hat{x}_D = (F_D^T W_D F_D)^{-1} F_D^T W_D \underline{y}_D. \quad (36)$$

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

The solution (36) however is not valid if $\dim(\underline{y}_D) < \dim(x_u) + 1$ for ToA/RSS and $\dim(\underline{y}_D) < \dim(x_u) + 2$ for TDoA. Compared to the original problem in (7) and (8), one extra measurement is required to solve the problem with one additional unknown χ . Hence, some of the SOF methods do not work if the system (7) is consistent with $m = n$.

- 2) Instead of completely discarding the information $f(x_u, b_c, \chi) = 0$, the second type of simplification partially exploits the constraint. The final solution can be seen as a combination of solutions obtained with multiple objective functions, and the expression is different for different methods. In contrast to the SOF methods, the direct methods applying the second type of simplifications will be referred to as the Multiple Objective Function (MOF) methods. Moreover, the MOF methods are capable of solving a consistent system of (7).

In the following sections, the SOF and MOF methods are reviewed in detail, and the results show that although the derivations of the documented direct methods can be greatly different, the SOF methods are essentially the same in the sense that they can be seen as one class of solutions to a unique problem with just applying different weighting schemes.

A. Single Objective Function Methods

The first SOF method was introduced in [6] to solve range difference equations of TDoA systems. Later on, other SOF methods are found in [9]–[11], [15], [16] for TDoA systems and [19]–[23] for ToA/RSS systems. According to the number of differencing operations applied to (7), and therefore different choices of A_D , B_D and C_D in (34), the SOF methods are further sub-categorized into three groups, which are discussed as follows.

1) *No Differencing*: Squaring on both sides of (8) gives:

$$\underline{y}_i^2 = x_u^T x_u - b_c^2 - 2x_i^T x_u + 2\underline{y}_i b_c + x_i^T x_i + 2d_i \underline{e}_i + \underline{e}_i^2, \quad i = 1 \dots m, \quad (37)$$

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

In accordance with (34), the SOF methods with no differencing are characterized by the following choices of $\underline{y}_{D,i} = \underline{y}_i^2 - x_i^T x_i$, $A_{D,i} = -2x_i^T$, $\underline{B}_{D,i} = 2\underline{y}_i$, $C_{D,i} = 1$, $\chi = x_u^T x_u - b_c^2$ and $\underline{e}_{D,i} = 2d_i \underline{e}_i + \underline{e}_i^2$. Note that with ToA/RSS systems, one has $\underline{B}_D = 0$ and $\chi = x_u^T x_u$. The method introduced in [9], [10], falls in this category. Although the authors introduced the method with two steps, the expression of the estimator can be given in one single equation, as shown with (36). Please note that the squaring implies two effects on the new error vector \underline{e}_D : 1) it has a non-zero mean, and 2) its PDF is not elliptically contoured, though it is well motivated that the error vector $\underline{e} =$ (before squaring) follows a Gaussian distribution. This note holds true for all the direct methods reviewed in this paper and we will come back to this for further discussion in V-D.

2) *Single Differencing*: Single differencing can be applied either to (37), ending up with a set of *squared-range differences* or directly to the equations (8) followed by squaring, ending up with a set of *squared range-differences* (please mind the different hyphenation).

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

$$\underline{y}_i^2 - \underline{y}_r^2 = -2(x_i - x_r)^T x_u + 2(\underline{y}_i - \underline{y}_r) b_c + x_i^T x_i - x_r^T x_r + 2d_i \underline{e}_i - 2d_r \underline{e}_r + \underline{e}_i^2 - \underline{e}_r^2, \quad i = 1 \dots m - 1. \quad (38)$$

Note that the common non-linear term $x_u^T x_u - b_c^2$ in (37) is canceled. The SOF methods with single differencing then take the form $\underline{y}_{D,i} = \underline{y}_i^2 - \underline{y}_r^2 - x_i^T x_i + x_r^T x_r$, $A_{D,i} = -2(x_i - x_r)^T$, $\underline{B}_{D,i} = 2(\underline{y}_i - \underline{y}_r)$, $C_{D,i} = 0$ and χ is canceled. In case of ToA/RSS systems, the single difference SOF methods using squared-range differences can be applied by simply choosing $\underline{B}_D = 0$. Compared to the case with no differencing, the number of equations is reduced by one, but the number of unknowns is also reduced by one from $n + 1$ to n , with χ canceled.

The single differencing SOF methods using squared-range differences include the ones in [19]–[23]. The methods in [19], [20] follow exactly the formulation of (38). Slight differences exist in [21], where instead of \underline{y}_r : $r = m$, the reference is chosen as the average of all m available squared ranges: $\underline{y}_r^2 = \frac{1}{m} \sum_{i=1}^m \underline{y}_i^2$; in [22], the differencing operation is repeated m times for $r = 1, \dots, m$; and in [23], the authors suggest to pick the shortest estimated range as the reference. However, the number of linearly independent squared-range difference equations in [19]–[23] is $m - 1$ where by ‘linearly independent’ we mean that none of the equations can be produced by any linear combination of the other equations. As shown later in this section, the methods described in [19]–[23] are equivalent.

Other than dealing with squared-range differences, it is also possible to obtain direct solution using squared range-differences. Squaring on both sides of the range difference equation (9) gives:

$$(\underline{y}_i - \underline{y}_r)^2 = -2(x_i - x_r)^T x_u - 2(\underline{y}_i - \underline{y}_r) d_r + x_i^T x_i - x_r^T x_r + 2d_i \underline{e}_i - 2d_r \underline{e}_r + \underline{e}_i^2 - \underline{e}_r^2, \quad i = 1 \dots m - 1. \quad (39)$$

In this case, with single differencing, one measurement is consumed to cancel the unknown clock offset b_c , instead of χ in the case of squared-range differences. Hence, this type of SOF method should not be applied to ToA/RSS systems where no clock offset exists. The single differencing SOF methods using squared range-differences are then characterized by the following choices: $\underline{y}_{D,i} = (\underline{y}_i - \underline{y}_r)^2 - x_i^T x_i + x_r^T x_r$, $A_{D,i} = -2(x_i - x_r)^T$, $\underline{B}_{D,i} = 0$, $C_{D,i} = -2(\underline{y}_i - \underline{y}_r)$ and $\chi = d_r = \sqrt{(x_u - x_r)^T (x_u - x_r)}$. The SOF method that uses (39) is the one introduced in [15], [16].

3) *Double Differencing*: Similar to the case of single differencing, double differencing can also be applied to two different types of equations, i.e. squared-range differences in (38) and squared range-differences in (39). However, the equations obtained by double differencing these two types of equations are identical.

Dividing both sides of (38) by $\underline{y}_i - \underline{y}_r$ gives:

$$\underline{y}_i + \underline{y}_r = \frac{-2(x_i - x_r)^T x_u + x_i^T x_i - x_r^T x_r + 2d_i \underline{e}_i - 2d_r \underline{e}_r + \underline{e}_i^2 - \underline{e}_r^2}{\underline{y}_i - \underline{y}_r} + 2b_c, \quad i = 1 \dots m - 1. \quad (40)$$

The subscript rr will be used to denote references used in the second round of differencing. Suppose $rr = m - 1$, subtracting $\underline{y}_{rr} + \underline{y}_r$ from the first $m - 2$ equations in (40) gives:

$$\begin{aligned} \underline{y}_i - \underline{y}_{rr} &= -2 \left[\frac{(x_i - x_r)^T}{\underline{y}_i - \underline{y}_r} - \frac{(x_{rr} - x_r)^T}{\underline{y}_{rr} - \underline{y}_r} \right] x_u + \left[\frac{x_i^T x_i - x_r^T x_r}{\underline{y}_i - \underline{y}_r} - \frac{x_{rr}^T x_{rr} - x_r^T x_r}{\underline{y}_{rr} - \underline{y}_r} \right] \\ &+ \left[\frac{2d_i \underline{e}_i - 2d_r \underline{e}_r + \underline{e}_i^2 - \underline{e}_r^2}{\underline{y}_i - \underline{y}_r} - \frac{2d_{rr} \underline{e}_{rr} - 2d_r \underline{e}_r + \underline{e}_{rr}^2 - \underline{e}_r^2}{\underline{y}_{rr} - \underline{y}_r} \right], \quad i = 1, \dots, m - 2. \quad (41) \end{aligned}$$

Although it seems that only one differencing operation $\underline{y}_i - \underline{y}_{rr}$ exists, the above equation is obtained from (38) where another differencing has already been applied to obtain squared-range differences.

Eq. (41) can also be obtained by first dividing both sides of (39) by $\underline{y}_i - \underline{y}_r$:

$$\underline{y}_i - \underline{y}_r = \frac{-2(x_i - x_r)^T x_u + x_i^T x_i - x_r^T x_r + 2d_i \underline{e}_i - 2d_r \underline{e}_r + \underline{e}_i^2 - \underline{e}_r^2}{\underline{y}_i - \underline{y}_r} - 2d_r, \quad i = 1 \dots m - 1, \quad (42)$$

and then choosing $\underline{y}_{rr} - \underline{y}_r$ as the reference and subtracting it from the first $m - 2$ equations in (42), which gives exactly (41).

Clearly, in both cases of double differencing, one equation is consumed to cancel the clock offset b_c and another one for the non-linear term χ . Therefore, double differencing should not be applied to a ToA/RSS system which does not have a clock offset, i.e. single differencing is already enough to obtain a set of equations that contains only x_u as the unknown, see (38).

In contrast to (34), the SOF methods with double differencing are characterized by the following choices:

$$\underline{y}_{D,i} = \underline{y}_i - \underline{y}_{rr} - \left[\frac{x_i^T x_i - x_r^T x_r}{\underline{y}_i - \underline{y}_r} - \frac{x_{rr}^T x_{rr} - x_r^T x_r}{\underline{y}_{rr} - \underline{y}_r} \right] \quad (43)$$

$$\underline{A}_{D,i} = -2 \left[\frac{(x_i - x_r)^T}{\underline{y}_i - \underline{y}_r} - \frac{(x_{rr} - x_r)^T}{\underline{y}_{rr} - \underline{y}_r} \right] \quad (44)$$

and $B_D = C_D = 0$. The methods that belong to this subcategory include the ones in [6], [11]. In [6] the author first introduced the method for a consistent set of equations with $\dim(\underline{y}) = m = \dim(x_u) + \dim(b_c) = n$ and then dealt with the overdetermined case with $m > n$. Note that only in the overdetermined case, the method belongs to SOF methods, while in the consistent case it belongs to MOF methods, which will be covered later in Section V-B.

4) *Relating SOF Methods*: If two sets of equations can be linearly related, the corresponding solutions can be seen as obtained using different measurement weighting schemes and therefore are equivalent. Let us check this point by first assuming another set of equations, in contrast to (34):

$$\underline{y}'_D = A'_D x_u + B'_D b_c + C'_D \chi + \underline{e}'_D, \quad (45)$$

and \underline{y}'_D can be related to \underline{y}_D in (34) with a transformation matrix T : $\underline{y}'_D = T \underline{y}_D$. Hence, it holds that $(\cdot)'_D = T(\cdot)_D$ for all coefficients on the right hand sides of (34) and (45). Using the notation $F'_D = [A'_D, B'_D, C'_D]$ and $x_D = [x_u^T, b_c, \chi]^T$, the least-squares estimator of (45) can be written as:

$$\hat{\underline{x}}'_D = \arg \min_{x_D} \|\underline{y}'_D - F'_D x_D\|_{W'_D}^2. \quad (46)$$

With a specific choice of

$$W'_D = (T W_D^{-1} T^T)^{-1}, \quad (47)$$

Eq. (46) yields:

$$\hat{\underline{x}}'_D = (F_D'^T W_D' F_D')^{-1} F_D'^T W_D' \underline{y}'_D \quad (48)$$

$$= (F_D^T W_D F_D)^{-1} F_D^T W_D \underline{y}_D = \hat{\underline{x}}_D, \quad (49)$$

if T is invertible: $(TW_D^{-1}T^T)^{-1} = T^{-T}W_D T^{-1}$. Hence, two sets of equations, that can be linearly related by any invertible transformation matrix, provide *equivalent* solutions, with the specific choice of (47).

Our case is more complicated, since T is not invertible. For example, in (38), the numbers of equations and unknowns are both reduced by one, as compared to (37). The matrix, transforming (37) to (38), then takes the form:

$$T = \begin{bmatrix} 1 & 0 & -1 \\ & \ddots & \vdots \\ 0 & 1 & -1 \end{bmatrix}_{(m-1) \times m}, \quad (50)$$

which is non-invertible.

However, it can still be proved that when W_D' , the weight matrix for (38), is related to W_D , the weight matrix for (37), as $W_D' = (TW_D^{-1}T^T)^{-1}$, these two sets of equations provide identical least-squares solutions. The proof can be done by rigorously computing the expressions of the least-squares solutions for x_u in the two cases, which takes quite cumbersome calculations and is therefore not shown in this review. Interested readers can complete the proof using the block matrix inversion lemma [53].

To this end, it is clear that all the mentioned SOF methods are linearly related, and provide identical solutions with special choices of weight matrices, see (47). In general, these methods can be seen as realizations of a unique method with different weighting schemes.

B. Multiple Objective Function Methods

The MOF methods employ more than one objective function, and the corresponding estimator is in general not a minimizer for any of the employed objective functions, but a ‘combination’ of them. As a result, the MOF methods and the SOF methods are generally not equivalent. Currently, there exist two types of MOF methods, which are reviewed in the following.

1) *MOF Type I*: Recall from (35) and (32), the simplified and rigorous least-squares solutions to the squared range equations are obtained by minimizing two different objective functions and the correspond-

ing solutions then take the forms:

$$\text{Simplified: } \hat{x}_D = \arg \min_{x_D} F_1(x_D), \quad (51)$$

$$\text{Rigorous: } \hat{x}_D = \arg \min_{x_D} F_1(x_D), \text{ subject to: } f(x_D) = 0. \quad (52)$$

where $F_1(x_D) = \|\underline{y}_D - F_D x_D\|_{W_D}^2$ and $x_D = [x_u^T, b_c, \chi]^T$. The rigorous least-squares solution (52) can be found by searching for a stationary point of the Lagrange function:

$$F_2(x_D, \lambda) = \|\underline{y}_D - F_D x_D\|_{W_D}^2 + \lambda f(x_D). \quad (53)$$

which cannot be obtained non-iteratively without simplification, due to the presence of a non-linear constraint $f(x_D) = 0$. Rather than directly using (51) as the final solution like the SOF methods do, the MOF methods derive the estimators by combining the following equations:

$$\partial_{[x_{D,1}, \dots, x_{D,k-1}, x_{D,k+1}, \dots]} F_1(x_D) = 0 \rightarrow \hat{x}_D = f_1(x_{D,k}), \quad (54)$$

$$\partial_\lambda F_2(x_D, \lambda) = 0 \rightarrow f(\hat{x}_D) = 0, \quad (55)$$

where in (54) one of the unknowns in x_D , assumed as $x_{D,k}$, is used to represent all other unknowns. As $\underline{y}_D - F_D x_D$ is a linear function of x_D , $f_1(\hat{x}_{D,k})$ is also a linear function which can be further written as $f_1(\hat{x}_{D,k}) = a + b\hat{x}_{D,k}$. Combining the above two equations gives:

$$f(a + b\hat{x}_{D,k}) = 0. \quad (56)$$

In positioning problems, the above equation is quadratic [6]–[8], [12]–[14], [17], [18], and two possible solutions of $\hat{x}_{D,k}$ can be obtained. Together with (54), two sets of $[\hat{x}_u^T, \hat{b}_c]^T$ (part of x_D) can be calculated, and the final solution can be chosen as:

$$[\hat{x}_u^T, \hat{b}_c]^T = \arg \min_{\hat{x}_{u,j}, \hat{b}_{c,j}} F(x_u, b_c), \quad j = 1, 2, \quad (57)$$

where $F(x_u, b_c)$ is the least-squares objective function defined in (11).

In case of an overdetermined system (7) with $m > n$, the MOF estimator (57) is not equivalent to the rigorous solution (52). Nevertheless, with a consistent set of equations, (57) and (52) are identical, since in this case:

$$\|\underline{y}_D - F_D x_D\|_{W_D}^2 = 0 \quad (58)$$

$$f(\hat{x}_D) = 0 \quad (59)$$

$$\hat{\lambda} = 0 \quad (60)$$

In [7], Bancroft used the squared-range equations (37) for estimation, and the constraint takes the form $f(x_D) = x_u^T x_u - b_c^2 - \chi = 0$. This method was introduced for both overdetermined and consistent cases, i.e. $m \geq n$. In [8], Schau et al. only covered the consistent case $m = n$, and the squared range-differences (39) are used instead with $\chi = d_r^2$. The constraint becomes $f(x_D) = \|x_u - x_r\|^2 - d_r^2 = 0$. Note that b_c is canceled in this case, and the authors made the assumption that the reference transmitter is placed at the origin, i.e. $x_r = 0$. Later double range differences are used in [17] to form a *consistent* set of equations, and the constraint in use is equivalent to $f(x_D) = \|x_u - x_{rr}\| - \|x_u - x_r\| - (\underline{y}_{rr} - \underline{y}_r) = 0$. Similar methods can also be found in [6], [12]–[14], [18].

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

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

The linear relations in the first step of the MOF methods that follow the above mechanics are equivalent, irrespective of whether they are obtained from squared ranges, squared-range differences or squared range-differences. In the second step, there exist an unlimited number of choices for the quadratic constraint. For example, in ToA/RSS cases, for any point x_c in space, one can form a quadratic constraint:

$$\chi = (x_u - x_c)^T (x_u - x_c). \quad (61)$$

In accordance with (34), the new measurements and the coefficients are characterized as $\underline{y}_{D,i} = \underline{y}_i^2 - x_i^T x_i + x_c^T x_c$, $A_{D,i} = -2(x_i - x_c)^T$, $\underline{B}_{D,i} = 0$, $C_{D,i} = 1$, and $\underline{e}_{D,i} = 2d_i \underline{e}_i + \underline{e}_i^2$. For TDoA cases, a similar explanation applies.

In the consistent case, these MOF methods with different constraints are equivalent. However in general when the system is overdetermined, MOF Type I methods with different choices of the constraint are not equivalent. The best choice for x_c in (61) is obviously $x_c = x_u$. In this case, the problem is linear with the non-linear term $\chi = 0$, and the MOF Type I method gives a rigorous least-squares estimator to (34). In practice, however, it is not possible to choose $x_c = x_u$ since x_u is not known.

2) *MOF Type II*: Another type of MOF method, e.g. [24], does not follow the mechanics described above. This method, however, can be seen as an improved version of the methods in [19]–[23], where the squared-range differences (38) are used. Since [19]–[23] all deal with ToA/RSS systems with $B_D = 0$, the method in [24] will be reviewed with this type of system first.

Grouping (38) with $\underline{y}_r = d_r + \underline{e}_r$, and $\underline{y}_r^2 = d_r^2 + 2d_r \underline{e}_r + \underline{e}_r^2$, with $d_r = \|x_u - x_r\|$, one can formulate

two solutions:

$$\begin{aligned} [\hat{x}_u^T, \hat{d}_r^2]^T &= \arg \min_{x_u, d_r^2} \left\| \underline{y}_D - A_D x_u \right\|_{W_D}^2 + \|\underline{y}_r^2 - d_r^2\|_{w_1}, \\ \text{subject to:} \quad &\|x_u - x_r\|^2 - d_r^2 = 0. \end{aligned} \quad (62)$$

and

$$\begin{aligned} [\hat{x}_u^T, \hat{d}_r]^T &= \arg \min_{x_u, d_r} \left\| \underline{y}_D - A_D x_u \right\|_{W_D}^2 + \|\underline{y}_r - d_r\|_{w_2}, \\ \text{subject to:} \quad &\|x_u - x_r\|^2 - d_r^2 = 0. \end{aligned} \quad (63)$$

where w_1 and w_2 equal to $Var(\underline{y}_r^2)^{-1}$ and $Var(\underline{y}_r)^{-1}$, respectively. Moreover, it is assumed that there is no correlation between \underline{y}_D and \underline{y}_r^2 .

Assuming the above two solutions are equal, one can obtain three closed form solutions for x_u , and one of them is equivalent to the SOF solution. Similar as the MOF I method, the final solution should be picked as the one that corresponds to the smallest $F(x)$ in (11). Note that the requirement for this method to work is $m > n$ [24].

In fact the method in [24] can be improved, because the assumption that $Cov(\underline{y}_{D,i}, \underline{y}_r^2) = 0$ is not very realistic. The covariance, however, becomes very small with a specific choice of the reference. By choosing the average $\underline{y}_{mean}^2 = \frac{1}{m} \sum_{i=1}^m \underline{y}_i^2$ instead of \underline{y}_r^2 , the covariance becomes:

$$C(\underline{y}_{D,i}, \underline{y}_{mean}^2) = \frac{1}{m} D(\underline{y}_r^2), \quad (64)$$

which is negligible for a large m . Note that the unknown in (62) and (63) then become $[\hat{x}_u^T, \hat{d}_{mean}^2]^T$ and $[\hat{x}_u^T, \hat{d}_{mean}]^T$, respectively, and the constraint becomes $x_u^T x_u - \frac{2}{m} \sum_{i=1}^m x_u^T x_i + \frac{1}{m} \sum_{i=1}^m x_i^T x_i = d_{mean}^2$.

In TDoA cases, the method can work in a similar way as described for the ToA/RSS case. However, we suggest not to use this method in TDoA cases, since it still may suffer from numerical problems, just like the SOF methods. Hence, the details for the method to work in TDoA cases are not reviewed here in this article.

C. Geometric Interpretation

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

Take the ToA/RSS case in Fig. 1 as an example, line l_i ($i = 1, 2$) is defined by the two intersection points of the circle centered at x_i and the circle centered at x_3 . Assuming $x_r = x_3$, the perpendicular distance from x_r to line l_i can be expressed as,

$$\left| \frac{(x_i - x_r)^T (x_u - x_r)}{\|x_i - x_r\|} \right| = l_i. \quad (65)$$

The unknown position then can be located at the intersection of all these lines, when the system is consistent or measurements are error-free. In an overdetermined system with measurement errors, these lines generally do not intersect at a unique point. The geometric interpretation should then consider Fig. 2, with a slight difference that the manifold $A(\cdot)$ should be linear.

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

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

with d_i the true distance between x_i and x_u . Thus, together with (65), the following equation can be obtained

$$d_i^2 - d_r^2 = -2(x_i - x_r)^T x_u + x_i^T x_i - x_r^T x_r, \quad i = 1, \dots, m - 1, \quad (66)$$

which is identical to the error free version of (38), with $b_c = 0$ for ToA/RSS cases.

In TDoA cases, each line is defined by the intersections of two hyperbolas. In [17], the equations linear in x_u that correspond to these lines are found based on the fact that, for any point x_u on the hyperbola i , the ratio between the distance from x_u to x_r (the common focal point for all i , e.g. $x_r = x_2$ in the case of Fig. 1) and the distance from x_u to the directrix (plane) is the eccentricity of the surface (see Fig. 4). This gives

$$\|x_u - x_r\| = e_{c,i} [d_{x,i} - n_i^T (x_u - x_r)], \quad i = 1, \dots, m - 1 \quad (67)$$

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

Any two of the equations defined by (67) can be used to cancel the left hand side term $\|x_u - x_r\|$. Picking $rr = m - 1$ as the reference, subtracting it from all other equations gives a set of $m - 2$ linear equations:

$$(e_{c,i} n_i - e_{c,rr} n_{rr})^T (x_u - x_r) = e_{c,i} d_{x,i} - e_{c,rr} d_{x,rr}, \quad i = 1, \dots, m - 2, \quad (68)$$

which, after rearranging, is identical to the error free version of (40).

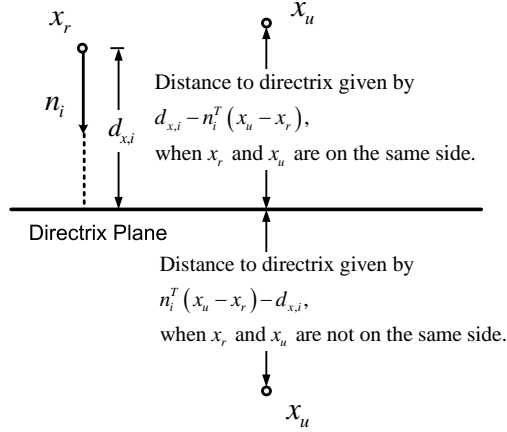


Fig. 4. According to plane geometry, quadratic curves (where x_u lies) can be defined in terms of a focal point (in our case, x_r), a directrix, and by their eccentricity. The directrix is a line in the plane and the eccentricity represents the ratio of the distance of a point on the curve to the focus and the directrix. The generalization to three dimensions is made by considering the directrix to be a plane. Here n_i is a unit vector pointing from x_r to the directrix. Please note that this example only applies to the TDoA case.

2) *MOF Methods Type I*: Here the geometric interpretation for the MOF method in [7] is shown for a 2-D ToA/RSS system with $\chi = x_u^T x_u$. In this case, (54) can be written in the form:

$$\begin{cases} x_{u,1} = a_1(x_{u,1}^2 + x_{u,2}^2) + b_1 \Rightarrow (x_{u,1} - \frac{1}{2a_1})^2 + x_{u,2}^2 = \frac{1-4a_1b_1}{4a_1^2} \\ x_{u,2} = a_2(x_{u,1}^2 + x_{u,2}^2) + b_2 \Rightarrow x_{u,1}^2 + (x_{u,2} - \frac{1}{2a_2})^2 = \frac{1-4a_2b_2}{4a_2^2} \end{cases} \quad (69)$$

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

Clearly, (69) represents two circles in the $x_{u,1}-x_{u,2}$ plane, centered at $(\frac{1}{2a_1}, 0)$ and $(0, \frac{1}{2a_2})$, with radius $\sqrt{\frac{1-4a_1b_1}{4a_1^2}}$ and $\sqrt{\frac{1-4a_2b_2}{4a_2^2}}$, respectively. The two intersections represent two possible solutions and the final estimate is chosen based on (57).

In a fully determined system, with $m = n = 2$, these two circles intersect at two points which are exactly the two intersections of the two measurement circles, centered at x_1 and x_2 with elements in \underline{y} as radius, see Fig 5. Without additional information about the unknown user position x_u , it is not possible to tell which intersection point corresponds to the true position of the user, since they both satisfies $F(x) = 0$, with $F(x)$ defined in (11).

When $m = 3 > n = 2$ and $\underline{e} = 0$, it is much easier to pick the final solution, since in this case,

the five circles, including two circles of (69) and three measurement circles, intersect at a unique point, which is the true user position, see Fig 6.

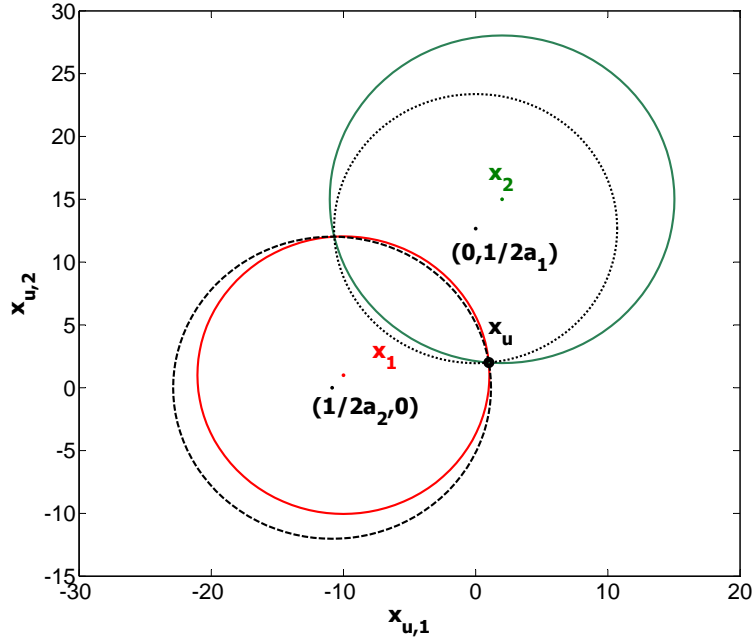


Fig. 5. 2D geometric view of the MOF I solution in determined case with 2 measurements. Solid: measurement circles; Dashed/Dotted: two circles represented by (69).

3) *MOF Methods Type II*: Here the geometric interpretation for an inconsistent ToA/RSS based system is given in Fig. 7.

The column vectors of the matrix A_D span a linear manifold in \mathbb{R}^m , which is the range space of $A_D : \mathfrak{R}(A_D)$. Without loss of generality, it is assumed that x_r is at the origin. The estimated residual \hat{e}_D is expressed as:

$$\hat{e}_D = y_D - A_D \hat{x}_u = \left[I - A_D (A_D^T W_D A_D + \hat{\lambda} I)^{-1} A_D^T W_D \right] y_D, \quad (70)$$

where the expression for \hat{x}_u is obtained by taking the first order derivative of either (62) or (63).

According to Section V-B2, three values of $\hat{\lambda}$ can be obtained non-iteratively and one of them is equal to zero. With $\hat{\lambda} = 0$, $\left[I - A_D (A_D^T W_D A_D + \hat{\lambda} I)^{-1} A_D^T W_D \right]$ is an orthogonal projector, and \hat{e}_D is orthogonal to $\mathfrak{R}(A_D)$. In this case, the solution \hat{x}_u follows from finding the vector $A_D x_u \in \mathfrak{R}(A_D)$ which has the least distance to y_D . This is identical to the solution of the SOF methods where no constraint

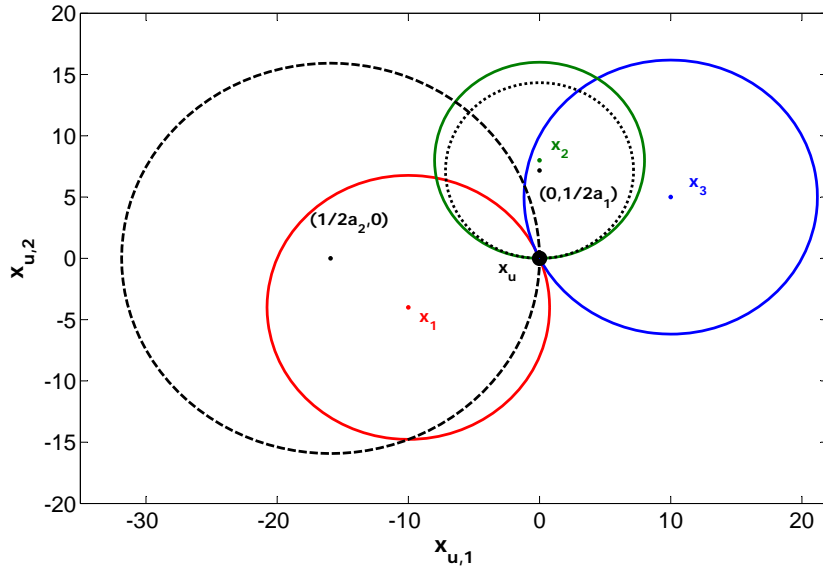


Fig. 6. 2D geometric view of the MOF I solution in consistent case with 3 measurements. Solid: measurement circles; Dashed/Dotted: two circles represented by (69).

is considered. In cases of $\hat{\lambda} \neq 0$, the residual vector is not orthogonal to $\Re(A_D)$ and the corresponding solution is different from the SOF estimate.

In the consistent cases, y_D lies in $\Re(A_D)$, and one always has $\hat{e}_D = 0$.

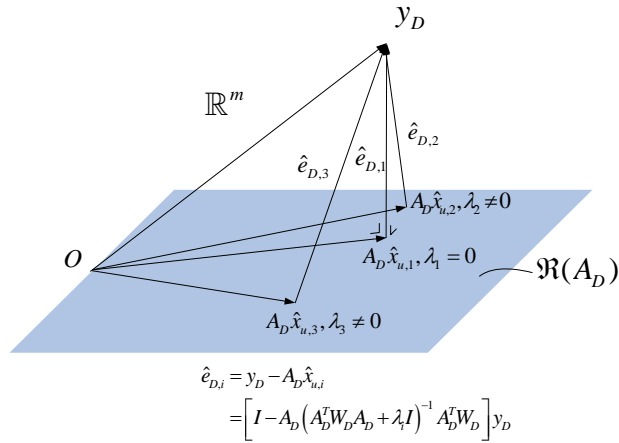


Fig. 7. The geometric interpretation of the solution of the Type II MOF method.

D. Error Analysis

1) *Bias and Variance*: For the SOF methods, the expressions for the first and second moments of a linear least-squares estimator are rather simple. In case of an error-free design matrix F_D :

$$\mu_{\hat{x}_D} = (F_D^T W_D F_D)^{-1} F_D^T W_D E\{\underline{y}_D - y_D\}, \quad (71)$$

$$Q_{\hat{x}_D \hat{x}_D} = (F_D^T W_D F_D)^{-1} F_D^T W_D Q_{y_D y_D} W_D F_D (F_D^T W_D F_D)^{-1}, \quad (72)$$

which can be directly used for e.g. ToA/RSS cases.

In TDoA cases, however, the design matrix \underline{F}_D contains measurements \underline{y} . The derivations for the first and second moments of \hat{x}_D then become much more complicated and some approximations are required. The details of the expressions in TDoA cases are skipped in this review, and interested readers can find the relevant results in [11].

For the MOF methods, the error analysis of the corresponding estimators is not available in the literature yet. One possible way to derive the MSE is to follow the general way described in III-D. This part, however, is left for future research and is not included here.

2) *Optimality*: Taking (37) as an example, one has $\underline{y}_{D,i} = y_i^2 - x_i^T x_i$, and hence:

$$\underline{y}_{D,i} - y_{D,i} = (e_i + y_i)^2 - y_i^2 \quad (73)$$

With the assumption that e_i follows a Gaussian distribution, i.e. $e_i \sim \mathcal{N}(\mu_{e,i}, \sigma_{e,i}^2)$, one has:

$$\sigma_{e,i}^{-2} (\underline{y}_{D,i} - y_{D,i} + y_i^2) \sim \chi^2(1, (\mu_{e,i} + y_i)^2 / \sigma_{e,i}^2) \quad (74)$$

where $\chi^2(1, (\mu_{e,i} + y_i)^2 / \sigma_{e,i}^2)$ represents a non-central chi-square distribution. Eq. (74) shows that the measurements $\underline{y}_{D,i}$ used by the direct methods are biased and the PDF of the measurement vector \underline{y}_D is not elliptically contoured. This means the corresponding estimator is biased, see (71), and the solution is neither equivalent to the ML estimator nor achieving CRLB, even with $\mu_{e,i} = 0$ and $W_D = Q_{y_D y_D}^{-1}$. In this sense, the direct methods do not provide optimal solutions. However, the direct methods are still very attractive for their low complexity, and a comparison of computational costs is given in VII. The outcomes of the direct methods can be used in two ways:

- 1) as initial estimates for the iterative methods, which usually leads to the same position estimates as if the true positions were fed to the iterative methods. This is shown at least with all the validation setups used in this paper, see VIII-D.
- 2) directly as final results, when the direct and iterative methods provide estimators with minor difference in accuracy. For example, in case the measurement standard deviation is less than 5

cm, the RMSE difference between estimates from direct methods and the Gauss-Newton method are less than 2.5 cm (Fig. 11).

In practice, how to use the direct methods of course depends on the applications and how the users want to balance between computation cost and accuracy. Moreover, $W_D = Q_{y_D y_D}^{-1}$ is still a valid choice, as the measurements with larger variances are given less weights during estimation.

E. Comparison of Direct Methods

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

The SOF methods are widely documented in the literature [6], [9]–[11], [15], [16], [19]–[23]. It has been shown that these methods are essentially equivalent. These methods are mutually related through their measurement weighting schemes, and they share the following strong and weak points:

- **Strong Point:** The final solution is uniquely obtained from a set of linear equations. As compared to the MOF methods, no extra quadratic equation needs to be solved and there is no need to check certain criteria to pick the final estimate out of several candidates.
- **Weak Point:** The non-linear constraint is neglected (loss of information). Thus, the SOF methods work only with a redundant set of equations (7) where $m > n$. Moreover, the SOF methods may face numerical problems in TDoA systems, as for instance an inverse of a close-to-singular matrix. This will be shown later in Section VIII.

The reviewed MOF methods include Type I: [6]–[8], [12]–[14], [17] and Type II: [24], which are in general not equivalent to each other. The first type of MOF methods has the following characteristics:

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

The second type of MOF method has the following strong and weak points:

- **Strong Point:** Improvement on accuracy is achieved non-iteratively, as compared to SOF methods, by combining multiple solutions that exploit the non-linear constraint.
- **Weak Point:** This method can only deal with the overdetermined case $m > n$ and the final solution is selected from three candidates obtained by solving an extra function.

VI. 1-DIMENSIONAL ITERATIVE METHODS

Based on the discussions in the previous section, the SOF methods and the MOF methods are capable of finding the position estimate non-iteratively. However, with the employed simplifications, neither SOF

TABLE III
SUMMARY OF REVIEWED DIRECT ALGORITHMS.

Methods	Characterization
SOF: No Differencing Eq. (37), [9], [10]	$\underline{y}_{D,i} = \underline{y}_i^2 - x_i^T x_i$ $A_{D,i} = -2x_i^T, \underline{B}_{D,i} = 2\underline{y}_i$ $C_{D,i} = 1, \chi = x_u^T x_u - b_c^2$
SOF: Single Differencing Eq. (38), [19]–[23]	$\underline{y}_{D,i} = \underline{y}_i^2 - \underline{y}_r^2 - x_i^T x_i + x_r^T x_r$ $A_{D,i} = -2(x_i - x_r)^T, \underline{B}_{D,i} = 2(\underline{y}_i - \underline{y}_r)$ $C_{D,i} = 0 \text{ and } \chi \text{ is canceled}$
SOF: Single Differencing Eq. (39), [15], [16]	$\underline{y}_{D,i} = (\underline{y}_i - \underline{y}_r)^2 - x_i^T x_i + x_r^T x_r$ $A_{D,i} = -2(x_i - x_r)^T, \underline{B}_{D,i} = 0$ $C_{D,i} = -2(\underline{y}_i - \underline{y}_r) \text{ and } \chi = \sqrt{(x_u - x_r)^T (x_u - x_r)}$
SOF: Double Differencing Eq. (41), [6], [11]	$\underline{y}_{D,i} = \underline{y}_i - \underline{y}_{rr} - \left[\frac{x_i^T x_i - x_r^T x_r}{\underline{y}_i - \underline{y}_r} - \frac{x_{rr}^T x_{rr} - x_r^T x_r}{\underline{y}_{rr} - \underline{y}_r} \right]$ $\underline{A}_{D,i} = -2 \left[\frac{(x_i - x_r)^T}{\underline{y}_i - \underline{y}_r} - \frac{(x_{rr} - x_r)^T}{\underline{y}_{rr} - \underline{y}_r} \right]$ $B_{D,i} = C_{D,i} = 0 \text{ and } \chi \text{ is canceled}$
MOF I: [6], [7], [12]–[14], [18]	can be characterized as all the listed SOF methods
MOF II: Eq. (62), (63) [24]	same characterization as SOF Single Differencing Eq. (38)

methods nor MOF methods provide rigorous least-squares solutions to the squared-range equations with or without differencing.

In this section, two different types of iterative techniques that bring improvements on estimation accuracy are reviewed:

- 1) The first type estimates the position by solving (76). Through eigenvalue decomposition, the solution is found by determining the roots of a high order polynomial.
- 2) The second type first transforms the multidimensional positioning problem to a 1-D optimization problem, and then iterates.

The interesting feature of both iterative techniques is that the solution is obtained with only *1-dimensional* iterations (1DI), i.e. during the iterations, only one single parameter needs to be calculated and updated.

The advantages of 1-D iteration are twofold:

- 1) The computations in each iteration, compared to a full n -D iterative case, are reduced by a factor of up to $1/n$ [27], [28].
- 2) The heavy computations required by the evaluation of $\partial_{xx^T}^2 F(x)$ in the Newton method is greatly relieved with $\dim(x) = 1$. Thus one can enjoy the quadratic rate of convergence brought by the Newton method with much less computations than in the case of $\dim(x) = n$.

A. 1DI Type I

The methods introduced in [25], [26] improve the direct methods by fully exploiting the non-linear constraint $f(x_D) = 0$, and the solution is obtained exactly as (32):

$$\hat{x}_D = \arg \min_{x_D} \|\underline{y}_D - F_D x_D\|_{W_D}^2, \text{ subject to: } f(x_D) = 0, \quad (75)$$

which can be obtained by searching for a stationary point of:

$$\|\underline{y}_D - F_D x_D\|_{W_D}^2 + \lambda f(x_D). \quad (76)$$

The procedure can be described as follows,

- 1) Taking the partial derivative of (76) w.r.t. x_D and setting it to zero, gives an expression of \hat{x}_D in $\hat{\lambda}$. By means of eigenvalue decomposition, each element of \hat{x}_D can be further written as an explicit expression in $\hat{\lambda}$, i.e. $\hat{x}_{D,k} = g_k(\hat{\lambda})$, $k = 1, \dots, \dim(x_D)$.
- 2) Taking the partial derivative of (76) w.r.t. λ and setting it to zero, gives $f(\hat{x}_D) = 0$.
- 3) Plugging $\hat{x}_{D,k} = g_k(\hat{\lambda})$ into $f(\hat{x}_D) = 0$ gives an explicit high order polynomial in $\hat{\lambda}$, which can be solved with the reviewed iterative search techniques. With the single unknown λ in this step, the iteration is performed in 1-D.
- 4) Once $\hat{\lambda}$ is available, \hat{x}_D can be obtained via $\hat{x}_{D,k} = g_k(\hat{\lambda})$.

The identical result can also be obtained without the eigenvalue decomposition in Step 1, but the explicit polynomial in λ in Step 3 cannot be obtained, and the solution needs to be calculated iteratively with full dimensions.

B. 1DI Type II

The idea of another type of 1DI method, introduced in [27], [28], can be described in general terms as follows:

- 1) Obtaining a mapping between $x = [x_u^T, b_c]^T$ and a scalar parameter $z: x = L(z)$, with $L: R \rightarrow R^n$ and plugging the relation into (7) gives:

$$\underline{y} = A(x) + \underline{e} = A(L(z)) + \underline{e}. \quad (77)$$

- 2) The least-squares solution to (77) is obtained as:

$$\hat{x} = L(\hat{z}) = L\left(\arg \min_z \|\underline{y} - A(L(z))\|_W^2\right), \quad (78)$$

If the mapping $x = L(z)$ is error-free, the estimator (78) is better than (11), in both the sense of accuracy and computational complexity, since the objective function in (78) contains less unknowns than (11) due to the additional information $x = L(z)$. Note that one reduces computational complexity most when z is a scalar (1-D vector).

Now the unsolved issue is how to find the mapping $L(z)$ in Step 1. In practice, an error-free mapping is usually not available, but for a positioning problem, one can always estimate it using the data at hand based on the SOF methods. Hence, the mapping in Step 1 becomes $L(z, \underline{e})$, which contains measurement errors, and (78) becomes

$$\hat{x} = L(\hat{z}, \underline{e}) = L\left(\arg \min_z \|\underline{y} - A(L(z, \underline{e}))\|_{Q_{yy}^{-1}}^2, \underline{e}\right). \quad (79)$$

Note that the first step of the Type II 1DI method is identical to the SOF methods or the first step of the Type I MOF methods. However, in the second step of the Type II 1DI method, the single parameter is iteratively estimated using the full non-linear model (77). In this way, improvement on the accuracy is achieved.

Another issue to be discussed is how to choose the scalar z . In [27], [28], the author suggests to choose z as the coordinate along the worst precision direction, which can be done using the eigenvalue decomposition [27]. However, the decomposition can be simply skipped by directly choosing the coordinate along the zenith direction (the height) [28], which typically is of worst precision in on-Earth satellite navigation and also in typical indoor positioning applications, where the room heights are generally smaller than lengths and widths.

C. Comparison and Error Analysis

Both of the two types of 1DI methods provide possible estimation accuracy improvements over the direct methods, but their differences can be described as follows, w.r.t. estimation accuracy and computational complexity.

- 1) Accuracy: The estimator of the Type I 1DI method, obtained with successful convergence, closely approximates the *rigorous* least-squares solution to (31). The estimator of the Type II 1DI method, however, is *not* a strict least-squares solution to the original problem (7), due to the estimation errors in $x_u = L(z, \underline{e})$.
- 2) Complexity: the eigenvalue decomposition is *skippable* with the Type II 1DI method to obtain a solution with 1-D iterations, but is *necessary* for the Type I 1DI method. In general, the decomposition itself is a non-linear problem and it has to be solved iteratively for a $m \times m$ matrix with $m \geq 3$. In [25], [26], the matrix to decompose is of 5×5 and 4×4 for 3-D TDoA and ToA/RSS positioning, respectively.

The error analysis for both types of 1DI methods can be done by following the general way described in [49]. The results can be found respectively in [26] for Type I (with simplifications) and [28] for Type II, which are quite complicated and therefore not included in this review.

VII. COMPUTATIONAL COSTS

In general the computation costs of different methods within a group, i.e. ID, direct and 1DI methods, are similar, exception is the MOF II method, which takes more computational cost than other direct methods. The computational cost differences between different groups are quite obvious. Taking the Gauss-Newton method and a direct method for example, in each iteration of the Gauss-Newton method, the computational cost is very close to a direct method. Then for k iterations, the computation cost of the Gauss-Newton method is k times of that of a direct method. The comparison of a direct method to a 1DI method is that a direct method does not iterate at all, i.e. no 1-D iterations. The comparison between the Gauss-Newton method and a 1DI method is that in each iteration, a 1DI method iterates on only 1 parameter, and thus takes about $1/n$ computations compared to an ID method.

In this section, the computational costs of the SOF method [54], the MOF Type I method [7], the 1DI Type II method [28], and the Gauss-Newton method are addressed in detail.

The flop-counts for the SOF method and for the MOF Type I are given as:

$$\text{flops}_{\text{SOF}} = n^3/3 + 2mn^2 + 2n^2 + 7mn + 4m + 3n - 2/3, \quad (80)$$

$$\text{flops}_{\text{MOF,TD\o A}} = n^3/3 + 2mn^2 + 9mn + 3n^2 + 3m + 9n + 14, \quad (81)$$

$$\text{flops}_{\text{MOF,ToA/RSS}} = n^3/3 + 9mn + 2mn^2 + 3n^2 + 7m + 9n + 14, \quad (82)$$

where we have assumed diagonal weight matrices. Please note that the evaluation of (57) is a big contributor in (81)-(82). In fact, the MOF I may get two candidates that are equal and there might be

other ways to pick the final solution than (57). Hence, the computations related to (57) may be skipped or reduced.

The derivations of (80) to (82) can be done based on the flop-counts derived for the Gauss-Newton method and the 1DI Type II method in [28]:

$$\text{flops}_{\text{1DI,TD0A}} = n^3/3 + 2mn^2 + 2n^2 + 7mn + 4m + 3n - 2/3 + k(3mn + 8m + 2n), \quad (83)$$

$$\text{flops}_{\text{1DI,ToA/RSS}} = n^3/3 + 2mn^2 + 2n^2 + 7mn + 4m + 3n - 2/3 + k(3mn + 6m + 2n), \quad (84)$$

$$\text{flops}_{\text{GN,TD0A}} = k(n^3/3 + 2mn^2 + 5mn + n^2 - 3m + 3n - 3), \quad (85)$$

$$\text{flops}_{\text{GN,ToA/RSS}} = k(n^3/3 + 2mn^2 + n^2 + 8mn + m + 2n - 1), \quad (86)$$

where k is the number of iterations. Please note that here n is the dimension of x , whereas in [28] n is the dimension of x_u . Moreover, the 1DI type II method requires some computations to complete the transform stage, which requires (one time) computations similar to the SOF method, but with 1 less unknown parameter, which corresponds to $n^3/3 + 2mn^2 + 2n^2 + 7mn + 4m + 3n - 2/3$ in (83) to (84).

In Figure 8, an example is given with $m = 5$, $n = 4$ (TDoA), $n = 3$ (ToA/RSS), and $k = 5$. The results are as expected that the SOF method is the most simple one, the MOF Type I method takes a bit more computations (worst case with two different candidates), the 1DI Type II method is a bit more computationally costly than the MOF Type II due to iterations, and the Gauss-Newton method is the most complicated, which takes about k times the computations of the SOF method.

VIII. PRACTICAL RESULTS

A. System Setup

The radio system used to generate range results to demonstrate the performance of the reviewed algorithms, consists of one transmitter and one receiver, i.e. the system is only capable of 1-D ranging. These limitations arise from the size, price and availability of the system hardware.

The setup of the radio UWB system is shown in Figure 9. The generator fires a Gaussian-like pulse with a duration of 50 ps. A 6 dB attenuator is placed before the transmit antenna to avoid multiple antenna reflections. The receiver part includes a sampling oscilloscope, a sampler unit and a PC. The sampling oscilloscope controls both the sampling unit and the pulse generator with trigger pulses, implying that the transmitter and the receiver are synchronized. The sampling unit samples the received signals using a stroboscopic sampling mechanism [55], which has a virtual sampling rate of 100 GHz. A bandpass filter in the 3-20 GHz bandwidth, to reduce out of band interference and noise, and an LNA, working in

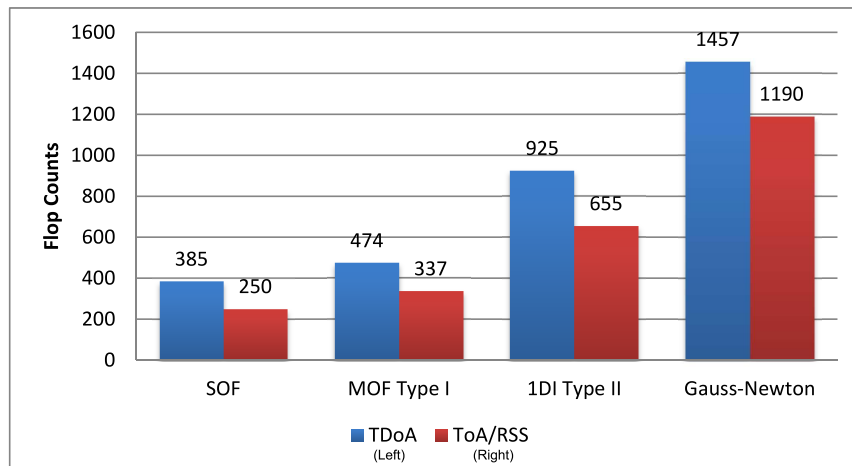


Fig. 8. A comparison between the computational costs of the SOF, MOF Type I, 1DI Type II and the Gauss-Newton methods. This figure is obtained with $m = 5$, $n = 4$ (TDoA), $n = 3$ (ToA/RSS), and $k = 5$, according to (80) to (86).

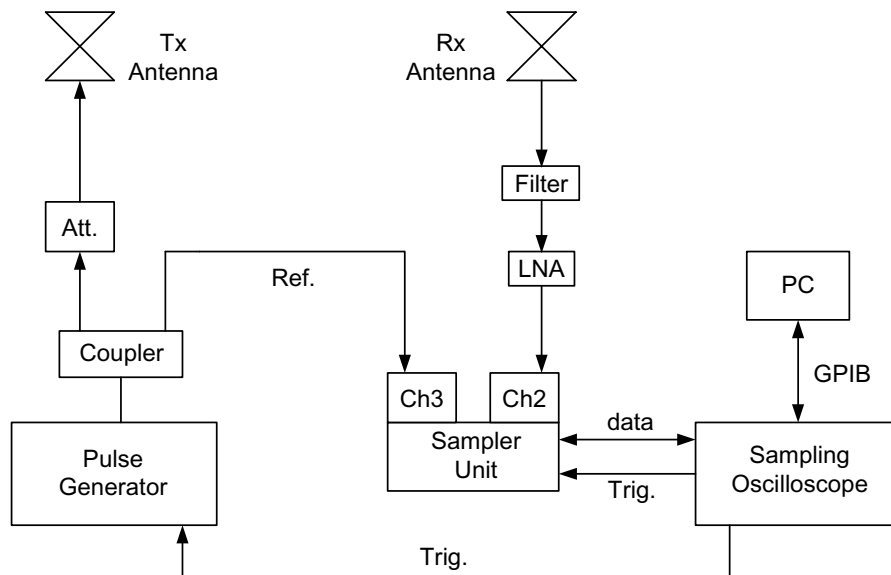


Fig. 9. Setup for UWB radio system.

the band between 0.1 and 12 GHz, follow the receiver antenna. As a result, it is possible to cover the bandwidth between 3.1 and 10.6 GHz allowed by the FCC for UWB radio transmissions [56]. Both the transmitter and the receiver use vertically polarized, omni-directional biconical antennas.

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

The ToA is determined as the arrival time of the first path of the CIR. The first path is defined here as the first local maximum of the envelope of the estimated channel, with an amplitude within 20 dB from the strongest peak. The distance between the transmitter and the receiver is then computed using (3).

The measurements were carried out in the faculty building of EEMCS of Delft University of Technology. Both the transmitter and the receiver antenna are placed 1.5 m above the floor. The distances between the transmitter and the receiver are between 2 and 15 m and 400 LoS measurements are collected in total. The corresponding true distances, measured with a laser-disto, are also recorded.

As the UWB radio measurement system is restricted to a single link (1 Tx, 1 Rx, synchronized), a 3-D positioning system is created virtually. Without loss of generality, the receiver is placed at the origin. From the total of 400 range estimates, a subset of 5 estimates is randomly chosen. For the i -th transmitter, spherical coordinates (d_i, θ_i, ϕ_i) are assigned, see Fig. 10, where θ_i and ϕ_i are randomly chosen from the intervals shown in Table IV. In this way, problematic geometries that make the positioning unsolvable, e.g. with all transmitters on a line, are avoided.

B. Ranging Results

1) *ToA*: The range error standard deviation (STD), calculated as

$$STD = \sqrt{\frac{1}{399} \sum_{l=1}^{400} \|\hat{d}_l - d_l\|^2}, \quad (87)$$

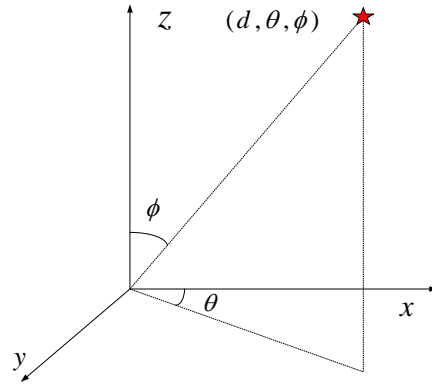


Fig. 10. Spherical coordinates of a single measurement. The receiver is at the origin and the transmitter at (d, θ, ϕ) .

TABLE IV
THE INTERVALS OF ZENITH ANGLE θ AND AZIMUTH ANGLE ϕ .

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

considering all 400 measurements obtained with a fixed bandwidth B , is given, cf. the values in Table V.

TABLE V
SIGNAL BANDWIDTH AND RANGING ERROR STD.

Bandwidth [GHz]	0.5	1.5	3.5	7.5
STD [cm]	11.0	3.5	1.8	1.5

TABLE VI
CHANNEL PARAMETERS AND RANGING ERROR STD.

Bandwidth [GHz]	1	1.5	3.5	7.5
σ_S [dB]	1.42	1.17	0.84	0.56
n_{fp}	1.86	1.88	1.88	1.87
STD [m]	1.34	1.05	0.78	0.48

C. RSS

The values for σ_S , n_p , and the STD of the range errors obtained under several of the tested bandwidths are summarized in Table VI.

D. Positioning Results

The tested methods include the SOF method, the MOF Type I method with $\chi = x_u^T x_u - b_c^2$, the MOF Type II method, the 1DI methods (both types), and the Gauss-Newton method. The empirical RMSE curves are plotted in Fig. 11 for the ToA case and in Fig. 13 for the TDoA case. On the horizontal axis, the range error STD is given.

In the ToA case, the Gauss-Newton method and the 1DI Type I method give best performance, since they closely approximate rigorous least-squares solutions to the corresponding models. The 1DI Type II method performs slightly worse than the rigorous solutions, but still better than the direct methods. The MOF Type II method brings improvements compared to the SOF method, but the absolute improvement is not very large since the ranging errors are only of centimeter level. With very precise range measurements (nearly) any method can achieve accurate positioning results; differences start to occur when it gets tougher with noisier measurements. In [24] with decimeter level or even meter level RSS ranging accuracy, the improvements are much more obvious, see Fig. 12. The MOF type I special (marked by ‘special’ in Fig. 11) performs very good with x_c chosen as the origin (exactly as x_u), since in this case the MOF type I solution is rigorous. On the contrary, the MOF Type I method performs worst with a regular choice of $x_c \neq x_u$. Again, we should note that in practice, it is not possible to choose x_c equal to x_u . In practice, x_c can be obtained in a similar way as how a initial estimate is picked: 1) as the outcome of a SOF method, or 2) as the geometrical center of the transmitters if one wants to reduce as much as possible the computation load.

In the TDoA case, the design matrices of the SOF method and the MOF Type II method are close to singular. The RMSEs of their estimates are hundreds of meters, and are not shown in Fig. 13. Other tested methods behave in a similar way as in the ToA case, only that the differences between various estimates are very small. Moreover, for the MOF Type I method, different choices of constraints hardly affect the final estimates. In Fig. 13, we show the result obtained with default choice $\chi = x_u^T x_u - b_c^2$.

E. Initial Estimate

For the iterative methods, i.e. the Gauss-Newton method and the 1DI methods, we have tested two different ways to obtain a initial estimate:

- 1) use the true values,
- 2) use the outcome of the MOF I [7]. Main reason is that this method works for both ToA/RSS and TDoA systems. Please note that for the 1DI Type I method, the initial estimate is always chosen as 0, see [25], [26].

Results show that exactly the same position estimates are obtained in either way. Moreover, using the outcome of a direct method as the initial guess makes the iterative methods converge faster than directly using the true values. This is due to the fact that the outcome of a direct method is in general closer to the final LS estimate, and this point is also verified in [57]. So, using the direct method to calculate the initial estimate for the iterative techniques is a very promising way.

IX. SUMMARY

In this article, range-based positioning algorithms are reviewed, including iterative descent methods, direct methods and methods with 1-D iterations.

The traditional iterative descent methods are widely used in non-linear estimation problems. They iteratively approximate a strict least-squares solution with an accuracy achieving the CRLB, in case of convergence. The drawbacks of these methods are that 1) the final estimates are obtained iteratively, i.e. no closed-form analytical expressions are available; 2) these methods all require an initial estimate to start with, which can be quite tricky for some applications, e.g. indoor localization.

One of the possible ways to provide an initial estimate for the iterative descent methods is to use a direct method, which provides a closed-form solution without any iterations. In this paper, two major types of the direct methods, namely SOF and MOF methods are reviewed. It has been shown that the SOF methods are strictly identical with a specific choice of the weight matrices. Compared to the iterative

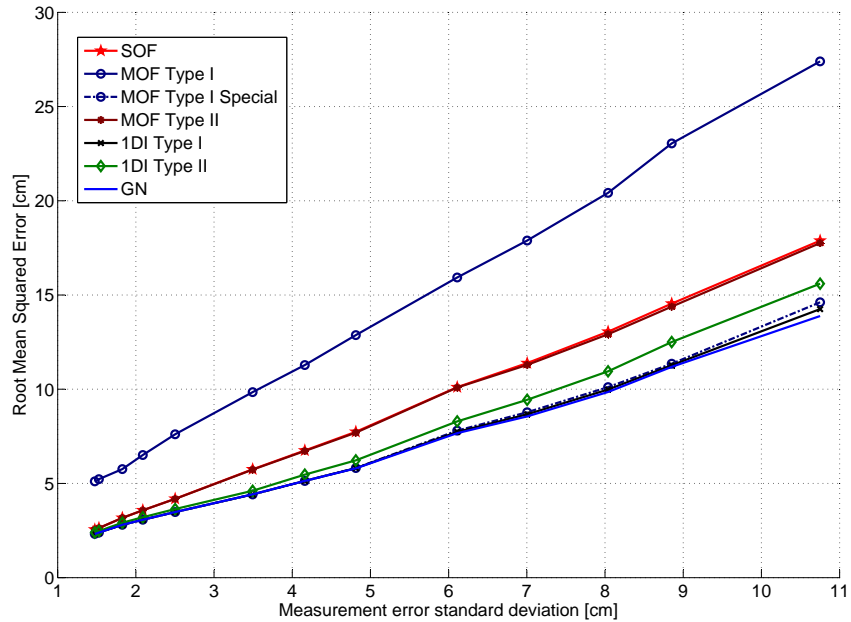


Fig. 11. RMSE of the estimates obtained using the actual ToA UWB data, for different methods.

descent methods, the advantage of the direct methods is clearly their simplicity (and small computational load), which however is achieved at the price of reduced accuracy.

There is also a third group of methods, referred in this paper as 1DI methods, which solve the problem with only 1-D iterations. These methods, on one hand, provide better accuracy than the direct solutions, and on the other hand, require less computations than applying iterative descent methods directly to the positioning problem. The Type I 1DI methods even approximates rigorous least-squares solutions to squared range equations which approximate the CRLB [26], but it requires a step of eigenvalue decomposition, which is itself a non-linear problem that needs to be solved iteratively in general.

A full summary of all reviewed methods can be found in Table VII, concerning computational complexity, positioning accuracy and requirements on the number of measurements. It should be pointed out that for a consistent system, there is no need to resort to iterative techniques and one should always use the Type I MOF methods, since the least-squares residual is zero.

This review paper intends to help system designers and researchers to have a better understanding of the existing range based positioning algorithms, their inter-relations and their performance, so that a balance between computational load and positioning accuracy can be found and the choice for a method can be made according to the requirements set by the application at hand.

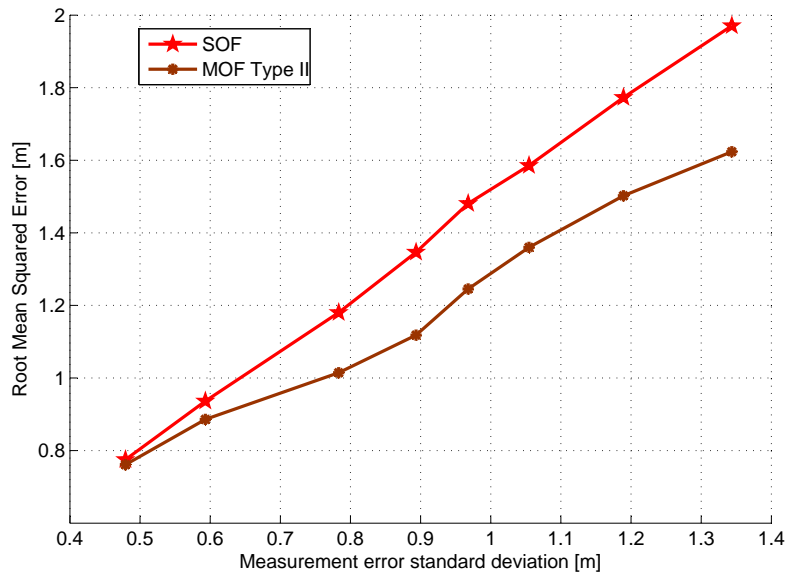


Fig. 12. RMSE of the estimates obtained using the actual RSS UWB data, for SOF and MOF type II methods.

REFERENCES

- [1] S. Gezici, Z. Tian, G. Giannakis, H. Kobayashi, A. Molisch, H. Poor, and Z. Sahinoglu, "Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks," *Signal Processing Magazine, IEEE*, vol. 22, no. 4, pp. 70–84, July 2005.
- [2] P. Misra and P. Enge, *Global Positioning System: Signals, Measurements, and Performances*, 2nd ed. Ganga-Jamuna Press, 2006.
- [3] B. Alavi and K. Pahlavan, "Modeling of the distance error for indoor geolocation," in *Proc. IEEE Int. Conf. Wireless Commun. Networking (WCNC)*, Mar. 2003, pp. 668–672.
- [4] G. Bellusci, G. J. M. Janssen, J. Yan, and C. C. J. M. Tiberius, "Model of the distance and bandwidth dependency of TOA-based UWB ranging error," in *Proc. IEEE Int. Conf. on Ultra-Wideband (ICUWB)*, Hannover, Germany, Sep. 10–12, 2008, pp. 193–196.
- [5] P. J. G. Teunissen, "Nonlinear least-squares," *Manuscripta Geodaetica*, vol. 15, no. 3, pp. 137–150, 1990.
- [6] R. O. Schmidt, "A new approach to geometry of range difference location," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 8, pp. 821–835, 1972.
- [7] S. Bancroft, "An algebraic solution for the GPS equations," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 21, no. 7, pp. 56–59, 1985.
- [8] H. C. Schau and A. Z. Robinson, "Passive source localization employing intersecting spherical surfaces from time-of-arrival differences," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, pp. 1223–1225, Aug. 1987.

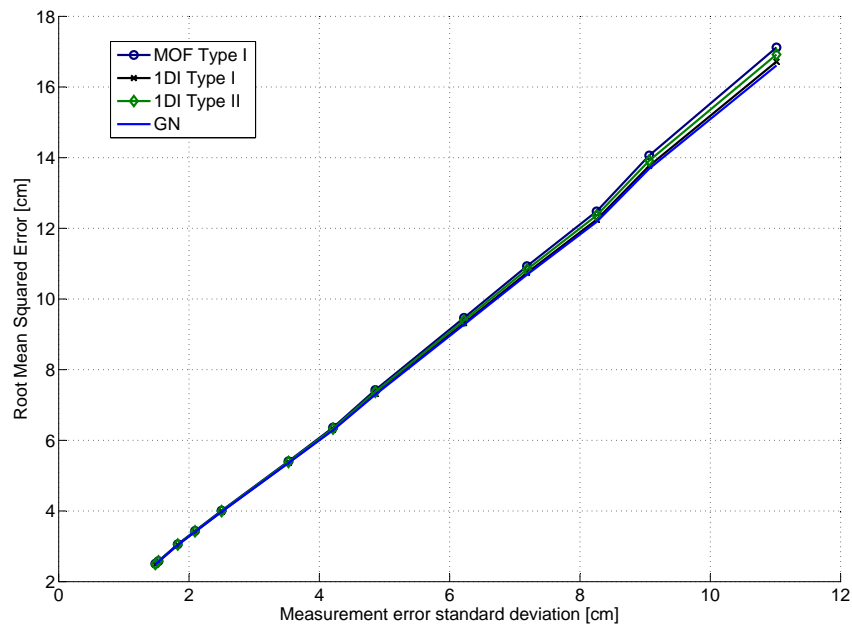


Fig. 13. RMSE of the estimates obtained using the actual TDoA UWB data, for different methods.

- [9] J. O. Smith and J. S. Abel, "The spherical interpolation method of source localization," *IEEE J. Ocean. Eng.*, vol. 12, pp. 246–252, Jan. 1987.
- [10] —, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, pp. 1661–1669, Aug. 1987.
- [11] B. Friedlander, "A passive location algorithm and its accuracy analysis," *IEEE J. Ocean. Eng.*, pp. 234–245, 1987.
- [12] L. Krause, "A direct solution to GPS-type navigation equations," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 23, no. 2, pp. 225–232, 1987.
- [13] B. Fang, "Simple solutions for hyperbolic and related position fixes," *IEEE Trans. Aerosp. Electron. Syst.*, pp. 734–753, 1990.
- [14] J. Chaffee and J. Abel, "On the exact solutions of pseudorange equations," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 4, pp. 1021–1030, 1991.
- [15] K. Ho and Y. Chan, "Solution and performance analysis of geolocation by TDOA," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, pp. 1311–1322, 1993.
- [16] Y. Chan and K. Ho, "A simple and efficient estimator for hyperbolic location," *IEEE Trans. Signal Process.*, vol. 29, pp. 1905–1915, 1994.
- [17] J. L. Leva, "An alternative closed-form solution to the GPS pseudo-range equations," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 4, pp. 1430–1439, 1996.
- [18] A. Kleusberg, "Analytical GPS navigation solution," in *Grafarend EW, Krumm FW. and Schwarze VS. (eds) Geodesy - the Challenge of the 3rd Millennium*, Heidelberg, 2003, pp. 93–96.
- [19] J. Beutel, "Geolocation in a picoradio environment," Master's thesis, UC Berkeley, 1999.

TABLE VII

SUMMARY OF THE REVIEWED RANGE BASED POSITIONING ALGORITHMS. HERE n IS THE DIMENSION OF x IN (7).

Methods	Complexity	Accuracy	Minimum Required Number of Measurements
ID methods	Iterative with full dimensions	Closely approximates strict solution that achieves CRLB	n
SOF methods	Non-iterative with unique solution	Simplified solution	$n + 1$
Type I MOF methods	Non-iterative with two candidate solutions	Simplified solution	n
Type II MOF method	Non-iterative with three candidate solutions	Simplified solution, but better than SOF solutions	$n + 1$
Type I 1DI method	Iterative with 1-D, requires eigenvalue decomposition	Closely approximates rigorous solution that achieves CRLB	n
Type II 1DI method	Iterative with 1-D, without eigenvalue decomposition	Simplified solution, but better accuracy than SOF, MOF methods.	n

- [20] J. J. Caffery, "A new approach to the geometry of TOA location," in *Proc. IEEE Vehic. Technol. Conf. (VTC)*, vol. 4, Boston, MA, Sep. 2000, pp. 1943–1949.
- [21] Z. Li, W. Trappe, Y. Zhang, and B. Nath, "Robust statistical methods for securing wireless localization in sensor networks," in *Proc. IEEE Int. Symposium on Info. Process. in Sensor Networks (IPSN)*, Los Angeles, CA, Apr. 15, 2005, pp. 91–98.
- [22] S. Venkatesh and R. M. Buehrer, "A linear programming approach to NLOS error mitigation in sensor networks," in *Proc. IEEE Int. Symposium on Info. Process. in Sensor Networks (IPSN)*, Nashville, TN, Apr. 2006, pp. 301–308.
- [23] I. Guvenc, S. Gezici, F. Watanabe, and H. Inamura, "Enhancements to linear least squares localization through reference selection and ML estimation," in *Proc. IEEE Int. Conf. Wireless Commun. Networking (WCNC)*, Las Vegas, NV, Mar. 31/Apr. 3, 2008, pp. 284–289.
- [24] J. Yan, C. C. J. M. Tiberius, G. Bellusci, and G. J. M. Janssen, "A novel non-iterative localization solution," in *IEEE GLOBECOM 09*, Honolulu, Nov. 30/Dec. 4, 2009.
- [25] Y. Huang and J. Benesty, "Real-time passive source localization: A practical linear-correction least-squares approach," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 8, pp. 943–955, Nov. 2001.
- [26] K. W. Cheung, H. C. So, W.-K. Ma, and Y. T. Chan, "Least squares algorithms for time-of-arrival-based mobile location," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 1121–1128, Apr. 2004.
- [27] J. Yan and G. Bellusci, "Low-complexity ultra-wideband indoor positioning," in *Proc. ION Global Navigation Satellite Systems (GNSS)*, Savannah, Sep. 22–25, 2009, pp. 1013–1024.

- [28] J. Yan, C. C. J. M. Tiberius, P. J. G. Teunissen, G. Bellusci, and G. J. M. Janssen, "A framework for low complexity least-squares localization with high accuracy," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4836–4847, 2010.
- [29] J. Hightower and G. Borriello, "Location systems for ubiquitous computing," *Computer*, vol. 34, no. 8, pp. 57–66, 2001.
- [30] K. Pahlavan, X. Li, and J. Makela, "Indoor geolocation science and technology," *Communications Magazine, IEEE*, vol. 40, no. 2, pp. 112–118, Feb 2002.
- [31] M. Vossiek, L. Wiebking, P. Gulden, J. Wiegardt, C. Hoffmann, and P. Heide, "Wireless local positioning," *Microwave Magazine, IEEE*, vol. 4, no. 4, pp. 77–86, Dec. 2003.
- [32] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 37, no. 6, pp. 1067–1080, Nov. 2007.
- [33] A. Sayed, A. Tarighat, and N. Khajehnouri, "Network-based wireless location: challenges faced in developing techniques for accurate wireless location information," *Signal Processing Magazine, IEEE*, vol. 22, no. 4, pp. 24–40, July 2005.
- [34] N. Patwari, J. Ash, S. Kyperountas, I. Hero, A.O., R. Moses, and N. Correal, "Locating the nodes: cooperative localization in wireless sensor networks," *Signal Processing Magazine, IEEE*, vol. 22, no. 4, pp. 54–69, July 2005.
- [35] F. Gustafsson and F. Gunnarsson, "Mobile positioning using wireless networks: possibilities and fundamental limitations based on available wireless network measurements," *Signal Processing Magazine, IEEE*, vol. 22, no. 4, pp. 41–53, July 2005.
- [36] G. Sun, J. Chen, W. Guo, and K. Liu, "Signal processing techniques in network-aided positioning: a survey of state-of-the-art positioning designs," *Signal Processing Magazine, IEEE*, vol. 22, no. 4, pp. 12–23, July 2005.
- [37] G. Mao, B. Fidan, and B. D. Anderson, "Wireless sensor network localization techniques," *Computer Networks*, vol. 51, pp. 2529–2553, 2007.
- [38] J. Ash and R. Moses, "Self-calibration of sensor networks," in *Handbook on Array Processing and Sensor Networks*, S. Haykin and K. Liu, Eds. IEEE-Wiley, 2010.
- [39] H. Hashemi, "The indoor radio propagation channel," *Proceedings of the IEEE*, vol. 81, no. 7, pp. 943–968, Jul 1993.
- [40] T. Rappaport, *Wireless Communications: Principles and Practice*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
- [41] A. Muqaibel, A. Safaai-Jazi, A. Attiya, B. Woerner, and S. Riad, "Path-loss and time dispersion parameters for indoor UWB propagation," *Wireless Communications, IEEE Transactions on*, vol. 5, no. 3, pp. 550–559, March 2006.
- [42] J. Choi, N.-G. Kang, Y.-S. Sung, and S.-C. Kim, "Empirical ultra wide band path loss model in office environments," in *Vehicular Technology Conference, 2006. VTC 2006-Spring. IEEE 63rd*, vol. 4, May 2006, pp. 1956–1960.
- [43] N. Alsindi, B. Alavi, and K. Pahlavan, "Empirical pathloss model for indoor geolocation using UWB measurements," *Electronics Letters*, vol. 43, no. 7, pp. 370–372, 2007.
- [44] G. Bellusci, G. J. M. Janssen, J. Yan, and C. C. J. M. Tiberius, "Low complexity ultra-wideband ranging in indoor multipath environments," in *Proc. IEEE/ION Position Location and Navigation Symposium (PLANS)*, Monterey, May 6–8, 2008.
- [45] L. Cong and W. Zhuang, "Nonline-of-sight error mitigation in mobile location," *Wireless Communications, IEEE Transactions on*, vol. 4, no. 2, pp. 560–573, March 2005.
- [46] K. Yu and Y. Guo, "Statistical NLOS identification based on AOA, TOA, and signal strength," *Vehicular Technology, IEEE Transactions on*, vol. 58, no. 1, pp. 274–286, Jan. 2009.
- [47] P. J. G. Teunissen, "Nonlinear inversion of geodetic and geophysical data: diagnosing nonlinearity," *Lecture notes in Earth Science*, no. 29, pp. 241–264, 1990.
- [48] Y. Qi, H. Suda, and H. Kobayashi, "On time-of-arrival positioning in a multipath environment," in *Vehicular Technology Conference, 2004. VTC2004-Fall. 2004 IEEE 60th*, vol. 5, Sept. 2004, pp. 3540–3544.

- [49] P. J. G. Teunissen, "First and second moments of non-linear least-squares estimators," *Bulletin Geodesique*, vol. 63, no. 3, pp. 253–262, 1989.
- [50] B. Efron and R. Tibshirani, "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy," *Statistical Science*, vol. 1, no. 1, pp. 54–77, 1986.
- [51] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of Applied Mathematics*, vol. 2, no. 2, pp. 164–168, 1944.
- [52] G. H. Golub and C. Van Loan, "An analysis of the total least squares problem," Ithaca, NY, USA, Tech. Rep., 1980.
- [53] G. H. Golub and C. F. van Loan, *Matrix Computations*. Baltimore, Md: Johns Hopkins University Press, 1989.
- [54] J. O. Smith and J. S. Abel, "The spherical interpolation method for closed-form passive source localization using range difference measurements," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 12, Apr. 1987, pp. 471–474.
- [55] Z. Irahauten, G. Bellusci, G. J. M. Janssen, and C. C. J. M. Tiberius, "Investigation of UWB ranging in dense indoor multipath environments," in *Proc. IEEE Int. Conf. Commun. Systems (ICCS)*, Singapore, Oct. 2006, pp. 1–5.
- [56] FCC. (2002) Revision of part 15 of the commission's rules regarding ultra-wideband transmission systems. [Online]. Available: <http://www.fcc.gov/oet/dockets/et98-153/>
- [57] J. Yan, C. C. J. M. Tiberius, G. Bellusci, and G. J. M. Janssen, "Feasibility of Gauss-Newton method for indoor positioning," in *Proc. IEEE/ION Position Location and Navigation Symposium (PLANS)*, Monterey, May 6–8, 2008.