On the Maximum Likelihood Approach for Source and Network Localization

Giuseppe Destino, *Student Member, IEEE*, and Giuseppe Abreu, *Senior Member, IEEE*

***Abstract*—We consider the source and network localization problems, seeking to strengthen the relationship between the *Weighted-Least-Square* (WLS) and the *Maximum-Likelihood* (ML) solutions of these problems. To this end, we design an optimization algorithm for source and network localization under the principle that: a) the WLS and the ML objectives should be the same; and b) the solution of the ML-WLS objective does not depend on any information besides the set of given distance measurements (observations). The proposed *Range-Global Distance Continuation* (R-GDC) algorithm solves the localization problems via iterative minimizations of smoothed variations of the WLS objective, each obtained by convolution with a Gaussian kernel of progressively smaller variances. Since the last (not smoothed) WLS objective derives directly from the ML formulation of the localization problem, and the R-GDC requires no initial estimate to minimize it, final result is maximum-likelihood approach to source and network localization problems. The performance of the R-GDC method is compared to that of state-of-the-art techniques such as semidefinite programming (SDP), nonlinear Newton least squares (NLS), and the Stress-of-a-MAjorizing-Complex-Objective-Function (SMACOF) algorithms, as well as to the Cramér-Rao Lower Bound (CRLB). The comparison reveals that the solutions obtained with the R-GDC algorithm is insensitive to initial estimates and provides a localization error that closely approaches that of the corresponding fundamental bounds. The R-GDC is also found to achieve a complexity order comparable to that of the SMACOF, which is known for its efficiency.**

***Index Terms*—Optimization methods, smoothing methods, wireless sensor networks.**

I. Introduction

LOCALIZATION has recently emerged as a major research topic in wireless communications, as the quest for ways to develop pervasive networks and enable ubiquitous services [2], [3]. For instance, in home and office environments, location information is used to identify the location of a device (personal computer, mobile phone, ID-card, etc.) and allow services such as smart-access to facilities and optimized file transfers. Personalized advertisement services and indoor navigation systems using context and location data in public spaces such as shopping malls, fairs and airports are already available; and in industrial and hospital scenarios, commercial positioning services are used to track assets and persons, study and improve logistics and, ultimately, minimize economical losses [4]–[6]. In addition to this existing usages, further applications of network localization are expected to arise in the next generation of mobile telephony using femtocells and featuring direct mobile-to-mobile communications, as well as emerging areas such Smart Grids.

In its most comprehensive form, location-awareness refers to knowledge on the location of all nodes in the network (i.e., *network localization*); and in its simpler form, to knowledge on the location of a specific node (i.e., *source localization*). While in outdoor and open space environments, pseudoranges obtained from satellite communications can be utilized for the design of localization systems, e.g., global positioning system (GPS), COMPASS, and GALILEO [7], in indoor scenarios, where the attenuation of satellite links is very strong, the alternative is to perform localization relying on the radio signals of short and medium range communication systems, e.g., ultrawideband (UWB), Wi-Fi, and RF-ID. In such cases, physical parameters that can be exploited include direction-of-arrival (DoA) [8], [9], received signal strength (RSS) [10] and time-of-arrival (ToA) [11].

Among these approaches, distance-based localization with ToA-based range estimates is the most dominant [12]–[22]. In fact, the terms ToA-based and distance-based localization are commonly used interchangeably. This choice is justified by theoretical evidence that ToA-based methods are inherently more accurate than the alternatives [13], [23].

Indeed, it has been observed that ToA-errors are often independent identically distributed (i.i.d) zero-mean Gaussian variables, which yields the advantage that optimal performances is achievable through maximum likelihood (ML) estimation. Typically, ML solutions are obtained as the global minimum of a weighted least square (WLS) objective function which is directly derived from the likelihood-function of the problem [12]–[22], [24]–[35]. Thus, the question of fundamental relevance to the WLS source and network localization problems is: *how the ML solution associated with a set of ranging measurements can be achieved?*

Notice that while the uniqueness of the solution is a general *asymptotic* property of ML estimation [36], in the context of localization a stronger condition is required for the sake of *localizability* [16]. Specifically, in the localization problem the ML solution associated with a *given* set of ranging measurements is implicitly required to be unique. Under this notion, a WLS localization algorithm must satisfy the two following conditions in order to be a ML estimator: a) the global minimum of the WLS cost-function must be directly related to the global optimum of the ML objective; b) the solution of the WLS algorithm must not depend on an initialization.

Manuscript received March 02, 2011; revised June 15, 2011; accepted June 15, 2011. Date of publication July 07, 2011; date of current version September 14, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Maria S. Greco. The material in this paper was presented at Int. Conf. Commun. (ICC 2009), SyCoLo Workshop.

The authors are with the Centre for Wireless Communications, University of Oulu, 90014, Oulu, Finland (e-mail: giuseppe.destino@ee.oulu.fi; giuseppe.abreu@ieee.org).

Digital Object Identifier 10.1109/TSP.2011.2161302

Several algorithms have been proposed which aim at finding ML solutions for the ML-WLS localization problem [22], [24]–[35]. To the best of our knowledge, however, with the exception of simulated annealing [24] which is prohibitively complex for large-scale networks, all existing methods violate either or both of the conditions outlined above.

For instance, the classic nonlinear-least square (NLS) algorithm, both centralized [25], [26] and distributed [22], [27], as well as the SMACOF [28], [29] algorithm, although popular due to their relatively low complexity and ease of implementation, are known to be sensitive to initial estimates. The initialization of these algorithms has received relatively little attention in the literature, where only a few alternatives based on the linearized-least square (LLS) and the multidimensional scaling (MDS) methods can be found [21], [26], [37], [38]. In small-scale networks these strategies are sufficient to allow accurate positioning with local optimization methods as the SMACOF and NLS algorithms. In large-scale networks, however, where scarce connectivity is a major problem to contend with, the LLS and the MDS initialization techniques are no longer reliable, transferring the problem to all initialization-sensitive approaches [28].

In turn, localization algorithms based on a global optimization such as the semidefinite programming (SDP) with convex constraints [16], [30], [31], as well as those relying on square-ranging [32], [33] operate over formulations which are not, by design, strictly derived from the likelihood function. In fact the strategy of squaring distances has been shown [39], [40] to amplify noise and introduce correlation between noise and variables, generally degrading localization accuracy.

Unlike these methods, the global distance continuation (GDC) approach proposed in [32] can be redesigned not to violate the aforementioned conditions. In this method, the solution of the localization problem is found by minimizing smoothed WLS-objectives, each obtained from the convolution of the original WLS-objective with a different Gaussian kernel. Since the first smoothed WLS-objective can always be made convex, this technique eliminates any sensitivity to initial estimates. However, in the form developed in [32]—hereafter referred to as SR-GDC—a suboptimum (non-ML) formulation of the WLS problem using *squared distances* was utilized.

In this paper we return to the principle of the GDC method for ToA-based localization problem, but without reformulating the WLS-objective with squared distances. Specifically, the algorithm proposed in this article, which will be referred to as Range-Global Distance Continuation (R-GDC), is a method to obtain ML-solutions for both the source and network localization problems (like the NLS and the SMACOF), with insensitivity to initial estimates (like the convex SDP and the SR-GDC), and at very low complexity (like the SMACOF).

In the process of deriving the R-GDC algorithm, we offer additional contributions including close-form expression of the Gaussian transform of the WLS-objective (or the Frobenious norm in general), and insight on important characteristics of this cost-function which are employed to inform the choice of the smoothing parameters. As a result, the proposed R-GDC algorithm is of low-complexity, robust to local minima, and enables accurate positioning in large and sparsely connected wireless networks, such as sensor networks. In addition to the aforementioned contributions, the fundamental limits of ML source localization is revised, in order to account for both large and small errors due to noise and geometrical ambiguities, recalling the alternative Stochastic Bound (SB) given in [41].

The reminder of the article is organized as follows. In Section II the ML formulation of the localization problem and state-of-the-art solutions are summarized. In Section III the derivation of all functions and tools related to the design of the proposed R-GDC algorithm are described, including: a) the closed-form expressions of the smoothed WLS-objective, its gradient-vector and Hessian matrix; b) a closed-form expression for the initial smoothing parameter $\lambda^{(0)}$; and c) formulas to calculate the *landmark* λ's for subsequent iterations (in the source-localization case). In Section IV-A, the fundamental limits of the ML localization are revised and an explicit expression of the SB for the source localization case is given. Finally, in Sections IV-B and V simulation results and conclusions are provided, respectively.

II. Maximum-Likelihood Formulation of the Localization Problem

Consider a network of N nodes deployed in a space of dimension η. Let $\mathbf{p}_i \in \mathbb{R}^\eta$ be a row-vector whose elements are the Euclidean coordinates of the ith node. The Euclidean distance between the ith and the jth node is defined as

$$d_{ij} \triangleq \|\mathbf{p}_i - \mathbf{p}_j\|_F \tag{1}$$

where $\|\cdot\|_F$ is the Frobenius norm.

Throughout the article, the estimate of a variate or parameter will be denoted by a "^", while an observation (measurement) of such variable/parameter will be denoted by a "~".

A node whose location is known *a priori* is referred to as an *anchor*, while a *target* is a node whose position is yet to be determined. When an index i refers to an anchor we have $\hat{\mathbf{p}}_i = \mathbf{p}_i \equiv \mathbf{a}_i$. We shall distinguish between the N_A anchors and the N_T targets by denoting their coordinate vectors with the symbols $\mathbf{a}$ and $\mathbf{z}$, respectively.

A measurement (ranging) of the distance d_{ij} between and ith and jth nodes is obtained[1] as

$$\tilde{d}_{ij} = \begin{cases} d_{ij} + n_{ij}, & \text{if either } \mathbf{p}_i = \mathbf{z}_i \text{ or } \mathbf{p}_j = \mathbf{z}_j, \\ d_{ij}, & \text{if both } \mathbf{p}_i = \mathbf{a}_i \text{ and } \mathbf{p}_j = \mathbf{a}_j \end{cases} \tag{2}$$

where n_{ij} is a statistical independent Gaussian random variable with zero mean and variance σ_{ij}^2.

Stacking the N_T coordinate row-vectors $\mathbf{z}_i$ into an $N_T \times \eta$ coordinate matrix $\mathbf{Z} \triangleq [\mathbf{z}_1; \cdots; \mathbf{z}_{N_T}]$, the network localization problem amounts to finding the estimate $\hat{\mathbf{Z}}$ that minimizes

$$\xi = \|\hat{\mathbf{Z}} - \mathbf{Z}\|_F .. \tag{3}$$

[1]In principle independent measurements of $d_{ij} \equiv d_{ji}$ could be obtained, such that $\tilde{d}_{ij}$ and $\tilde{d}_{ji}$ would be available, affected by i.i.d noise samples n_{ij} and n_{ji}, respectively. However, in the ToA case, two-way or three-way ranging is typically performed due to advantages in terms of synchronization and accuracy [42]. Therefore we consider, without loss of generality, that a single sample $\tilde{d}_{ij}$ is associated with each pair of nodes (i, j).

However, since $\mathbf{Z}$ is not measurable, (3) cannot be utilized as an optimization criterion. The alternative is to maximize the likelihood function of $\hat{\mathbf{Z}}$,

$$p(\hat{\mathbf{Z}}|\tilde{\mathbf{D}}) \triangleq \prod_{ij\in\mathcal{H}} \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp\left(-\frac{\left(\tilde{d}_{ij}-\hat{d}_{ij}\right)^2}{2\sigma_{ij}^2}\right) \quad (4)$$

where $\tilde{\mathbf{D}} \triangleq [\tilde{d}_{ij}]$ is a sample of the Euclidean Distance Matrix (EDM) and $\mathcal{H}$ indicates the set of ij-indexes corresponding to the anchor-to-target and target-to-target connected links.

This approach is the well-known ML formulation[2] of the localization problem, which can be rewritten as a WLS minimization problem [13], [21], [27], [31]

$$\min_{\hat{\mathbf{Z}}\in\mathbb{R}^{N_T\times\eta}} f_{\mathrm{R}}(\hat{\mathbf{Z}}) \quad (5)$$

with

$$\begin{aligned} f_{\mathrm{R}}(\hat{\mathbf{Z}}) &\triangleq \sum_{ij\in\mathcal{H}} w_{ij}(\tilde{d}_{ij}-\hat{d}_{ij})^2 \\ &= \sum_{ij\in\mathcal{H}} w_{ij}(\tilde{d}_{ij}-\|\mathbf{a}_i-\hat{\mathbf{z}}_j\|_F)^2 \\ &\quad + \sum_{ij\in\mathcal{H}} w_{ij}(\tilde{d}_{ij}-\|\hat{\mathbf{z}}_i-\hat{\mathbf{z}}_j\|_F)^2 \end{aligned} \quad (6)$$

where w_{ij} is a weight [29], [43] related to the "concern" [44] over the term $(\tilde{d}_{ij}-\hat{d}_{ij})$, and the explicit functional relationship between $f_{\mathrm{R}}(\hat{\mathbf{Z}})$ and the unknown variables $\hat{\mathbf{z}}$ is highlighted by separating the terms related to the anchor-to-target and target-to-target links.

The optimal weighing strategy in the ML-sense, is $w_{ij}=\frac{1}{\sigma_{ij}^2}$, as shown in [13]. This, however, requires *a priori* knowledge on the statistics of the ranging errors (specifically the exact value of σ_{ij}^2), which is not always available. Alternative and more practical weighing mechanisms have been proposed, for instance in [29] and [43]. Specifically, in [29] w_{ij} is computed using a nonlinear regression model that emphasizes short distances on long ones. In [43], w_{ij} is formed of a *dispersion* component, which captures the effect of noise while maximizing the diversity of the information, and a *penalty* component, which penalizes the assumption of bias-free measurements on the basis of a hypothesis testing.

Notice also that the ML-WLS objective is not a convex function [45]. Knowledge of the location of non convex regions of the objective can therefore be useful, especially if a relationship between those regions and the set of distances $\{d_{ij}\}$ can be established. This is achieved in what follows.

Lemma L1 (Concavity of $f_R(\hat{\mathbf{Z}})$ Around Anchors): Consider the set of ranging measurement $\{\tilde{d}_{it}\}$ from a number N_A of anchors $\mathbf{a}_i$ to a single target, where $t \triangleq N_A+1$, and define the set $B_i \triangleq \{\hat{\mathbf{z}} \mid \|\mathbf{a}_i-\hat{\mathbf{z}}\| \le \varrho\}$. Then

$$\exists \varrho \mid \nabla^2_{\hat{\mathbf{z}}} f_{\mathrm{R}}(\hat{\mathbf{z}}) \preceq 0,\ \forall \hat{\mathbf{z}} \in B_i. \quad (7)$$

[2]Notice that the ML approach to localization is only meaningful when the network is localizable [16], which implies that the topology is not ambiguous or, equivalently that the maximum of the asymptotic likelihood function is identifiable [36].

Proof: In the case of single target, the Hessian of $f_{\mathrm{R}}(\hat{\mathbf{z}})$ is

$$\nabla^2_{\hat{\mathbf{z}}} f_{\mathrm{R}}(\hat{\mathbf{z}}) = 2\left(\sum_{i=1}^{N_A} w_{it}\frac{\hat{d}_{it}-\tilde{d}_{it}}{\hat{d}_{it}}\right)\cdot\mathbf{I}_\eta + 2\sum_{i=1}^{N_A} w_{it}\frac{\tilde{d}_{it}}{\hat{d}_{it}}\mathbf{v}_i\mathbf{v}_i^T \quad (8)$$

where $\hat{\mathbf{z}} \in B_i$, $\mathbf{v}_i \triangleq \frac{(\mathbf{a}_i-\hat{\mathbf{z}})}{\hat{d}_{it}}$ and $t = N_A+1$.

Isolating the ith term out of the sums yields,

$$\begin{aligned} \frac{\nabla^2_{\hat{\mathbf{z}}} f_{\mathrm{R}}(\hat{\mathbf{z}})}{2} &= c_{it}\cdot(\boldsymbol{\Upsilon}_i-\mathbf{I}_\eta) \\ &+ \underbrace{\left(w_{it}+\sum_{j\neq i}^{N_A} w_{jt}\left(1-\frac{\tilde{d}_{jt}}{\hat{d}_{jt}}\right)\right)\cdot\mathbf{I}_\eta + \sum_{j\neq i}^{N_A} w_{jt}\frac{\tilde{d}_{jt}}{\hat{d}_{jt}}\mathbf{v}_j\mathbf{v}_j^T}_{\mathbf{R}_i}, \end{aligned} \quad (9)$$

with $c_{it} \triangleq \frac{w_{it}\cdot\tilde{d}_{it}}{\|\mathbf{a}_i-\hat{\mathbf{z}}\|}$ and

$$\boldsymbol{\Upsilon}_i \triangleq \begin{bmatrix} \cos^2\theta_i & \sin\theta_i\cos\theta_i \\ \sin\theta_i\cos\theta_i & \sin^2\theta_i \end{bmatrix} \quad (10)$$

where θ_i is the angle between the vectors centered at $\hat{\mathbf{z}}$ pointing towards $\mathbf{a}_i$ and $\hat{\mathbf{z}}+[1,0]$, respectively.

Notice that the eigenvalues of the matrix difference $(\boldsymbol{\Upsilon}_i-\mathbf{I}_\eta) \preceq 0$ are -1 and 0), and that the coefficient c_{it} is positive and grows unboundedly with decreasing ϱ. In contrast, the matrix $\mathbf{R}_i$ is finite for all $\hat{\mathbf{z}} \in B_i$ as long as there is no $\mathbf{a}_j \in B_i$, which is ensured by a sufficiently small ϱ. In conclusion, the eigenspectrum of $\nabla^2_{\hat{\mathbf{z}}} f_{\mathrm{R}}(\hat{\mathbf{z}})$ approaches that of $c_{it}\cdot(\boldsymbol{\Upsilon}_i-\mathbf{I}_\eta)$ asymptotically (with decreasing ϱ). □

Lemma L1 establishes that the ML-WLS objective has concavity regions in the vicinity of the anchors *even* if all d_{ij} are error-free. Further nonconvexity is certain to arise under the influence of noise, the number of targets, their location and their connectivity, so that the number of minima in fact grows quickly in a network localization scenario. In the sequel we revise three different state-of-the-art ML-WLS localization algorithms with aim at illustrating how typical methods attempt to circumvent the problem of optimizing over the nonconvex ML objective given in (6).

A. Semi-Definite Programming Algorithm

The first state-of-the-art localization technique to be discussed in this section addresses the network localization problem (localization of multiple-devices in a mesh network). The algorithm is based on an SDP optimization technique and can be understood as a variation of ML estimation technique, in which the original multivariate optimization problem is casted as a matrix-proximity problem.

We refer specifically to the scheme proposed in [31], where Biswas et. al start from the ML formulation of the localization problem and cast it as a matrix proximity problem, in which the objective is to find the EDM-estimate $\hat{\mathbf{D}} \triangleq [\hat{d}_{ij}]$ of rank $\eta+2$ closest to the observed EDM-sample $\tilde{\mathbf{D}}$, in the Frobenius norm sense. Mathematically,

$$\begin{aligned} \min_{\hat{\mathbf{D}}} \quad & \left\|\mathbf{W}\circ(\tilde{\mathbf{D}}-\hat{\mathbf{D}})\right\|_F^2, \\ \text{s.t.} \quad & \mathrm{rank}(\mathbf{V}^T\hat{\mathbf{D}}^{\circ 2}\mathbf{V}) = \eta \\ & \hat{\mathbf{D}}^2 \in \mathbb{EDM} \end{aligned} \quad (11)$$

where $\mathbb{EDM}$ is the EDM space, $\mathbf{W} \in \mathbb{R}^{N\times N}$ is a weight matrix, and $\mathbf{V} \in \mathbb{R}^{N\times N}$ is defined in [45].

Unfortunately the rank-constraint inserted in(11)—required to ensure that the solution is confined to the dimension of the space where the nodes are located—makes this optimization difficult to solve. In order to circumvent this difficulty, the rank constraint is relaxed, which allows for the following "convexized" reformulation of SDP-based localization [31]

$$\begin{aligned}\min_{\hat{\mathbf{K}},\{\hat{\mathbf{B}}_{ij}\}} \quad & \sum_{ij\in\mathcal{H}} w_{ij}\,\varepsilon_{ij} \\ \text{s.t.} \quad & [-\tilde{d}_{ij}\ 1]\,\hat{\mathbf{B}}_{ij}\,[-\tilde{d}_{ij}\ 1]^T = \varepsilon_{ij},\ \forall ij \\ & [\mathbf{0}_\eta\ \mathbf{e}_i - \mathbf{e}_j]\,\hat{\mathbf{K}}\,[\mathbf{0}_\eta\ \mathbf{e}_i - \mathbf{e}_j]^T = \nu_{ij},\ i,j \geq N_A \\ & [\mathbf{a}_i\ -\mathbf{e}_j]\,\hat{\mathbf{K}}\,[\mathbf{a}_i\ -\mathbf{e}_j]^T = \nu_{ij},\ i \leq N_A,\ \forall j \\ & \hat{\mathbf{B}}_{ij} \triangleq \begin{bmatrix} 1 & b_{ij} \\ b_{ij} & \nu_{ij} \end{bmatrix} \succeq 0 \\ & \hat{\mathbf{K}} \triangleq \begin{bmatrix} \mathbf{I}_\eta & \hat{\mathbf{Z}}^T \\ \hat{\mathbf{Z}} & \hat{\mathbf{Y}} \end{bmatrix} \succeq 0 \end{aligned} \tag{12}$$

where $\mathbf{0}_\eta$ is a vector of zeros and $\mathbf{e}_i \in \mathbb{R}^{N_T}$ the only nonzero element is a 1 at the ith element.

The convexized SDP formulation of (12) can be optimally solved using standard convex SDP optimization software, such as SDPA, CSDP, SDPT3, SeDuMi[3].

Unlike the optimization problem given in (11), however, the solution of (12) is no longer related to the ML formulation described in (5), thus the aforementioned SDP method is not an exact ML estimation technique. Moreover, the computational-complexity of the SDP approach quickly grows with the number of constraints (proportional to the number of links) as well as with the level of incompleteness which increases the discrepancy between EDM-samples and the EDM space. Indeed, the cost of this algorithm was shown in [31] to reach the order $\mathcal{O}(N_T^6)$, which makes it prohibitive in large-scale network localization problems.

B. Square-Range Least-Square Algorithm (Source Loc)

The second state-of-the-art localization algorithm of interest is the exact optimization method proposed by Beck *et al.* in [33]. This technique tackles the localization problem of a single target (aka, source localization) and relies on a variation of the ML localization problem given in (5), in which range measurements are squared. Specifically, the problem solved in [33] is

$$\min_{\hat{\mathbf{z}}\in\mathbb{R}^\eta} \sum_{i=1}^{N_A} \left(\tilde{d}_i^2 - \hat{d}_i^2\right)^2 \tag{13}$$

where the index j is omitted since $j = 1$, and $w_i = 1,\ \forall i$ since $\sigma_i = \sigma,\ \forall i$.

Again, since the range measurements $\tilde{d}_i$ and their estimates $\hat{d}_i$ are *squared*, this problem differs from the original ML source localization problem. Nevertheless, in [33] it was shown that the exact solution of (13) can be derived as follows. First, compute the matrices $\mathbf{Q}$, $\mathbf{H}$, $\mathbf{b}$ and $\mathbf{c}$ as

$$\mathbf{Q} \triangleq \begin{bmatrix} -2\mathbf{a}_1^T & 1 \\ \vdots & \vdots \\ -2\mathbf{a}_{N_A}^T & 1 \end{bmatrix}, \quad \mathbf{b} \triangleq \begin{bmatrix} \tilde{d}_1^2 - \|\mathbf{a}_1\|_F^2 \\ \vdots \\ \tilde{d}_{N_A}^2 - \|\mathbf{a}_{N_A}\|_F^2 \end{bmatrix} \tag{14}$$

$$\mathbf{H} \triangleq \begin{bmatrix} \mathbf{I}_\eta & \mathbf{0}_\eta \\ \mathbf{0}_1 & 0 \end{bmatrix}, \quad \mathbf{c} \triangleq \begin{bmatrix} \mathbf{0}_\eta \\ -0.5 \end{bmatrix}. \tag{15}$$

Next, compute α^* as the root of

$$\hat{\mathbf{r}}(\alpha)^T \cdot \mathbf{H} \cdot \hat{\mathbf{r}}(\alpha) + 2\mathbf{c}^T \cdot \hat{\mathbf{r}}(\alpha)^T = 0,\ \text{with}\ \alpha \in \left(-\frac{1}{\alpha_1}, \infty\right) \tag{16}$$

where α_1 is the maximum of the generalized eigenvalues [46] of $\mathbf{Q}\mathbf{Q}^T$ and

$$\hat{\mathbf{r}}(\alpha) \triangleq \left(\mathbf{Q}\cdot\mathbf{Q}^T + \alpha\mathbf{H}\right)^{-1}\left(\mathbf{Q}^T\mathbf{b} - \alpha\mathbf{c}\right). \tag{17}$$

Finally, the minimizer of the objective in (13) is given by the η first components of $\hat{\mathbf{r}}(\alpha^*) \in \mathbb{R}^{\eta+1}$.

In spite of the fact that the Beck *et al.*'s solution is exact, the formulation of the problem is suboptimal to the ML formulation given by (5), since squaring the distances amplifies errors on the measurements, fundamentally increasing the achievable mean square location error [39].

C. Square-Range Global Distance Continuation Algorithm

A low-complexity algorithm relying on the same distance-squaring approach described above was proposed in [32] to address the localization of multiple targets (aka, network localization). The approach consists of solving

$$\min_{\hat{\mathbf{Z}}\in\mathbb{R}^{N_T\times\eta}} f_{SR}(\hat{\mathbf{Z}}). \tag{18}$$

with

$$f_{SR}(\hat{\mathbf{Z}}) \triangleq \sum_{ij\in\mathcal{H}} w_{ij}\left(\tilde{d}_{ij}^2 - \|\hat{\mathbf{p}}_i - \hat{\mathbf{p}}_j\|_F^2\right)^2 \tag{19}$$

using a method referred to as GDC.

The GDC method can be summarized in three steps: *smoothing*, *minimization* and *continuation*.

In the smoothing step the entire objective is transformed into a function with a higher degree of differentiability (smoothed), obtained by means of a convolution of with a Gaussian kernel. In the minimization step each of these smoothed functions is minimized using an efficient steepest-decent algorithm [47]. Finally, the continuation refers to the process of tracing the global minimum, which in practice is typically performed by initializing the minimization of the next smoothed objective with the latest solution. Mathematically, the SR-GDC relies on the following Definitions and Theorem.

Definition D1 (Gaussian Kernel): The **Gaussian kernel** is defined as

$$g(u,\lambda) \triangleq e^{\frac{-u^2}{\lambda^2}}. \tag{20}$$

Definition D2 (Smooth Function): Consider a variable $\boldsymbol{z} \in \mathbb{R}^n$ and a multivariate continuous function $f : \mathbb{R}^n \to \mathbb{R}$.

[3]SeDuMi runs in Matlab© and uses the Self-Dual method for solving general convex optimization problems, etc.

The smoothed version (or **smoothed function**) of $f(\boldsymbol{z})$, denoted $\langle f\rangle_{\lambda}(\boldsymbol{z})$, is defined as

$$\langle f\rangle_{\lambda}(\boldsymbol{z}) \triangleq \frac{1}{\pi^{\frac{n}{2}}\lambda^n}\int_{\mathbb{R}^n} f(\boldsymbol{u})\cdot\exp\left(-\frac{\|\boldsymbol{z}-\boldsymbol{u}\|_F^2}{\lambda^2}\right)\mathrm{d}\boldsymbol{u} \quad (21)$$

where $\lambda \in \mathbb{R}^+$ is a parameter that controls the degree of smoothing ($\lambda \gg 0$ strong smoothing).

Theorem T1 (Continuation Principle): Let $\mathcal{L} = \{\lambda^{(k)}\}$ with $\{1 \leq k \leq K\}$ be any decreasing sequence of λ's converging to zero, i.e., $\lambda^{(K)} = 0$. If $\boldsymbol{z}^{(k)}$ is a global minimizer of $\langle f\rangle_{\lambda^{(k)}}(\boldsymbol{z})$ and $\{\boldsymbol{z}^{(k)}\}$ converges to $\boldsymbol{z}^*$, then $\boldsymbol{z}^*$ is the global minimum of $f(\boldsymbol{z})$.

Proof: See [32]. □

Referring to the above and the square range WLS objective defined in equation(19), the kth iteration of the SR-GDC technique proposed in [32] can be summarized as follows:

$$\hat{\mathbf{Z}}^{(k)} = \min_{\hat{\mathbf{Z}}\in\mathbb{R}^{N_T\times\eta}} \langle f_{SR}\rangle_{\lambda^{(k)}}(\hat{\mathbf{Z}}),\ 1\leq k\leq K \quad (22)$$

with

$$\langle f_{SR}\rangle_{\lambda}(\hat{\mathbf{Z}}) = f_{SR}(\hat{\mathbf{Z}}) + \sum_{ij\in\mathcal{H}}\left(2w_{ij}\left(3+(\eta-1)\lambda^2\hat{d}_{ij}^2\right)\right)+C \quad (23)$$

where the sum-form given above results from the fact that $f_{SR}(\hat{\mathbf{Z}})$ is a decomposable function as shown in [32], and the constant C is given by

$$C \triangleq \sum_{ij\in\mathcal{H}} w_{ij}\left(\eta(\eta+2)\lambda^4 + 2\eta\tilde{d}_{ij}^2\lambda^2\right). \quad (24)$$

In [32], the authors realized that thanks to the square ranging approach, convenient properties of the Gaussian transform can be invoked in order to derive the simple form of the smoothed objectives $\langle f_{SR}\rangle_{\lambda}(\hat{\mathbf{Z}})$ shown in (23), which in turn ensure the low-complexity of the SR-GDC algorithm.

As previously discussed, however, this strategy also eliminates the direct relationship with the ML-WLS objective, and therefore the SR-GDC algorithm is fundamentally not an ML algorithm as it violates the first condition for ML-compliance (see Section I).

Nevertheless, insensitivity to initial estimates—which is the second criterion for ML-compliance—is an intrinsic feature of the GDC method, in so far as the convexity of the first smoothed version of any objective can be ensured by the choice of a sufficiently large $\lambda^{(1)}$.

The GDC-based localization algorithm is therefore the subject of the next section and the main contribution of this article.

III. Range-Global Distance Continuation ALG

In order to apply the GDC method to the ML-WLS network localization problem, two main challenges need be met:

- A closed-form expression for the smoothed range-WLS objective $\langle f_{\mathrm{R}}\rangle_{\lambda}(\hat{\mathbf{Z}})$, along with its Gradient and Hessian, needs be derived;
- A least upper bound, or *supremum* on the initial smoothing parameter $\lambda^{(1)}$ needs be found.

The first challenge goes to the complexity of the algorithm, while the second goes to the fundamental issue of establishing compliance with the initial-estimate insensitivity implied by the ML approach. In addition to the above, the practical implementation of the Range-Global Distance Continuation (R-GDC) can also benefit from a set of landmark smoothing parameters $\{\lambda^{(1)},\ldots,\lambda^{(K)}\}$ to improve the efficiency of the continuation step. These goals are addressed below.

A. Smoothed ML-Function, Gradient and Hessian

For the sake of simplicity, it will be assumed hereafter that $\eta = 2$ (two-dimensional networks). We emphasize, however, that the results can be easily extended to the 3D case.

Given the restriction to the two-dimensional space, it will prove convenient to indicate the cartesian axis by the letters x and y, such that for example $\mathbf{p} = [p_x, p_y]$.

Theorem T2 (Smoothed ML Objective): The smoothed version of the ML-WLS objective defined in (6) is given by

$$\langle f_{\mathrm{R}}\rangle_{\lambda}(\hat{\mathbf{Z}}) = \sum_{ij\in\mathcal{H}} w_{ij}\left(\lambda^2+\tilde{d}_{ij}^2+\hat{d}_{ij}^2-\lambda\sqrt{\pi}\tilde{d}_{ij}\,{}_1F_1\left(\tfrac{3}{2};1;\tfrac{\tilde{d}_{ij}^2}{\lambda^2}\right)\exp\left(\tfrac{-\hat{d}_{ij}^2}{\lambda^2}\right)\right) \quad (25)$$

where $\Gamma(a)$ is the gamma function and ${}_1F_1(a;b;c)$ is the confluent hypergeometric function [48].

Proof: Consider the objective function in (6), and apply equation (21) to obtain

$$\begin{aligned}\langle f_{\mathrm{R}}\rangle_{\lambda}(\hat{\mathbf{Z}}) &= \tfrac{1}{\pi}\int_{\mathbb{R}^\eta}\sum_{ij\in\mathcal{H}} w_{ij}\left(\tilde{d}_{ij}-\|\hat{\mathbf{p}}_i-\hat{\mathbf{p}}_j+\lambda\mathbf{u}\|_F\right)^2\exp(-\|\mathbf{u}\|_F^2)\mathrm{d}\mathbf{u}\\ &= \tfrac{1}{\pi}\sum_{ij\in\mathcal{H}} w_{ij}(\mathcal{I}_1+\mathcal{I}_2-2\tilde{d}_{ij}\mathcal{I}_3),\end{aligned} \quad (26)$$

where $\mathbf{u} \triangleq [u_x, u_y]$ and

$$\mathcal{I}_1 \triangleq \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\tilde{d}_{ij}^2\exp(-u_x^2-u_y^2)\mathrm{d}u_x\mathrm{d}u_y, \quad (27)$$

$$\mathcal{I}_2 \triangleq \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left((\Delta_{x_{ij}}+\lambda u_x)^2+(\Delta_{y_{ij}}+\lambda u_y)^2\right)\exp(-u_x^2-u_y^2)\,\mathrm{d}u_x\mathrm{d}u_y, \quad (28)$$

$$\mathcal{I}_3 \triangleq \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\sqrt{(\Delta_{x_{ij}}+\lambda u_x)^2+(\Delta_{y_{ij}}+\lambda u_y)^2}\,\exp(-u_x^2-u_y^2)\,\mathrm{d}u_x\mathrm{d}u_y, \quad (29)$$

in which $\hat{\mathbf{p}}_i - \hat{\mathbf{p}}_j = [\Delta_{x_{ij}}, \Delta_{y_{ij}}]$ with $\Delta_{x_{ij}} \triangleq \hat{p}_{x_i} - \hat{p}_{x_j}$ and $\Delta_{y_{ij}} \triangleq \hat{p}_{y_i} - \hat{p}_{y_j}$.

In the remainder of this proof the subscript $_{ij}$ is temporarily omitted for notational convenience.

The integrals $\mathcal{I}_1$ and $\mathcal{I}_2$ shown above admit the following trivial solutions [32]:

$$\mathcal{I}_1 = \pi\tilde{d}^2 \quad (30)$$

$$\mathcal{I}_2 = \pi(\hat{d}^2+\lambda^2). \quad (31)$$

In turn, $\mathcal{I}_3$ can be solved as follows. First, consider the change of variables

$$\Delta_x + \lambda u_x = \rho\cos\theta, \quad (32)$$

$$\Delta_y + \lambda u_y = \rho\sin\theta. \quad (33)$$

The integral $\mathcal{I}_3$ can then be rewritten as

$$\mathcal{I}_3 = \frac{1}{\lambda^2}\int_0^{+\infty}\int_0^{2\pi}\rho^2\exp\left(\frac{-\rho^2-\hat{d}^2+2\rho(\Delta_x\cos\theta+\Delta_y\sin\theta)}{\lambda^2}\right)\mathrm{d}\rho\mathrm{d}\theta$$
$$= \frac{1}{\lambda^2}\exp\left(\frac{-\hat{d}^2}{\lambda^2}\right)\cdot\int_0^{+\infty}\rho^2\exp\left(\frac{-\rho^2}{\lambda^2}\right)\mathrm{d}\rho$$
$$\times\underbrace{\int_0^{2\pi}\exp\left(\frac{2\rho(\Delta_x\cos\theta+\Delta_y\sin\theta)}{\lambda^2}\right)\mathrm{d}\theta}_{\mathcal{I}_4}. \quad (34)$$

Choosing $\{a=0, p=0, q=0, b=\frac{-2\rho\Delta_x}{\lambda^2}, c=\frac{-2\rho\Delta_y}{\lambda^2}\}$ and performing the change of variables $x=\theta+\pi$ in [49, Eq. 3.338-4.6, pp. 336], yields the following closed-form solution for $\mathcal{I}_4$

$$\mathcal{I}_4 \triangleq \int_0^{2\pi}\exp\left(\frac{2\rho(\Delta_x\cos\theta+\Delta_y\sin\theta)}{\lambda^2}\right)\mathrm{d}\theta = I_0\left(\frac{2\rho\hat{d}}{\lambda^2}\right) \quad (35)$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind and 0th order.

Using (35) into (34) yields

$$\mathcal{I}_3 = \pi\int_0^{+\infty}\frac{\rho^2}{\frac{\lambda^2}{2}}\exp\left(-\frac{\rho^2+\hat{d}^2}{\lambda^2}\right)I_0\left(\frac{\rho\hat{d}}{\frac{\lambda^2}{2}}\right)\mathrm{d}\rho, \quad (36)$$

It is recognized that the integral above is the first moment of a Rice distribution (see [50, Eq. (2.1-140), pp.46] with $\left\{\sigma^2=\frac{\lambda^2}{2}, s=\hat{d}, r=\rho\right\}$), such that the closed-form of $\mathcal{I}_3$ is

$$\mathcal{I}_3 = \pi\lambda\Gamma\left(\frac{3}{2}\right){}_1F_1\left(\frac{3}{2};1;\frac{\hat{d}^2}{\lambda^2}\right)\exp\left(\frac{-\hat{d}^2}{\lambda^2}\right). \quad (37)$$

Substituting (30), (31) and (37) into (26) finally yields (25). □

Since the GDC method requires repeated optimization of various smoothed objectives $\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{Z}})$ with decreasing $\lambda^{(k)}$, the complexity of evaluating (25) is of crucial importance. Indeed, as $\lambda\to 0$, the confluent hyperbolic function increases unboundedly, while the exponential term vanishes. However this issue can be circumvented by using the alternative expressions [48, Eq. 13.1.2, pp. 504]

$${}_1F_1\left(\frac{3}{2};1;s\right) = 1+\sum_{m=1}^{+\infty}\left(s^m\cdot\prod_{t=1}^{m}\left(\frac{1}{2t^2}+\frac{1}{t}\right)\right) \quad (38)$$

and [48, Eq. 13.5.1, pp. 508]

$${}_1F_1\left(\frac{3}{2};1;s\right) = \frac{2e^s}{\sqrt{\pi}}\sum_{p=0}^{P-1}\frac{s^{\frac{1}{2}-p}}{p!}\prod_{t=0}^{p-1}\left(t-\frac{1}{2}\right)^2$$
$$-\frac{s^{\frac{-3}{2}}}{2\sqrt{\pi}}\sum_{r=0}^{R-1}\frac{(-s)^{-r}}{r!}\prod_{t=0}^{r-1}\left(\frac{3}{2}+t\right)^2$$
$$+\mathcal{O}(|s|^{-R})+\mathcal{O}(|s|^{-P}). \quad (39)$$

Notice that the product in (38) decreases fast with t (and therefore m), such that the series convergences quickly for *small* s. In contrast, the residues $\mathcal{O}(|s|^{-R})$ and $\mathcal{O}(|s|^{-P})$ in (39) decrease geometrically with R and P, respectively (see [48, Eq. 13.5.3-4, pp.508]), such that the series converges quickly for *large* s. Therefore, using these two expressions with $s\triangleq\frac{\hat{d}^2}{\lambda^2}$, (25) can be evaluated efficiently for any values of $\hat{d}^2$ and λ^2.

In order to perform numeric optimization of the objective in theorem T2, however, closed-forms of the Gradient and Hessian of $\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{Z}})$ are also required. Thus, the following results are in order.

Lemma L2 (Gradient of the Smoothed ML Objective): The closed-form of the gradient of the smoothed objective $\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{Z}})$, denoted by $\nabla_{\hat{\mathbf{Z}}}\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{Z}})$, is

$$\nabla_{\hat{\mathbf{Z}}}\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{Z}}) \triangleq \sum_{ij\in\mathcal{H}} w_{ij}\cdot\left(2-\frac{\sqrt{\pi}\tilde{d}_{ij}}{\lambda}e_1^{\frac{-\hat{d}_i^2}{\lambda^2}}F_1\left(\frac{3}{2};2;\frac{\hat{d}_i^2}{\lambda^2}\right)\right)$$
$$\times(\mathbf{e}_{ij}\otimes(\hat{\mathbf{z}}_j-\hat{\mathbf{z}}_i)) \quad (40)$$

where $\otimes$ indicates the Kronecker product and $\mathbf{e}_{ij}\in\mathbb{R}^{N_T}$ are row-vectors with the ith and the jth element equal to 1 and -1, respectively.

Proof: It will prove convenient to define the vectorized form of a matrix. Specifically

$$\vec{\boldsymbol{z}}\triangleq\mathrm{vec}(\mathbf{Z})=[\mathbf{z}_1,\ldots,\mathbf{z}_{N_T}]. \quad (41)$$

Thanks to this definition, the kth element of the $N_T\times\eta$ matrix $\mathbf{Z}$, counted continuously from left to right, top to bottom, can be simply denoted $\vec{z}_k$. Next, define

$$f_{ij}(\vec{\boldsymbol{z}};\lambda)\triangleq\lambda^2+\tilde{d}_{ij}^2+\hat{d}_{ij}^2$$
$$-\lambda\sqrt{\pi}\tilde{d}_{ij}\,{}_1F_1\left(\frac{3}{2},1,\frac{\hat{d}_{ij}^2}{\lambda^2}\right)\exp\left(\frac{-\hat{d}_{ij}^2}{\lambda^2}\right), \quad (42)$$

such that $\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{Z}})=\sum_{ij\in\mathcal{H}}w_{ij}f_{ij}(\vec{\boldsymbol{z}};\lambda)$ and the gradient $\nabla_{\hat{\mathbf{Z}}}\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{Z}})$ can be written concisely as

$$\nabla_{\hat{\mathbf{Z}}}\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{Z}})\triangleq\sum_{ij\in\mathcal{H}}w_{ij}\nabla_{\hat{\mathbf{Z}}}f_{ij}(\vec{\boldsymbol{z}};\lambda). \quad (43)$$

In order to derive a closed-form expression for $\nabla_{\hat{\mathbf{Z}}}f_{ij}(\vec{\boldsymbol{z}};\lambda)$, consider the representation of $f_{ij}(\vec{\boldsymbol{z}};\lambda)$ as a composite function $h(g(\vec{\boldsymbol{z}});\lambda)$, with $h:\mathbb{R}\to\mathbb{R}$, $g:\mathbb{R}^{N_T\times\eta}\to\mathbb{R}$, so that each element of $\nabla_{\hat{\mathbf{Z}}}f_{ij}(\vec{\boldsymbol{z}};\lambda)$ can be obtained via the chain rule

$$\frac{\partial h(g(\vec{\boldsymbol{z}});\lambda)}{\partial\vec{z}_k}=\frac{\mathrm{d}h(u;\lambda)}{\mathrm{d}u}\frac{\partial g(\vec{\boldsymbol{z}})}{\partial\vec{z}_k} \quad (44)$$

where $u=g(\vec{\boldsymbol{z}})$ and $k=\{1,\ldots,\eta N_T\}$.

From equation (44) it is evident that the gradient of $f_{ij}(\hat{d}_{ij};\lambda)$ is given by a scalar corresponding to the term $\frac{\mathrm{d}h(u;\lambda)}{\mathrm{d}u}$, multiplied by the vector $\left[\frac{\partial g(\vec{\boldsymbol{z}})}{\partial\vec{z}_1},\ldots,\frac{\partial g(\vec{\boldsymbol{z}})}{\partial\vec{z}_{\eta N_T}}\right]$.

Specifically

$$\nabla_{\hat{\mathbf{Z}}}f_{ij}(\hat{d}_{ij};\lambda)=S_1(\hat{d}_{ij}^2;\lambda)\nabla_{\hat{\mathbf{Z}}}(\hat{d}_{ij}^2) \quad (45)$$

where, from $g_{ij}(\vec{\boldsymbol{z}}) = \hat{d}_{ij}^2$ we obtain

$$\nabla_{\hat{\mathbf{Z}}}(\hat{d}_{ij}^2) \triangleq \left[\frac{\partial \hat{d}_{ij}^2}{\partial \vec{z}_1}, \ldots, \frac{\partial \hat{d}_{ij}^2}{\partial \vec{z}_{\eta N_T}}\right] = 2\mathbf{e}_{ij} \otimes (\hat{\mathbf{z}}_i - \hat{\mathbf{z}}_j) \quad (46)$$

while from $h = f_{ij}(u;\lambda)$ we have

$$S_1(u;\lambda) \triangleq \frac{\mathrm{d} f_{ij}(u;\lambda)}{\mathrm{d}u} = 1 + \frac{\sqrt{\pi}\tilde{d}_{ij}}{\lambda} \exp\left(\frac{-u}{\lambda^2}\right)\left({}_1F_1\left(\frac{3}{2}, 1, \frac{u}{\lambda^2}\right) - \frac{\mathrm{d}}{\mathrm{d}u}{}_1F_1\left(\frac{3}{2}, 1, \frac{u}{\lambda^2}\right)\right), \quad (47)$$

where for simplicity we commit a slight abuse of notation by omitting the subindex ij in S_1.

Equation (47) can be simplified utilizing the recurrence formula [48, pp.508, 13.5.1]

$$\frac{(b-a)}{b}\,{}_1F_1\,(a;b+1;u) =_1 F_1\,(a;b;u) - \frac{\mathrm{d}}{\mathrm{d}u}\,{}_1F_1\,(a;b;u) \quad (48)$$

such that

$$S_1(u;\lambda) = 1 - \frac{\sqrt{\pi}\tilde{d}_{ij}}{2\lambda} \exp\left(\frac{-u}{\lambda^2}\right)_1 F_1\left(\frac{3}{2}; 2; \frac{u}{\lambda^2}\right). \quad (49)$$

Substituting (46) and (49) into (45), and the result into (43) yields (40). □

Lemma L3 (Hessian of the Smoothed ML Objective): The Hessian of $\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{Z}})$, denoted by $\nabla^2_{\hat{\mathbf{Z}}}\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{Z}})$, is

$$\nabla^2_{\hat{\mathbf{Z}}}\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{Z}}) = \sum_{ij\in\mathcal{H}} w_{ij}\cdot\Big(S_1(\hat{d}_{ij}^2;\lambda)\nabla^2_{\hat{\mathbf{Z}}}(\hat{d}_{ij}^2) + S_2(\hat{d}_{ij}^2;\lambda)\nabla^T_{\hat{\mathbf{Z}}}(\hat{d}_{ij}^2)\nabla_{\hat{\mathbf{Z}}}(\hat{d}_{ij}^2)\Big), \quad (50)$$

where $\nabla^2_{\hat{\mathbf{Z}}}(\hat{d}_{ij}^2) \in \mathbb{R}^{\eta N_T \times \eta N_T}$ and it is given by

$$\nabla^2_{\hat{\mathbf{Z}}}(\hat{d}_{ij}^2) = 2(\mathbf{e}_{ij} \otimes \mathbf{I}_\eta)^T(\mathbf{e}_{ij} \otimes \mathbf{I}_\eta) \quad (51)$$

and

$$S_2(\hat{d}_{ij}^2;\lambda) = \frac{\tilde{d}_{ij}\sqrt{\pi}}{8\lambda^3} \exp\left(\frac{-\hat{d}_{ij}^2}{\lambda^2}\right)_1 F_1\left(\frac{3}{2}; 3; \frac{\hat{d}_{ij}^2}{\lambda^2}\right). \quad (52)$$

Proof: Proceeding as in the proof of Lemma 2, the Hessian of $\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{Z}})$ can be written as

$$\nabla^2_{\hat{\mathbf{Z}}}\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{Z}}) \triangleq \sum_{ij\in\mathcal{H}} w_{ij}\nabla^2_{\hat{\mathbf{Z}}} f_{ij}(\vec{\boldsymbol{z}};\lambda) \quad (53)$$

where kq-th element of $\nabla^2_{\hat{\mathbf{Z}}} f_{ij}(\hat{d}_{ij}^2;\lambda)$ is obtained from the chain rule

$$\frac{\partial^2 h(g(\vec{\boldsymbol{z}}))}{\partial \vec{z}_k \partial \vec{z}_q} = \frac{\mathrm{d}h(u)}{\mathrm{d}u}\frac{\partial^2 g(\vec{\boldsymbol{z}})}{\partial \vec{z}_k \partial \vec{z}_q} + \frac{\mathrm{d}^2 h(u)}{\mathrm{d}^2 u}\frac{\partial g(\vec{\boldsymbol{z}})}{\partial \vec{z}_k}\frac{\partial g(\vec{\boldsymbol{z}})}{\partial \vec{z}_q}. \quad (54)$$

From (54) it can be seen that the Hessian of $f_{ij}(\hat{d}_{ij}^2;\lambda)$ is a sum of two matrices, the first given by $\nabla^2_{\hat{\mathbf{Z}}}(\hat{d}_{ij}^2) = \left[\frac{\partial^2 g(\vec{\boldsymbol{z}})}{\partial \vec{z}_k \partial \vec{z}_q}\right]_{\forall(k,q)}$, multiplied by the scalar $S_1(u;\lambda)$ found in (49); and the second given by the cross-product of the vector $\nabla_{\hat{\mathbf{Z}}}(\hat{d}_{ij}^2)$ described in (46), multiplied by another scalar corresponding to the term $\frac{\mathrm{d}^2 h(u)}{\mathrm{d}^2 u}$. Specifically

$$\nabla^2_{\hat{\mathbf{Z}}} f_{ij}(\hat{d}_{ij};\lambda) = S_1(\hat{d}_{ij}^2;\lambda)\nabla^2_{\hat{\mathbf{Z}}}(\hat{d}_{ij}^2) + S_2(\hat{d}_{ij}^2;\lambda)\nabla^T_{\hat{\mathbf{Z}}}(\hat{d}_{ij}^2)\nabla_{\hat{\mathbf{Z}}}(\hat{d}_{ij}^2), \quad (55)$$

with

$$\nabla^2_{\hat{\mathbf{Z}}}(\hat{d}_{ij}^2) = \left[\frac{\partial^2 g(\vec{\boldsymbol{z}})}{\partial \vec{z}_k \partial \vec{z}_q}\right]_{\forall(k,q)} = 2(\mathbf{e}_{ij} \otimes \mathbf{I}_\eta)^T(\mathbf{e}_{ij} \otimes \mathbf{I}_\eta) \quad (56)$$

and

$$S_2(u;\lambda) \triangleq \frac{\mathrm{d}^2 f_{ij}(u;\lambda)}{\mathrm{d}u^2} = \frac{\tilde{d}_{ij}\sqrt{\pi}}{2\lambda^3} \exp\left(!\frac{-u}{\lambda^2}\right)\left(\frac{\mathrm{d}}{\mathrm{d}u}{}_1F_1\left(\frac{3}{2}; 2; \frac{u}{\lambda^2}\right) - {}_1F_1\left(\frac{3}{2}; 2; \frac{u}{\lambda^2}\right)\right) = \frac{\tilde{d}_{ij}\sqrt{\pi}}{8\lambda^3} \exp\left(\frac{-u}{\lambda^2}\right)_1 F_1\left(\frac{3}{2}; 3; \frac{u}{\lambda^2}\right) \quad (57)$$

where we again omit the subindex ij in S_2 for simplicity and employed the recursive relation of (48) [48, pp. 508, 13.5.1] in the last equality.

Substituting (46), (49), (56), and (57) into (55) and the result into (53) finally yields (50). □

B. The Initial Smoothing Parameter $\lambda^{(1)}$

In this subsection we address the second challenge outlined in the begining of Section III, namely, the derivation of a least upper bound on the set of smoothing parameters that ensures that first smoothed objective is convex in the whole domain, thus lending the R-GDC algorithm complete insensitivity to initial estimates, as required by the ML principle.

Definition D3 (Set of Convexizing Smoothing Parameters): The set of **convexizing smoothing parameters** is defined by

$$\mathscr{L} \triangleq \{\lambda | \nabla^2_{\hat{\mathbf{Z}}}\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{Z}}) \succeq 0,\ \forall \hat{\mathbf{Z}} \in \mathbb{R}^{N_T \times \eta}\}. \quad (58)$$

Definition D4 (Critical Smoothing Parameter λ^*): Let $\lambda^* \in \mathscr{L}$ be a smoothing parameter such that

$$\nexists \lambda < \lambda^* | \nabla^2_{\hat{\mathbf{Z}}}\langle f_{\mathrm{R}}\rangle_{\lambda^*}(\hat{\mathbf{Z}}) \succeq 0,\ \forall \hat{\mathbf{Z}} \in \mathbb{R}^{N_T \times \eta}. \quad (59)$$

The parameter λ^*, which can alternatively be defined as $\lambda^* \triangleq \inf\{\mathscr{L}\}$, will hereafter be referred to as the **critical smoothing parameter**.

A closed-form and simple expression for the critical smoothing parameter λ^* is hard to determine. However, the GDC method requires that various smoothed objectives $\langle f_{\mathrm{R}}\rangle_{\lambda^{(k)}}(\hat{\mathbf{Z}})$ be minimized with decreasing smoothing parameters $\lambda^{(1)} > \cdots > \lambda^{(k)} > \cdots > \lambda^{(K)} = 0$, such that the first smoothed objective $\langle f_{\mathrm{R}}\rangle_{\lambda^{(1)}}(\hat{\mathbf{Z}})$ is convex and the last $\langle f_{\mathrm{R}}\rangle_{\lambda^{(K)}}(\hat{\mathbf{Z}}) = f_{\mathrm{R}}(\hat{\mathbf{Z}})$. Referring to Definition D3 and to the set $\mathcal{L} \triangleq \{\lambda^{(k)}\}$ defined in Theorem T1, it is sufficient and necessary that $\lambda^{(1)} \in \mathscr{L} \implies \lambda^{(1)} \geq \lambda^*$, such that $\lambda^{(1)}$ can be referred to as a *majorizer* of λ^*. In the sequel, a simple closed-form expression for a *majorizer* of λ^* is derived.

Theorem T3 (Initial Smoothing Parameter: Source Loc): Let $\{\tilde{d}_{it}\}$ be the set of ranging measurement between the anchors $\{\mathbf{a}_i\}$ and a single target. Then

$$\left\{\lambda \geq \frac{\sqrt{\pi}}{2}\frac{\sum_{i=1}^{N_A} w_{it}\tilde{d}_{it}}{\sum_{i=1}^{N_A} w_{it}} \triangleq \hat{\lambda}_S^*\right\} \Longrightarrow \left\{\nabla^2_{\hat{\mathbf{z}}}\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{Z}}) \succeq 0,\, \forall\, \hat{\mathbf{Z}} \in \mathbb{R}^{\eta N_T}\right\}. \quad (60)$$

Proof: Invoking (50), and noticing that in the source localization case $\nabla^T_{\hat{\mathbf{z}}}(\hat{d}^2_{jt})\nabla_{\hat{\mathbf{z}}}(\hat{d}^2_{jt})$ and $\nabla^2_{\hat{\mathbf{z}}}(\hat{d}^2_{jt})$ reduce to $\frac{4}{\hat{d}^2_{jt}}\boldsymbol{\Upsilon}_j$ and $2\mathbf{I}_\eta$, as can be found by setting $\mathbf{e}_{jt} = 1$ in (56) and (46), respectively, we obtain

$$\nabla^2_{\hat{\mathbf{z}}}\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{z}}) \succeq \underbrace{\left(2\sum_{i=1}^{N_A} w_{it}\left(1 - \frac{\sqrt{\pi}}{2\lambda}\tilde{d}_{it}\right)\right)\mathbf{I}_\eta}_{\boldsymbol{\Psi}} + \underbrace{\sum_{i=1}^{N_A}\frac{4w_{it}}{\hat{d}^2_{it}}S_2(\hat{d}^2_{it};\lambda)\cdot\boldsymbol{\Upsilon}_i}_{\bar{\boldsymbol{\Upsilon}}}, \quad (61)$$

where $\boldsymbol{\Upsilon}_i$ are as in (10) for $\eta = 2$ and can be found in [51] for $\eta = 3$, and the inequality results from the minorization of $S_1(\hat{d}^2_{it};\lambda)$ by $\left(1 - \frac{\sqrt{\pi}}{2\lambda}\tilde{d}_{it}\right)$ in light of Lemma L4 and (49). Next, invoke the Wenyl bounds [52, Eq. (2.3)], which yields

$$\gamma_1(\boldsymbol{\Psi} + \bar{\boldsymbol{\Upsilon}}) \geq \gamma(\boldsymbol{\Psi}) + \gamma_1(\bar{\boldsymbol{\Upsilon}}) \quad (62)$$

where $\gamma_k(\cdot)$ denote the kth smallest eigenvalue of a matrix and $\gamma(\boldsymbol{\Psi}) \triangleq \gamma_1(\boldsymbol{\Psi}) = \cdots = \gamma_\eta(\boldsymbol{\Psi})$, since $\boldsymbol{\Psi}$ has a single eigenvalue of multiplicity η.

Finally, notice that in order for inequality (61) to hold it is sufficient it that $\gamma_1(\boldsymbol{\Psi}+\bar{\boldsymbol{\Upsilon}})$ is nonnegative, which under the fact that $\bar{\boldsymbol{\Upsilon}}$ is positive semidefinite, i.e., $\gamma_1(\bar{\boldsymbol{\Upsilon}}) \geq 0$, is equivalent to

$$\sum_{i=1}^{N_A} w_{it}\left(1 - \frac{\sqrt{\pi}\tilde{d}_{it}}{2\lambda}\right) \geq 0, \quad (63a)$$

$$\lambda \geq \frac{\sqrt{\pi}}{2}\cdot\frac{\sum_{i=1}^{N_A} w_{it}\tilde{d}_{it}}{\sum_{i=1}^{N_A} w_{it}} \triangleq \hat{\lambda}_S^*. \quad (63b)$$

□

In addition to an analytical and simple expression for the majorizer of $\hat{\lambda}^*$, which can be used as the initial smoothing parameter in the R-GDC algorithm, Theorem T3 also hints on a strategy to select the remaining $\lambda^{(k)}$'s. To elaborate, let $\{\tilde{d}_{it}\}$ be ordered without loss of generality so that $\tilde{d}_{1t} \geq \cdots \geq \tilde{d}_{N_A t}$. Then, a set $\mathcal{L}_S = \{\lambda^{(k)}\}$ of $K = N_A+1$ "landmark" smoothing parameters are

$$\hat{\lambda}_S^* \geq \cdots \geq \frac{\sqrt{\pi}}{2}\frac{\sum_{i=k}^{N_A} w_{it}\tilde{d}_{it}}{\sum_{i=k}^{N_A} w_{it}} \geq \cdots \geq \frac{\sqrt{\pi}}{2}\tilde{d}_{1t} \geq 0 \quad (64)$$

such that $\lambda^{(k)} \triangleq \frac{\sqrt{\pi}}{2}\frac{\left(\sum_{i=k}^{N_A} w_{it}\tilde{d}_{it}\right)}{\left(\sum_{i=k}^{N_A} w_{it}\right)}$.

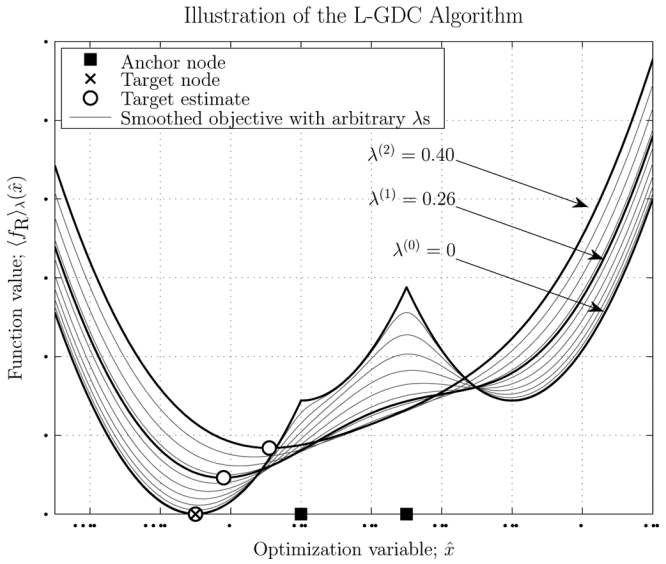


Fig. 1. Example of the R-GDC method in a one-dimensional scenario with $N_A = 2$ anchors and a single target. Bold lines indicate the smoothed objective functions obtained with the sequence of $\lambda^{(k)}$'s given in (64), while gray lines are for intermediate smoothing parameters. It can be seen that utilizing the proposed strategy only $N_A + 1$ iterations are required to achieve the global optimum.

The sequence shown above is monotonically descending, involves less information to be evaluated as k increases and, as can be inferred form the proof of Theorem T3, leads to successively less smooth objectives $\langle f_{\mathrm{R}}\rangle_{\lambda^{(k)}}(\hat{\mathbf{Z}})$, as desired.

Obviously, arbitrarily many additional smoothing parameters can be selected in between the ones given above as well, although exhaustively numerical tests have consistently shown that to be typically unnecessary (see Section IV-B).

An illustration of the effect of choosing the aforementioned landmark smoothing parameters in a scenario with $N_A = 2$ anchors—marked by dark squares—is shown in Fig. 1. The figure shows plots of smoothed objectives $\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{Z}})$, with various smoothing parameters, including those given in the sequence (64) (bold lines) as well as others assuming intermediary values (gray lines).

The minima of corresponding objectives are marked with a "◯," while the actual target location is marked with a "×," which coincides with the last "◯" obtained at the last round of the R-GDC algorithm. It can be observed that the landmark smoothing parameters captures the progressively less smooth behavior of $\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{Z}})$, allowing for the global minimum of $f_{\mathrm{R}}(\hat{\mathbf{Z}})$ to be attained with a small number of R-GDC continuations.

Theorem T4 (Initial Smoothing Parameter: Network Loc):

$$\left\{\lambda \geq \frac{\sqrt{\pi}}{2}\max_{ij\in\mathcal{H}}\tilde{d}_{ij}\right\} \Longrightarrow \{\nabla^2_{\hat{\mathbf{Z}}}\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{Z}}) \succeq 0,\, \forall\, \hat{\mathbf{Z}} \in \mathbb{R}^{\eta N_T}\}. \quad (65)$$

Proof: Recall from (53) that the Hessian matrix of $\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{Z}})$ is given by a positive sum of the matrices $\nabla^2_{\hat{\mathbf{Z}}} f_{ij}(\hat{d}_{ij};\lambda)$, such that a sufficient condition for $\nabla^2_{\hat{\mathbf{x}}}\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\mathbf{Z}}) \succeq 0$ is that $\nabla^2_{\hat{\mathbf{Z}}} f_{ij}(\hat{d}_{ij};\lambda) \succeq 0$, $\forall ij$. Next, invoke (55) and notice that the matrices $\nabla^T_{\hat{\mathbf{Z}}}(\hat{d}^2_{ij})\nabla_{\hat{\mathbf{Z}}}(\hat{d}^2_{ij})$ and $\nabla^2_{\hat{\mathbf{Z}}}(\hat{d}^2_{ij})$ are positive semidefinite since the first is the Gramian of the gradient vector $\nabla_{\hat{\mathbf{Z}}}(\hat{d}^2_{ij})$ and the second is the Hessian of a norm function [44].

Therefore a sufficient condition to ensure that $\nabla^2_{\hat{\mathbf{Z}}} f_{ij}(\hat{d}_{ij};\lambda) \succeq 0$ is that the functions $S_1(u;\lambda)$ and $S_2(u;\lambda)$ given in (49) and (57), respectively, are both positive for any u. In order to prove that $S_1(u;\lambda)$ is positive consider the following inequalities

$$1 \geq \frac{\sqrt{\pi}\tilde{d}_{ij}}{2\lambda}\exp\left(\frac{-u}{\lambda^2}\right)_1 F_1\left(\frac{3}{2};2;\frac{u}{\lambda^2}\right), \tag{66a}$$

$$\lambda \geq \frac{\sqrt{\pi}\tilde{d}_{ij}}{2}\exp\left(\frac{-u}{\lambda^2}\right)_1 F_1\left(\frac{3}{2};2;\frac{u}{\lambda^2}\right) \geq \frac{\sqrt{\pi}\tilde{d}_{ij}}{2}, \tag{66b}$$

where the last inequality holds due to Lemma L4, ii).

As for $S_2(u;\lambda)$, referring to (57), it is clear that $S_2(u;\lambda)$ is positive since it is a product of positive terms, namely, the exponential function, the Hypergeometric function with parameters $a=\frac{3}{2}, b=3$ (see [48, pp.514]) and the coefficient $\sqrt{\pi}(8\lambda^3)^{-1}\tilde{d}_{ij}$, which is positive since negative distance measurements $\tilde{d}_{ij}$ are discarded due to the physical impossibility of $d_{ij}<0$. Finally, since the last inequality of (66) must hold for all ij, the implication (65) requires that

$$\lambda \geq \frac{\sqrt{\pi}}{2}\max_{ij\in\mathcal{H}}\tilde{d}_{ij} \triangleq \hat{\lambda}^*_N. \tag{67}$$

□

Notice that the majorizer given in (67) is also a majorizer of the critical smoothing parameter for the source localization case.

In other words, if the expression in equation(67) is used for a set of distance measurements $\{\tilde{d}_{it}\}$ associated to a single target, we have $\hat{\lambda}^*_N \geq \hat{\lambda}^*_S \geq \lambda^*$.

As for the remaining smoothing parameters, in the network localization case the number of pairwise distance estimates $\tilde{d}_{ij}$ grows geometrically with the number of nodes, such that the approach of calculating $\lambda^{(k)} = \frac{\sqrt{\pi}}{2}\tilde{d}_k$, as inspired by Theorem T4, is prohibitive. Instead, we therefore utilize

$$\mathcal{L}_N = \left\{\frac{K-k+1}{K}\hat{\lambda}^*_N\right\}. \tag{68}$$

IV. Performance Analysis and Results

A. Fundamental Limits on Localization Error

This subsection is dedicated to the analysis of the fundamental performances limits of localization algorithms. In particular, we briefly revise the best-known and most suitable lower bounds on the localization errors in both single-target (source) and network scenarios. Fundamental limits for the localization problem can be obtained under two underlying approaches, namely, *deterministic* and *stochastic* (often Bayesian). In the first, lower bounds are formulated under the assumption of *full* (deterministic) knowledge on the location of the target(s).

An example of such bounds is the well-known CRLB,

$$\text{CRLB} \triangleq \sqrt{\text{trace}(\mathbf{J}^\dagger)}, \tag{69}$$

where $\text{trace}(\cdot)$ denotes the trace, † denotes pseudo-inverse and $\mathbf{J}\in\mathbb{R}^{\eta N_T\times\eta N_T}$ is the Fisher Information Matrix (FIM) whose kqth element is, in the context hereby, given by

$$[\mathbf{J}]_{kq} \triangleq \int_\Omega \frac{1}{p(\hat{\mathbf{Z}}|\tilde{\mathbf{D}})}\frac{\partial p(\hat{\mathbf{Z}}|\tilde{\mathbf{D}})}{\partial \vec{\hat{z}}_k}\frac{\partial p(\hat{\mathbf{Z}}|\tilde{\mathbf{D}})}{\partial \vec{\hat{z}}_q}\mathrm{d}\tilde{\mathbf{D}} \tag{70}$$

where $p(\hat{\mathbf{Z}}|\tilde{\mathbf{D}})$ is the likelihood function defined in (4), $\vec{\hat{z}}_k$ is the kth element of the vectorized coordinate matrix $\hat{\mathbf{Z}}$ (see Lemma L2), and Ω is the space of observations $\tilde{\mathbf{D}}$.

Closed form expressions for the FIM components and resulting CRLB on network localization errors can be easily found in the literature, e.g., [23], [53], [54].

Despite being popular and relatively easy to evaluate, the suitability of the CRLB as a measure of the best performance of ML-WLS localization algorithms is, however, somewhat limited. One fundamental limitation of the CRLB, for instance, is that it is known to be loose compared to the performance of the ML estimator [55], [56] in the presence of large perturbations. This is of critical relevance to localization problems since location ambiguity [57], [58] caused by certain geometric configurations of the nodes' locations (topology) may lead to large localization errors not accounted for by small scale error bounds such as the CRLB and the Bhattacharyya bounds [59].

For the sake of illustration, in this article only an example is offered in the next subsection (see Figs. 2 and 3), which exposes the ambiguity problems and the performance mismatch between the Mean Square Error (MSE) achieved with an ideal ML-based localization algorithm and the CRLB. The location ambiguity problem [57], [58] is in fact a topic that also deserves further investigation, especially in the context of mesh networks.

While other deterministic bounds—most noticeably the Barankin, the Reuven-Messeur (simplified Barankin) and the Abel bounds [55]—do exist, which aim at improving accuracy at the low SNR regime, such bounds still prove unable to fully capture the effect of ambiguity in localization problems and therefore offer little improvement over the CRLB in this particular application. This issue reveals a second and more fundamental limitation of deterministic bounds when applied to localization problems, namely, the very assumption of exact prior knowledge of $\mathbf{Z}$ implied. Indeed, if any *a priori* information on $\mathbf{Z}$ is available, the "localization" problem is better defined as a *tracking* problem for which not only entirely different algorithmic approaches exist [60]–[62] but also more suitable bounds, such as the Bayesian Abel, the Bayesian Borbrovsky-Zakai, and the Weiss-Weinstein bounds, can be considered [56]. Non-surprisingly, the latter bounds belong to the category of stochastic bounds in which the assumption of prior knowledge on $\mathbf{Z}$ is "softened" in the sense that $\mathbf{Z}$ is assumed to follow a certain distribution which incidentally may be even uniform (that is, no prior knowledge assumed!).

Amongst many stochastic bounds found in the literature, the one proposed by Wang *et al.* in [41], [63] is of particular interest, since it foregoes entirely the use of a prior distribution $p(\hat{\mathbf{Z}})$, replacing it instead by a minimum-entropy estimate $p(\hat{\mathbf{Z}}|\mathbf{D})$, constructed directly from the exact EDM associated to the problem. This bound can summarized as follows.

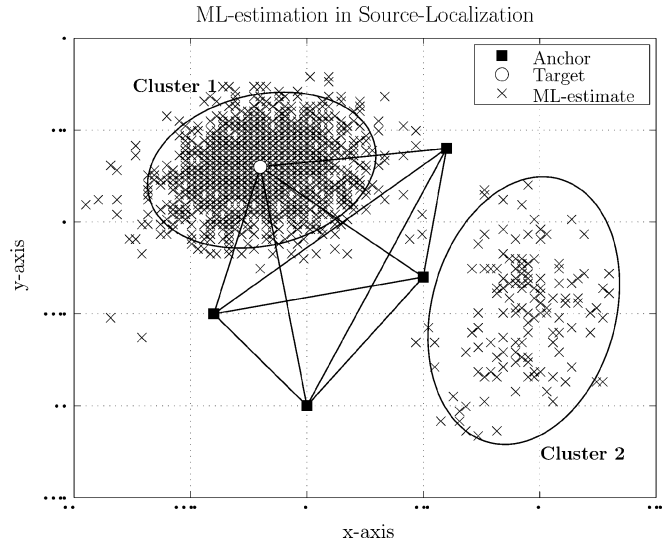


Fig. 2. Localization of a single target in a two-dimensional network with $N_A = 4$ anchors. Multiple ML estimations, obtained via exhaustive search under different realizations of Gaussian random noise with the same deviation $\sigma_{ij} = 0.25$ are shown (marked by "×'s"). The ML estimates form two distinct clusters, illustrating the flip-ambiguity phenomenon.

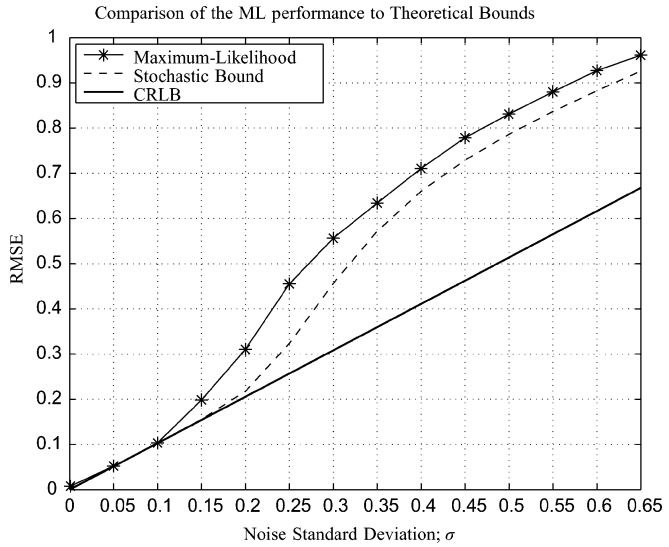


Fig. 3. Comparison of the ML-RMSE against the CRLB and the stochastic bound of Theorem T5 corresponding to the scenario depicted in Fig. 2. Results are shown for a range noise standard deviations σ_{ij}, with each point obtained out of $L = 2000$ realizations. While the CRLB fails to capturing the effect of flip-ambiguity, the stochastic bound of Theorem T5, in turn, is substantially tighter.

Theorem T5 (Wang Stochastic Bound): Let $\vec{\boldsymbol{z}}^* \triangleq \mathrm{vec}(\max_{\hat{\mathbf{Z}}} p(\hat{\mathbf{Z}}|\tilde{\mathbf{D}}))$ denote the (row) vectorized ML estimate of the variable $\hat{\mathbf{Z}}$ under an observation $\tilde{\mathbf{D}}$, governed by the likelihood function $p(\hat{\mathbf{Z}}|\tilde{\mathbf{D}})$. Likewise, let $\vec{\boldsymbol{u}}$ be an auxiliary variable whose distribution $p(\vec{\boldsymbol{u}}|\mathbf{D})$ is the minimum-entropy estimate of $\vec{\boldsymbol{z}}^*$. Then

$$E_{\vec{\boldsymbol{z}}^*}\left[(\vec{\boldsymbol{z}}^* - \vec{\boldsymbol{z}})^T \cdot (\vec{\boldsymbol{z}}^* - \vec{\boldsymbol{z}})\right] \geq \int_{\mathbb{R}^{\eta N_T}} (\vec{\boldsymbol{u}} - \vec{\hat{\boldsymbol{u}}})^T \cdot (\vec{\boldsymbol{u}} - \vec{\hat{\boldsymbol{u}}})\, p(\vec{\boldsymbol{u}}|\mathbf{D}) \mathrm{d}\vec{\boldsymbol{u}} \tag{71}$$

in which $E_{\vec{\boldsymbol{z}}^*}[\cdot]$ denotes expectation with respect to the distribution of $\vec{\boldsymbol{z}}^*$ and

$$\vec{\hat{\boldsymbol{u}}} \triangleq \int_{\mathbb{R}^{\eta N_T}} \vec{\boldsymbol{u}}\, p(\vec{\boldsymbol{u}}|\mathbf{D}) \mathrm{d}\vec{\boldsymbol{u}} \tag{72}$$

Proof: See [41], [63]. □

B. Simulation Results and Comparisons

In this section, we evaluate the performance of the proposed R-GDC algorithm in both source and network localization scenarios, comparing it to alternative methods, namely, the SDP, the SMACOF and the NLS algorithms as well as to the theoretical bounds discussed in Section IV-A.

First consider Fig. 2, which shows an example of source localization via Maximum-likelihood[4]

It can be seen that the ML estimates (marked by "×'s") are grouped in two clusters, identified by ellipses computed from the eigenspectrum of the covariance matrices associated with each cluster (see [64, Sec. IV]), which illustrates the ambiguity phenomenon typical of localization problems.

Definition D5 (Root Mean Square Error): Let $\hat{\mathbf{Z}}_\ell^*$ be the solution yielded by the ℓth out of L minimization algorithms under a given realization $\tilde{\mathbf{D}}$. Then, the **root mean square error** (RMSE) is given by

$$\bar{\xi}(\ell) \triangleq \sqrt{E_{\tilde{\mathbf{D}}}\left[\|\hat{\mathbf{Z}}_\ell^* - \mathbf{Z}\|_F^2\right]}. \tag{73}$$

In Fig. 3, we compare the RMSE achieved with the ML method for the network depicted in Fig. 2 against the CRLB and the stochastic bound of Theorem T5, as a function of the noise deviation. The results indicate that for a sufficiently large σ (low-SNR region), the RMSE achieved with the ML algorithm is biased by the errors associated with the outliers forming cluster 2. The phenomenon is captured by the SB of Theorem T5, but not by the CRLB. Based on this result, the SB will be considered as the main benchmark for a source localization problem since it is the one that best captures the real ML performance. In the sequel, we study the ergodic performance of the aforementioned localization algorithms under the assumption of random networks, comparing the proposed R-GDC against alternative methods, in particular the SDP [30], the SMACOF [28], [29], and the NLS [47].

Efficient versions of implementations of the SMACOF, NLS and R-GDC algorithms were implemented as follows. In the R-GDC technique, the smoothed objectives are minimized using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) Quasi-Newton method, which requires no matrix inversion [47]. The SMACOF is implemented using the Guttman-transform to minimize the majorized objective in closed-form, and a fast-convergent implementation of the NLS method is used in which the subspace trust-region technique with pre-computed

[4]The minimization of the ML objective function was performed numerically via exhaustive search. on a network formed with a target (marked by "○") at the coordinates $[0.6, 0.4]$, and N_A anchor nodes (marked by "■") at $[0, -1; 0.5, -0.3; -0.4, -0.5; -0.2, 0.3]$. In this simulation, the network is maintained fixed while the several ranging experiments are performed using the ranging error model given in (2), with a noise standard deviation $\sigma_{ij} = 0.25$ for all anchor-target links.

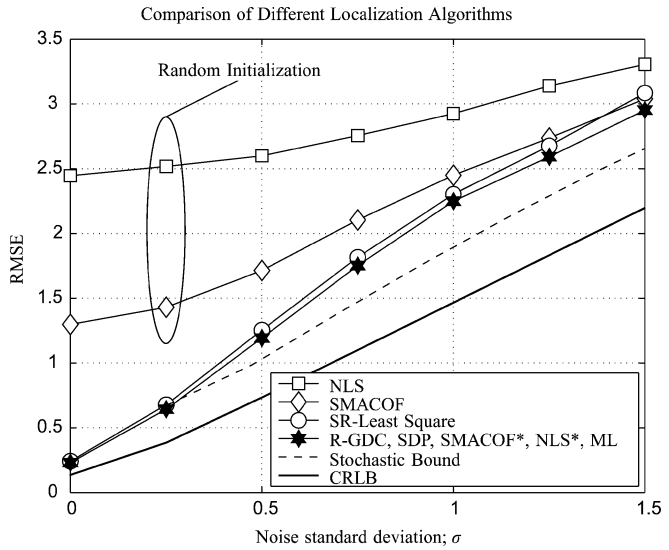


Fig. 4. Performance of different localization algorithms in scenarios with $N_A = 4$ anchors and $N_T = 1$ target randomly distributed in a 10-by-10 square. The dashed line indicates the stochastic bound of Theorem T5, while the solid line without markers indicates the CRLB. Each point is obtained out of 100 network configurations and 300 distance measurement realizations.

Jacobians [47] is incorporated. Pseudo codes of the SMACOF, NLS and R-GDC algorithms are given in the Appendixes.

In order to improve the accuracy of the SMACOF and NLS algorithms, we employ initializations obtained via the combination of the Nyström and shortest-path completion algorithms described in [65]. Hereafter we distinguish the two initialization strategies, namely the random and the Nyström-based, by labeling the name of the algorithm with a "*" when the latter is used.

In addition to *localization accuracy*, measured by the RMSE defined in Definition D5, the performances of these algorithms are evaluated using also the following criteria.

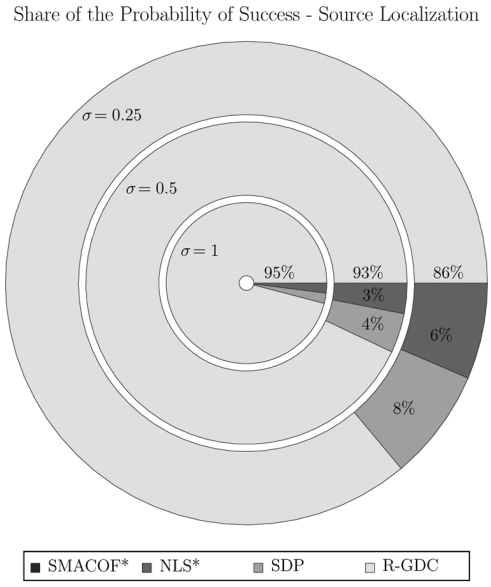
Definition D6 (Minimization Effectiveness): The **minimization effectiveness** of the ℓth algorithm over the set of L alternatives is defined as

$$\Pi(\ell) \triangleq \Pr\left\{ f_{\mathrm{R}}(\hat{\mathbf{Z}}_\ell^*|\tilde{\mathbf{D}}) = \min_{\forall q} f_{\mathrm{R}}(\hat{\mathbf{Z}}_q^*|\tilde{\mathbf{D}}) \right\}. \quad (74)$$

Definition D7 (Minimizer Deviation): The **minimizer deviation** of the ℓth algorithm over the set of L alternatives is defined as

$$\zeta(\ell) \triangleq E_{\tilde{\mathbf{D}}}\left[||f_{\mathrm{R}}(\hat{\mathbf{Z}}_\ell^*|\tilde{\mathbf{D}}) - \min_{\forall q} f_{\mathrm{R}}(\hat{\mathbf{Z}}_q^*|\tilde{\mathbf{D}})||_F^2 \right]. \quad (75)$$

In plain words, the *minimization effectiveness* of a given algorithm is the likelihood that *it* is the best minimizer *amongst the methods compared*; while the *minimizer deviation*, quantifies the average error of the objective at the solution attained by each algorithm with respect to the that attained by the best minimizer. Let us start the comparisons with Fig. 4, in which the ergodic RMSE performances of the SMACOF, NLS, SDP and R-GDC algorithms are shown. These results of all algorithms were obtained on the basis of the same data, constructed out of random networks where $N_A = 4$ anchors and a single target



Average Minimizer Deviation - Source Localization

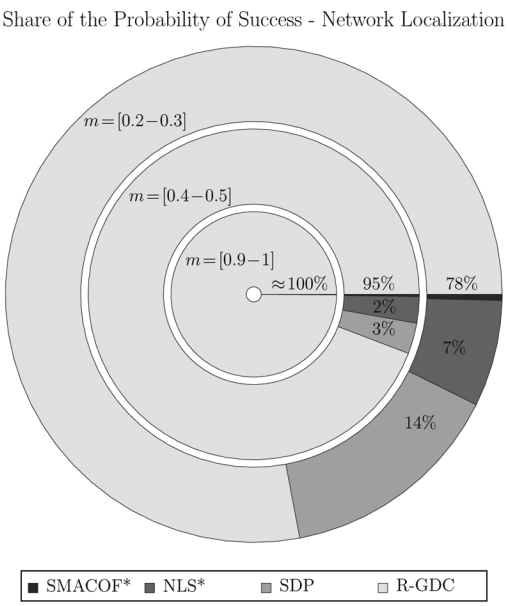
σ	SMACOF*	NLS*	SDP	R-GDC
0.25	3.5e-02	4.5e-03	1.0e-04	4.0e-04
0.50	5.5e-02	7.9e-03	6.0e-04	7.0e-04
0.75	6.0e-02	1.0e-03	1.3e-03	1.3e-03
1.00	6.9e-02	1.0e-02	1.8e-03	1.5e-03
1.25	7.2e-02	1.1e-02	1.8e-03	1.3e-03
1.50	7.7e-02	1.1e-02	1.1e-03	1.0e-03
1.75	8.2e-02	1.2e-02	1.4e-03	1.4e-03
2.00	8.7e-02	1.3e-02	1.2e-03	1.6e-03
2.25	9.2e-02	1.3e-02	1.0e-03	1.1e-03
2.50	9.8e-02	1.4e-02	1.0e-03	1.0e-03

Fig. 5. Ring-chart indicating the minimization effectiveness $\Pi(\ell)$ of various localization algorithms in networks with $N_A = 4$ anchors and $N_T = 1$ under various noise standard deviations $\sigma \in [0, 1]$. Similar ratios of $\Pi(\ell)$ to those obtained with $\sigma = 1$ are observed for $\sigma > 1$. All minimization techniques operate under the same tolerances of 10^{-8} on the variations of objective function and the estimates.

are uniformly distributed in a space of 10×10 meters. The standard deviation of the ranging errors are varied within the interval $\sigma \in (0, 1.5)$.

For a global perspective on the accuracy of these algorithms, the ergodic CRLB and the stochastic bound of Theorem T5 are also shown. Fig. 4 indicates that the R-GDC, the SDP, and the optimized SMACOF* and NLS* have similar accuracies. Notice, however, that the R-GDC have much lower computational complexity compared to the SDP – (as will be shown later in this section)—and that the solution of the R-GDC is independent of the initialization, unlike the SMACOF* and NLS*. Indeed, the plots show that the performance of the SMACOF and the NLS are comparable to those achieved by the R-GDC only when initialized by the Nyström method.

In order to emphasize the advantage of the R-GDC method, consider the comparison of these algorithms in terms of their minimization effectiveness and minimizer deviations, offered in Fig. 5. The ring chart reveals the clear advantage of the proposed R-GDC method in terms of minimization effectiveness, which in fact is found to be more substantial as the SNR deteriorates. In turn, the accompanying table reveals that the minimizer deviation of the R-GDC algorithm is typically at the order



Average Minimizer Deviation - Network Localization

m	SMACOF*	NLS*	SDP	R-GDC
(0.2,0.3)	6.3e-01	5.6e-01	6.1e-01	5.0e-02
(0.3,0.4)	7.4e-01	5.7e-01	9.9e-02	4.3e-02
(0.4,0.5)	5.9e-01	2.9e-01	8.5e-02	1.3e-02
(0.5,0.6)	4.8e-01	2.4e-01	4.3e-02	8.6e-03
(0.6,0.7)	4.6e-01	4.9e-02	2.0e-02	4.4e-03
(0.7,0.8)	5.1e-01	3.0e-02	1.1e-02	2.3e-03
(0.8,0.9)	5.2e-01	7.9e-03	7.8e-03	1.3e-03
(0.9,1)	6.7e-01	3.2e-03	3.0e-03	1.0e-05

Fig. 6. Ring-chart indicating the minimization effectiveness $\Pi(\ell)$ of various localization algorithms in networks with $N_A = 4$ anchors and $N_T = 10$ under various values of meshness ratios $m \in [0, 1]$. All minimization techniques operate under the same tolerances of 10^{-8} on the variations of objective function and the estimates.

of 10^{-3}, similar to that of the SDP. The few instances when the NLS and SDP appear to have better minimization effectiveness in high SNR, as indicated in the ring chart (e.g., 6% and 8% with $\sigma = 0.25$), are attributed to numerical issues since the average minimizer deviation is 10^{-4}.

The comparison shows that the R-GDC is consistently the best, achieving the lowest point of the ML-WLS objective amongst the alternatives.

Next consider the performance of the algorithms studied in the network localization scenario. To this end, we repeat the experiments described above, but with $N_T = 10$ targets. It is well-known that the performances of localization algorithms are less sensitive to noise in network localization scenarios then in source localization scenarios, due the "diversity" available in the network case [22], [43]. Consequently, when comparing the localization algorithms of interest a fixed noise standard deviation of $\sigma = 1$ was maintained, while the connectivity range was varied from 8 to 25. The relative amount of connectivity achieved by a network (i.e., the number of facets formed in the network graph) under a given connectivity range can be measured by its *meshness ratio*, defined as [43], [66].

Definition D8 (Meshness Ratio): Let E_F and $E(r)$ denote the complete set of links of a fully connected network and the set

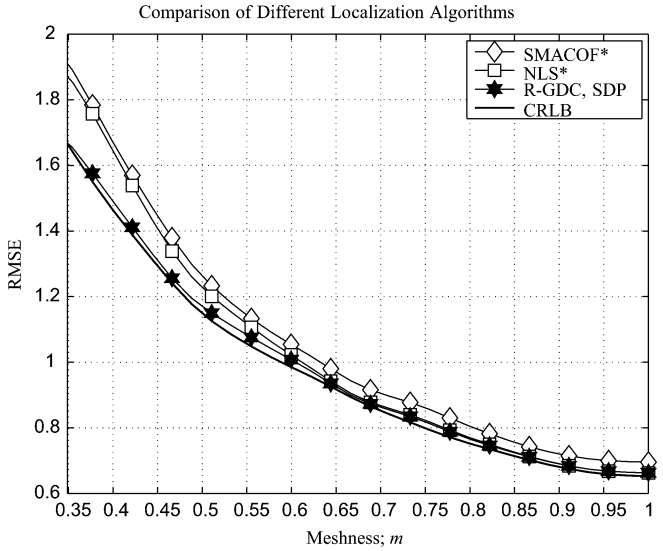


Fig. 7. Performance of different localization algorithms in scenarios with $N_A = 4$ anchors and $N_T = 10$ targets randomly distributed in a 10-by-10 square. Due to the lower sensitivity to noise experienced by network localization scenarios (which benefit from the redundancy introduced by the larger number of pairwise distances measurements), plots are shown as function of the meshness ratio. The noise standard deviation is $\sigma = 1$ and, for each m a total of 500 network configurations and 5 noisy distance measurement realizations were generated.

of links under a given connectivity range r. Then, the **meshness ratio** is defined as

$$m(r) \triangleq \frac{(|E(r)| - N + 1)}{(|E_F| - N + 1)} \tag{76}$$

where $|\cdot|$ denotes the cardinality of a set.

Fig. 6 is the equivalent of Fig. 5 for the network localization case, with the meshness ratio m playing the same role the noise standard deviation played. The advantage of the R-GDC method over the alternatives is once again clearly demonstrated. In fact the minimization efficiency of R-GDC reaches nearly 100% at high meshness-ratio. The impact of the meshness ratio on the performance of all algorithms, in terms of minimizer deviation can also be seen from the accompanying table. It can be seen that similarly to the source localization case, the R-GDC algorithm is the only one to achieve a minimizer deviation close 10^{-5}.

Compare the numbers in the table of Fig. 6 against the plots of Fig. 7, where the RMSE achieved by the algorithms are compared against one another and against the CRLB.

Both results indicate that R-GDC outperforms all alternatives, including the more computationally demanding SDP[5]. Next, in Fig. 8, we compare the performance of the **optimized** SMACOF* and NLS* algorithms against the R-GDC, as a function of the number of targets N_T with $m = 1$. Amongst these algorithms, the SMACOF* proves the least accurate, while the NLS* manages to approach the performance of R-GDC, albeit at the penalty of a higher complexity.

[5]In terms of absolute value of the cost function at the provided solution, the performances of the SDP and R-GDC are almost identical. This is why in Fig. 7 the curves of both algorithms overlap completely to the naked eye and therefore are shown with the same marker. Compared to one another, however, the R-GDC exhibiting a very minute, but consistent superiority, as evidenced by the Pie-charts of Figs. 5 and 6.

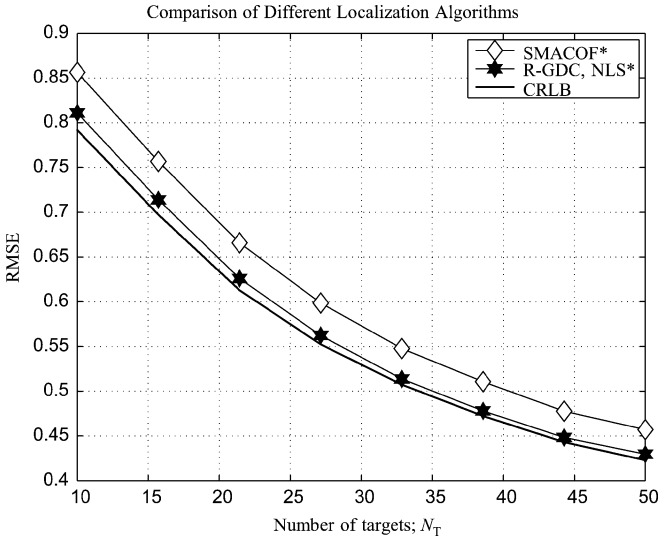


Fig. 8. Performance of localization algorithms in scenarios with $N_A = 4$ anchors are illustrated as a function of the number of targets N_T. The noise standard deviation is $\sigma = 1$ and the meshness ratio $m = 1$ (fully connected). The SDP is omitted due to the much larger complexity, which can be of the order $\mathcal{O}(N^6)$ as shown in [31].

Indeed, all R-GDC results for network localization scenarios were obtained with only 3 equi-spaced smoothing parameters, specifically $\lambda^{(1)}$ as in Theorem T4, $\lambda^{(2)} = \frac{\lambda^{(1)}}{2}$, and $\lambda^{(3)} = 0$.

In order to compare the complexity of the R-GDC against the alternatives, consider the following definition.

Definition D9 (Complexity Exponent): Let $\mathcal{C}(M)$ denote the complexity of a given optimization algorithm on M variables. Then, its associated **complexity exponent** is defined as

$$\varphi(M) \triangleq \log_M(E_{\tilde{\mathbf{D}},\mathbf{Z}}[\mathcal{C}(M)]) \tag{77}$$

where the expectation over $\tilde{\mathbf{D}}$ and $\mathbf{Z}$ refer to different noise and topology realizations, respectively.

From Definition D9 it follows that under the standard $\mathcal{O}$-notation the average complexity order of a given algorithm is $\mathcal{O}(M^{\varphi})$. The complexities[6] of the SMACOF*, NLS* and R-GDC algorithms are estimated as follows: [47]

$$\mathcal{C}_{\mathrm{SMACOF}}(M;t) \approx N^3 + t\left(N^2(2\eta+\frac{1}{2}) - N(\eta+\frac{1}{2})\right) \tag{78}$$

$$\mathcal{C}_{\mathrm{NLS}}(M;t) \approx t\left(\frac{M^3}{\eta^2} + (\frac{M^2}{\eta^2}+\frac{M}{\eta})(2N_A-1)+M\right) + r(t)\left(M^3 + \frac{N}{2}(N-1)\right) \tag{79}$$

$$\mathcal{C}_{\mathrm{R-GDC}}(M;t) \approx \frac{1}{2}\sum_{k=1}^{K} t_k\left(N(N-1)C_2 + 16M^2 - 2M\right) + r(t_k)N(N-1)C_1 \tag{80}$$

where t (t_k refers to the tth iteration in the kth minimization of the R-GDC method) and $r(t)$ refer, respectively, to the number of iterations of the main and inner loops, which are required to meet a prescribed stopping criterion that, as detailed in Appendixes C, B, and D, are given by the tolerance χ_Z on the variation of $\hat{\mathbf{X}}$, the tolerance χ_F on the variation of the objective function and the maximum number of iterations T_{MAX}.

[6]Details are provided in the Appendixes C, B and D.

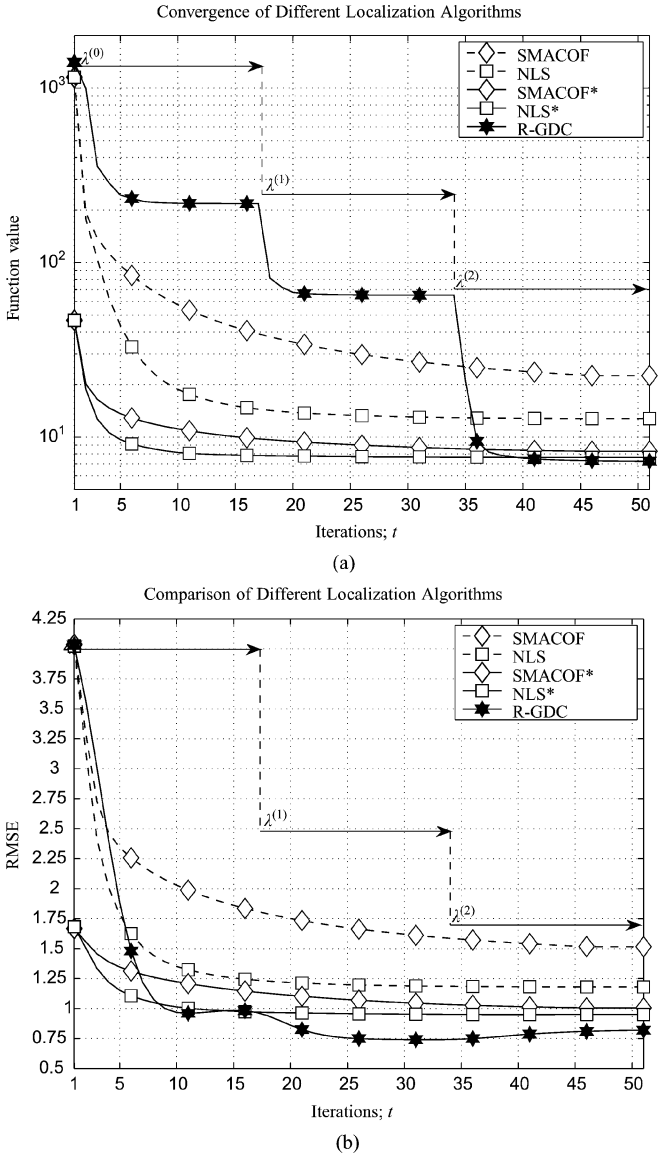


Fig. 9. Convergence of R-GDC, SMACOF, NLS, SMACOF* and NLS* algorithms. All algorithms operate under the same maximum number of iterations $T_{\mathrm{MAX}} = 51$, tolerances of 10^{-8} on the variations of objective function and the estimates. Each point is obtained out of 1000 network configurations with meshness $m \in [0.3, 0.4]$. R-GDC performs $K = 3$ iterations. (a) Function value versus Iterations. (b) RMSE versus Iterations.

Notice that the complexity of the algorithms of interest depend not only on their structure, which is mostly related to the number of variables M, but also on conditions such as noise and robustness to initial estimates, which in turn are captured by the number of iterations t. On the other hand, the convergence to a minimum is guaranteed by the utilization of standard optimization mechanisms [47], [67].

The convergence behavior of the R-GDC, SMACOF, NLS, SMACOF*, and NLS* algorithms are shown in Fig. 9.

The results are obtained with $T_{\mathrm{MAX}} = 51$, $\chi_Z = 10^{-8}$ and $\chi_F = 10^{-8}$, and together shown that the R-GDC outperforms

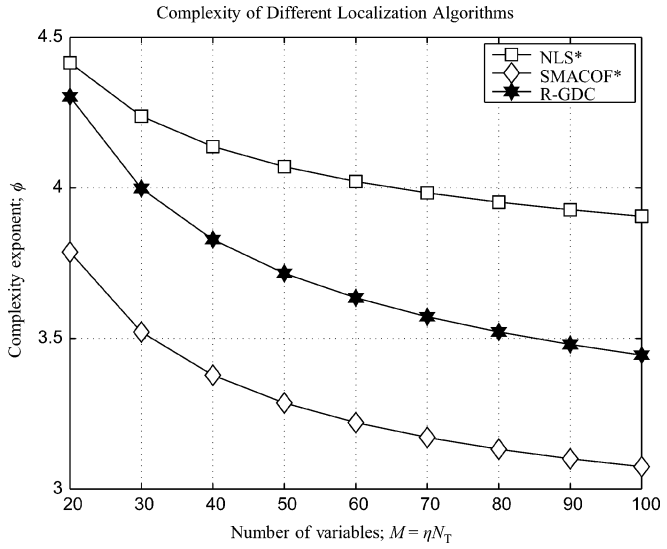


Fig. 10. Complexity exponents of different localization algorithms, as a function of the number of targets. All algorithms operate under the same maximum number of iterations $T_{\mathrm{MAX}} = 40$, tolerances of 10^{-8} on the variations of objective function and the estimates. Each point is obtained out of 100 network configurations and 6 noisy distance measurement realizations.

the alternatives (in terms of localization accuracy) already after as few as 9 iterations. Notice that the staircase convergence behavior of the R-GDC (in terms of objective minimization) seen in Fig. 9 is consistence with the strategy of the algorithm. This is in fact anther desirable feature of the method since it indicates that further gains in complexity reduction can be achieved by relaxing the continuation principle (see Theorem T1) down to a few iterations *per* smoothed objective.

In Fig. 10, instead, the complexity exponent of the R-GDC, NLS, and SMACOF algorithms computed with equations (79) through (78) and Definition D9 are compared against one another[7]. The results show that the R-GDC and SMACOF have comparable computational complexity, which is about $\mathcal{O}(M^{3.5})$ for very large networks.

Recall, however, that the SMACOF is also found to be the worst-performing algorithm in the comparisons (see Fig. 5 through 7), and that in fact only the significantly more complex SDP technique—which is also not ML-compliant—is comparable to the R-GDC in terms of accuracy and statistical coherence. In conclusion, based on this and the other results, it can be safely be stated that the proposed R-GDC algorithm exhibits superior performance to that of the SDP, with complexity inferior to that of the classic NLS, and comparable to that of the SMACOF.

V. Conclusions

We consider the distance-based ML formulation of the source and network localization problems, and propose an efficient and effective algorithm to minimize the corresponding WLS objective.

Unlike preceding methods, the proposed algorithms, dubbed the Range-Global Distance Continuation (R-GDC) holds true to the ML principle by adhering to the formulation directly derived form the likelihood function and by yielding, given an observation, the same result despite of initialization. The R-GDC method is shown via extensive comparison to outperform existing the best and most popular alternatives, in particular, the SDP, the NLS, and the SMACOF techniques, while exhibiting a computational complexity comparable to that of the SMACOF algorithm, which is widely known for its efficiency.

[7]Notice, however, that the results relative to the SDP technique are omitted since, for large scale networks the runtime of the algorithm was very long given that the computational complexity grows as $\mathcal{O}(N_T^6)$

Appendix A
Properties of $S_3(a,b,u)$

Lemma L4 (Properties of $s_3(a,b,u)$): Define the auxiliary function $S_3(a,b,u) \triangleq \exp(-u)\cdot_1 F_1(a;b;u)$, where $(a,b,u) > 0$. This function has the following properties:

i) Monotonicity: $sign(\frac{\mathrm{d}}{\mathrm{d}u}S_3(a,b,u)) = sign(a-b)$;
ii) Confinement: $0 \leq S_3(a,b,u) \leq 1$ for $a < b$.

Proof: Define the auxiliary function $S_3(a,b,u) \triangleq \exp(-u)\cdot_1 F_1(a;b;u)$, where $(a,b,u) > 0$. This function has the following properties:

i) Monotonicity: $sign(\frac{\mathrm{d}}{\mathrm{d}u}S_3(a,b,u)) = sign(a-b)$;
ii) Confinement: $0 \leq S_3(a,b,u) \leq 1$ for $a < b$.

In order to prove the first property, we study the derivative of $S_3(a,b,u)$. Invoking the recurrence relation of (48), we obtain

$$\frac{\partial S_3(a,b,u)}{\partial u} = \frac{(a-b)}{b}\exp(-u)\cdot_1 F_1(a;b+1;u). \quad (81)$$

Since $\exp(-u)\cdot_1 F_1(a;b+1;u)$ is always positive, (81) implicates in property i. Finally, recognize [48, p. 514] that $S_3(a,b,0) =_1 F_1(a;b+1;0) = 1$ and that $\exp(-u)\cdot_1 F_1(a;b+1;u)$ is always positive, such that $\lim_{u\longrightarrow\infty} S_3(a,b,u) = 0^+$ for $a < b$. □

Appendix B
Pseudocode of the SMACOF Algorithm

Algorithm 1 SMACOF localization

1:Get the matrices $\tilde{\mathbf{D}} \in \mathbb{R}^{N\times N}$, $\mathbf{W} \in \mathbb{R}^{N\times N}$ and $\hat{\mathbf{Z}}^0 \in \mathbb{R}^{N_T\times\eta}$

2:Set the Stop-Criteria $\chi_Z = 10^{-8}$, $\chi_F = 10^{-8}$ and $T_{\mathrm{MAX}} = 51$;

3:Compute $\mathbf{H} \in \mathbb{R}^{N\times N}$ as

$$h_{ij} \triangleq \begin{cases} \sum_{q=1}^{N} w_{iq} - w_{ii}, & \text{if } i = j, \\ -w_{ij}, & i \neq j; \end{cases}$$

4:Compute $\mathbf{H}^{\dagger}$;

5:$\hat{\mathbf{X}}^0 \triangleq [\mathbf{a}_1;\cdots;\mathbf{a}_{N_A};\hat{\mathbf{Z}}^0] \in \mathbb{R}^{N\times\eta}$;

6:$t = 0$;

7:**repeat**

8:$t \leftarrow t+1$;

9:Compute $\mathbf{A} \in \mathbb{R}^{N\times N}$ as

$$a_{ij} = \begin{cases} \sum_{\substack{i=1\\ i\neq j}}^{N} a_{ij}, & i = j, \\ w_{ij}^2 \cdot \frac{\tilde{d}_{ij}}{\|\hat{\mathbf{x}}_i^{t-1} - \hat{\mathbf{x}}_j^{t-1}\|_F}, & i \neq j; \end{cases}$$

10:$\hat{\mathbf{X}}^t \leftarrow \mathbf{H}^\dagger \cdot \mathbf{A} \cdot \hat{\mathbf{X}}^{t-1}$;

11:Evaluate $f_{\mathrm{R}}(\hat{\mathbf{X}}^t)$;

12:Compute $\Delta_{\mathbf{X}} \triangleq \|\hat{\mathbf{X}}^t - \hat{\mathbf{X}}^{t-1}\|_F$, $\Delta_{\mathbf{f}} \triangleq |f_{\mathrm{R}}(\hat{\mathbf{X}}^t) - f_{\mathrm{R}}(\hat{\mathbf{X}}^{t-1})|$;

13:**until** $(\Delta_{\mathbf{X}} \leq \chi_X)\,\&\&\,(t \geq T_{\mathrm{MAX}})\,\&\&\,(\Delta_{\mathbf{f}} \leq \chi_F)$.

As in the NLS algorithm, the initial estimate $\hat{\mathbf{Z}}^0$ used in the SMACOF-based localization technique is obtained via the combination of the Nyström and shortest-path completion algorithms described in [65].

In order to reduce the computational complexity of this method, we employ the Guttman transform for the update of the estimate $\hat{\mathbf{X}}$. In so doing, the total number of flops accounts for a) the calculation of the pseudoinverse $\mathbf{H}^\dagger$, the update $\hat{\mathbf{X}}$, and c) the evaluation of the objective function. Specifically, the total computational cost can be approximated as follows

$$\mathcal{C}_{\mathrm{SMACOF}}(M;t) \approx \underbrace{N^3}_{\text{line 4}} + t\left(\underbrace{\eta N(2N-1)}_{\text{line 10}} + \underbrace{\frac{N}{2}(N-1)}_{\text{line 11}}\right) \tag{82}$$

$$= N^3 + t\left(N^2(2\eta+\frac{1}{2}) - N(\eta+\frac{1}{2})\right). \tag{83}$$

Appendix C
Pseudocode of the NLS Algorithm

Algorithm 1 NLS localization

1:Get the matrices $\tilde{\mathbf{D}} \in \mathbb{R}^{N\times N}$, $\mathbf{W} \in \mathbb{R}^{N\times N}$ and $\hat{\mathbf{Z}}^0 \in \mathbb{R}^{N_T\times\eta}$

2:Set the Stop-Criteria $\chi_Z = 10^{-8}$, $\chi_F = 10^{-8}$ and $T_{\mathrm{MAX}} = 51$;

3:**repeat**

4:$t \leftarrow t+1$;

5:Evaluate $[\hat{\mathbf{g}}^{t-1}]_u = \hat{g}_{ij}^{t-1} \triangleq \sqrt{w}_{ij}(\tilde{d}_{ij} - \|\hat{\mathbf{p}}_i^{t-1} - \hat{\mathbf{p}}_j^{t-1}\|_F)$, $\forall i = 1,\ldots,N$, $j > N_A$;

6:Evaluate $\mathbf{J}_{t-1} \triangleq \frac{\partial \hat{\mathbf{g}}^{t-1}}{\partial \hat{\mathbf{z}}}$;

7:Compute $\mathbf{J}_{t-1}^T \cdot \mathbf{g}^{t-1}$;

8:**repeat**

9:$r \leftarrow r+1$;

10:Compute $\mathbf{\Delta_z}$ via the preconditioned conjugate gradient (PCG) procedure;

11:Evaluate $f_{\mathrm{R}}(\hat{\mathbf{Z}}^{t-1} + \mathbf{\Delta_z})$;

12:**until** $f_{\mathrm{R}}(\hat{\mathbf{Z}}^{t-1} + \mathbf{\Delta_z}) < f_{\mathrm{R}}(\hat{\mathbf{Z}}^{t-1})$

13:$\hat{\mathbf{Z}}^t \leftarrow \hat{\mathbf{Z}}^{t-1} + \mathbf{\Delta_z}$;

14:Compute $\Delta_{\mathbf{f}} \triangleq |f_{\mathrm{R}}(\hat{\mathbf{X}}^t) - f_{\mathrm{R}}(\hat{\mathbf{X}}^{t-1})|$;

15:**until** $(\mathbf{\Delta_z} \leq \chi_Z)\,\&\&\,(t \geq T_{\mathrm{MAX}})\,\&\&\,(\Delta_{\mathbf{f}} \leq \chi_F)$

The initial estimate $\hat{\mathbf{Z}}^0$ can be computed via the combination of the Nyström and shortest-path completion algorithms described in [65]. To the bet of our knowledge, this technique is both computationally efficient and the estimate is sufficiently reliable.

In order to estimate the computational complexity of the NLS-based localization algorithm, we consider: a) the total number of flops for the evaluations of the objective function function, the Jacobian evaluations and b) the number of elementary operations required to perform the PCG operation. Specifically, the estimated computational cost is given by

$$\mathcal{C}_{\mathrm{NLS}}(M;t) \approx t\left(\underbrace{\frac{M^2}{\eta^2} + \frac{M}{\eta}(2N_A-1)}_{\text{line 5+line 6}} + \underbrace{\frac{M^3}{\eta^2} + \frac{2M^2}{\eta}(N_A-1) + M}_{\text{line7}}\right) + r(t)\left(\underbrace{M^3}_{\text{line 10}} + \underbrace{\frac{N}{2}(N-1)}_{\text{line 11}}\right) \tag{84}$$

$$= t\left(\frac{M^3}{\eta^2} + \left(\frac{M^2}{\eta^2} + \frac{M}{\eta}\right)(2N_A-1) + M\right) + r(t)\left(M^3 + \frac{N}{2}(N-1)\right). \tag{85}$$

Appendix D
Pseudocode of the R-GDC Algorithm

Algorithm 1 R-GDC localization

1:Get the matrices $\tilde{\mathbf{D}} \in \mathbb{R}^{N\times N}$, $\mathbf{W} \in \mathbb{R}^{N\times N}$;

2:Compute the set of smoothing parameters $\{\lambda^{(k)}\}$ with $\lambda^{(1)} = \frac{\sqrt{\pi}}{2}\max_{ij\in\mathcal{H}} \tilde{d}_{ij}$;

3:Choose a random initial point $\hat{\mathbf{Z}}^0 \in \mathbb{R}^{N_T\times\eta}$;

4:Set the Stop-Criteria $\chi_Z = 10^{-8}$, $\chi_F = 10^{-8}$, and $T_s = \lceil \frac{T_{\mathrm{MAX}}}{K} \rceil$;

5:**for** $k = 1$ to K **do**

6:$\lambda \leftarrow \lambda^{(k)}$;

7:$t \leftarrow 0$, $\mathbf{B}^0 \leftarrow \mathbf{I}$;

8:**repeat**

9:$t \leftarrow t+1$;

10:Evaluate $\nabla_{\hat{\mathbf{Z}}}\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\boldsymbol{z}}^{t-1})$;

11:Compute the search direction $\mathbf{p}^t = -\mathbf{B}^{t-1} \cdot \nabla_{\hat{\mathbf{Z}}}\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\boldsymbol{z}}^{t-1})$;

12:**repeat**

13:$r \leftarrow r+1$;

14:Select the step-length α^r (backtracking line-search method)

15:$\hat{\boldsymbol{z}}^r = \hat{\boldsymbol{z}}^{t-1} + \alpha^r \mathbf{p}^t$;

16:Evaluate $\langle f_{\mathrm{R}}\rangle_\lambda(\hat{\boldsymbol{z}}^r)$;

17:**until** $\langle f_{\rm R}\rangle_\lambda(\hat{\boldsymbol{z}}^r) < \langle f_{\rm R}\rangle_\lambda(\hat{\boldsymbol{z}}^{t-1})$;

18:$\hat{\boldsymbol{z}}^t \leftarrow \hat{\boldsymbol{z}}^r$; $\mathbf{s} = \hat{\boldsymbol{z}}^t - \hat{\boldsymbol{z}}^{t-1}$;

19:$\mathbf{y} = \nabla_{\hat{\mathbf{Z}}}\langle f_{\rm R}\rangle_\lambda(\hat{\boldsymbol{z}}^t) - \nabla_{\hat{\mathbf{Z}}}\langle f_{\rm R}\rangle_\lambda(\hat{\boldsymbol{z}}^{t-1})$;

20:$\mathbf{B}^t = \mathbf{B}^{t-1} - \frac{\mathbf{B}^{t-1}\mathbf{s}\mathbf{s}^T\cdot\mathbf{B}^{t-1}}{\mathbf{s}^T\mathbf{B}^{t-1}\mathbf{s}} + \frac{\mathbf{y}\mathbf{y}^T}{\mathbf{y}^T\mathbf{s}}$;

21:Compute $\Delta_{\mathbf{z}} \triangleq \|\hat{\mathbf{Z}}^t - \hat{\mathbf{Z}}^{t-1}\|_F$, $\Delta_{\mathbf{f}} \triangleq |\langle f_{\rm R}\rangle_\lambda(\hat{\boldsymbol{z}}^t) - \langle f_{\rm R}\rangle_\lambda(\hat{\boldsymbol{z}}^{t-1})|$;

22:**until** $(\boldsymbol{\Delta}_{\mathbf{z}} \le \chi_Z)\,\&\&\,(t \ge T_s)\,\&\&\,(\Delta_{\mathbf{f}} \le \chi_F)$

23:$\mathbf{Z}^0 \leftarrow \mathbf{Z}^t$;

24:**end for**

We assume that the total number of iterations are the same for all algorithms, therefore, the kth minimization in the R-GDC algorithm is restricted to $T_s = \left\lceil \frac{T_{\rm MAX}}{K} \right\rceil$ sub-iterations, where $\lceil\rceil$ denote the integer supremum. Furthermore, the R-GDC algorithm does not require any specific initial estimate, indeed, we utilize a random starting point $\hat{\mathbf{Z}}^0$.

In order to evaluate the computational complexity, the minimization of the smoothed objective is implemented via Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm with a line-search. In so doing, no matrix inversions are involved and an estimate of the overall number of flops reduce to the those required for a) the calculation of the search direction, b) the evaluation of the gradient and the objective function and c) in the update of the matrix $\mathbf{B} \in \mathbb{R}^{M\times M}$. Specifically, the estimated computational cost is given by

$$\begin{aligned}
&\mathcal{C}_{\rm R-GDC}(M;t)\\
&\quad\approx \sum_{k=1}^{K} t_k \left(\underbrace{\frac{N}{2}(N-1)C_2}_{\text{line 10}} + \underbrace{2M^2 - M}_{\text{line 11}} + \underbrace{6M^2}_{\text{line 21}} \right)\\
&\qquad + r(t_k)\underbrace{\frac{N}{2}(N-1)C_1}_{\text{line 14}} \qquad (86)\\
&\quad= \frac{1}{2}\sum_{k=1}^{K} t_k \left((N^2-N)C_2 + 16M^2 - 2M\right)\\
&\qquad + r(t_k)(N^2-N)C_1 \qquad (87)
\end{aligned}$$

where $r(t_k)$ is the number of iterations in the inner loop obtained with $\lambda^{(k)}$, C_1 and C_2 are the costs for the calculation (i.e., (38) for $s \le 10$ with $m = 50$ and (39) with $R = 6$ otherwise) of the Hypergeometric functions of (25) and (40), respectively.

REFERENCES

[1] G. Destino and G. Abreu, "Solving the source localization prolem via global distance continuation," in *Proc. IEEE Int. Conf. Commun.*, 2009.

[2] C. Harrison, J. Wiese, and A. K. Dey, "Achieving ubiquity: The new third wave," *IEEE MultiMedia*, vol. 17, no. 3, pp. 8–12, Jul.–Sep. 2010.

[3] S. Poslad, *Ubiquitous Computing: Smart Devices, Environments and Interactions, ser. 978-0-470-03560-3*. New York: Wiley, 2009.

[4] J. Hightower and G. Borriello, "Location systems for ubiquitous computing," *IEEE Comput.*, vol. 34, no. 8, pp. 57–66, Aug. 2001.

[5] M. Vossiek, L. Wiebking, P. Gulden, J. Wieghardt, C. Hoffmann, and P. Heide, "Wireless local positioning," *IEEE Microw. Mag.*, vol. 4, no. 4, pp. 77–86, Dec. 2003.

[6] K. Scott and R. Benlamri, "Context-aware services for smart learning spaces," *IEEE Trans. Learn. Technol.*, vol. 3, no. 3, pp. 214–227, Jul. –Sep. 2010.

[7] E. Kaplan and C. Hegarty, *Understanding GPS: Principles and Applications*, 2nd ed. Reading, MA: Artech House, 2006.

[8] M. Hyder and K. Mahata, "Direction-of-arrival estimation using a mixed $\ell_{2,0}$ norm approximation," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4646–4655, Sep. 2010.

[9] P. Stoica and K. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 38, no. 7, pp. 1132–1143, Jul. 1990.

[10] S. D. Chitte, S. Dasgupta, and Z. Ding, "Distance estimation from received signal strength under log-normal shadowing: Bias and variance," *IEEE Signal Process. Lett.*, vol. 16, no. 3, pp. 216–218, 2009.

[11] D. Dardari, C. Chong, and M. Z. Win, "Threshold-based time-of-arrival estimators in UWB dense multipath channels," *IEEE Trans. Commun.*, vol. 56, no. 8, pp. 1366–1378, Aug. 2008.

[12] M. Spirito, "On the accuracy of cellular mobile station location estimation," *IEEE Trans. Veh. Technol.*, vol. 50, no. 3, 2001.

[13] N. Patwari, R. J. O. Dea, and Y. Wang, "Relative location estimation in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2137–2148, Aug. 2003.

[14] Q. Yihong, H. Suda, and H. Kobayashi, "On time-of-arrival positioning in a multipath environment," in *Proc. IEEE 60th Veh. Technol. Conf. (VTC'04 Fall)*, 2004, vol. 5, pp. 3540–3544.

[15] 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," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 70–84, 2005.

[16] A. M.-C. So and Y. Ye, "Theory of semidefinite programming for sensor network localization," in *Proc. ACM-SIAM 16th Ann. Symp. Discr. Algorithms*, 2005, pp. 405–414.

[17] S. Al-Homidan and H. Wolkowicz, "Approximate and exact completion problems for euclidean distance matrices using semidefinite programming," *Linear Algebra and Its Appl.*, vol. 406, no. 1–3, pp. 109–141, Sep. 2005.

[18] G. Destino and G. T. F. de Abreu, "Sensor localization from WLS optmization with closed-form Gradient and Hessian," in *Proc. IEEE 49th Ann. Globecom Conf. (GLOBECOM'06)*, Dec. 1, 2006, pp. 1–6.

[19] A. J. Weiss and J. S. Picard, "Network localization with biased range measurements," *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, pp. 298–304, Jan. 2008.

[20] K. Yu and Y. Guo, "Improved positioning algorithms for nonline-of-sight environments," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2342–2353, Jul. 2008.

[21] I. Guvenc, S. Gezici, F. Watanabe, and H. Inamura, "Enhancements to linear least squares localization through reference selection and ML estimation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar.–Apr. 2008, pp. 284–289.

[22] H. Wymeersch, J. Lien, and M. Win, "Cooperative localization in wireless networks," *IEEE Proc.*, vol. 97, no. 2, pp. 427–450, 2009.

[23] Y. Shen and M. Z. Win, "Fundamental limits of wideband localization—Part I: A general framework," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4956–4980, 2010.

[24] A. A. Kannan, M. Guoqiang, and B. Vucetic, "Simulated annealing based localization in wireless sensor network," in *Proc. IEEE Conf. Local Comput. Netw. 30th Anniv.*, Sydney, Australia, 2005, pp. 513–514.

[25] P. E. Gill and W. Murray, "Algorithms for the solution of the nonlinear least-squares problem," *SIAM J. Numer. Anal.*, vol. 15, no. 5, pp. 977–992, 1978.

[26] Y. Shang and W. Ruml, "Improved MDS-based localization," in *Proc. IEEE 23rd Ann. Joint Conf. IEEE Comput. Commun. Soc. (INFOCOM'04)*, Hong-Kong, China, Mar. 2004, vol. 4, pp. 2640–2651.

[27] B. Denis, L. He, and L. Ouvry, "A flexible distributed maximum log-likelihood scheme for UWB indoor positioning," in *Proc. IEEE 4th Workshop on Position., Navigat. Commun.*, 2007, pp. 77–86.

[28] J. de Leeuw, "Convergence of the majorization method for multidimensional scaling," *J. Classific.*, vol. 5, pp. 163–180, 1988.

[29] J. A. Costa, N. Patwari, and A. O. Hero, "Distributed multidimensional scaling with adaptive weighting for node localization in sensor networks," *ACM J. Sens. Netw.*, vol. 2, no. 1, pp. 39–64, Feb. 2006.

[30] P. Biswas, T.-C. Liang, K.-C. Toh, T.-C. Wang, and Y. Ye, "Semidefinite programming approaches for sensor network localization with noisy distance measurements," *IEEE Trans. Autom. Sci. Eng.*, vol. 3, no. 4, pp. 360–371, Oct. 2006.

[31] P. Biswas, T.-C. Liang, K.-C. Toh, and T.-C. Wang, "Semidefinite programming based algorithms for sensor network localization with noisy distance measurements," *ACM Trans. Sens. Netw.*, vol. 2, no. 2, pp. 188–220, May 2006.
[32] J. More and Z. Wu, "Global continuation for distance geometry problems," *SIAM J. Optimiz.*, vol. 7, pp. 814–836, 1997.
[33] A. Beck, P. Stoica, and J. Li, "Exact and approximate solutions for source localization problems," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1770–1778, May 2008.
[34] C. T. Huang, C. H. Wu, Y. N. Lee, and J. T. Chen, "A novel indoor RSS-based position location algorithm using factor graphs," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 3050–3058, 2009.
[35] G. Destino and G. Abreu, "Reformulating the least-square source localization problem with contracted distances," in *Proc. IEEE 43th Asilomar Conf. Signals, Syst. Comput.*, 2009.
[36] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, ser. Signal Process. Ser.. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[37] D. Macagnano and G. Abreu, "Improved MDS-based multi-target tracking algorithm," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2009, pp. 1–6.
[38] A. Savvides, H. Park, and M. B. Srivastava, "The n-hop multilateration primitive for node localization problems," *Mobile Netw. Appl.*, vol. 8, pp. 443–451, 2003.
[39] E. Larsson and D. Danev, "Accuracy comparison of LS and squared-range LS for source localization," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 916–923, Feb. 2010.
[40] 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.
[41] H. Wang, L. Yip, K. Yao, and D. Estrin, "Lower bounds of localization uncertainty in sensor networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '04)*, May 17–21, 2004, vol. 3, pp. 917–20.
[42] L. Joon-Yong and R. Scholtz, "Ranging in a dense multipath environment using an UWB radio link," *IEEE J. Sel. Areas in Commun.*, vol. 20, pp. 1667–1683, Dec. 2002.
[43] G. Destino and G. T. F. De Abreu, "Weighing strategy for network localization under scarce ranging information," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3668–3678, 2009.
[44] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[45] J. Dattorro, *Convex Optimization and Euclidean Distance Geometry*. New York: Meboo, 2005.
[46] G. H. Golub and C. F. van Loan, *Matrix Computations*, 3rd ed. New York: Johns Hopkins Univ. Press, 1996.
[47] J. Nocedal and S. Wright, *Numerical Optimization*. New York: Springer, 2006.
[48] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 10th ed. New York: Dover, 1965.
[49] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 6th ed. New York: Academic, 2000.
[50] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2000.
[51] P. Moon and D. E. Spencer, *Field Theory Handbook, Including Coordinate Systems, Differential Equations, and Their Solutions*. New York: Springer-Verlag, 1988.
[52] I. C. F. Ipsen and B. Nadler, "Refined perturbation bounds for eigenvalues of Hermitian and non-Hermitian matrices," *SIAM J. Matrix Anal. Appl.*, vol. 31, no. 1, pp. 40–53, Feb. 2009.
[53] J. Chaffee and J. Abel, "GDOP and Cramer-Rao bound," in *Proc. IEEE Position Location Navigat. Symp., ser. 0-7803-1435-2*, Apr. 11–15, 1994, pp. 663–668.
[54] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal, "Locating the nodes: Cooperative localization in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 54–69, 2005.
[55] J. S. Abel, "A bound on the measn-square-estimate error," *IEEE Trans. Inf. Theory*, vol. 39, no. 5, pp. 1675–1680, Sep. 1993.
[56] A. Renaux, P. Forster, P. Larzabal, C. Richmond, and A. Nehorai, "A fresh look at the Bayesian bounds of the Weiss-Weinstein family," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5334–5352, Nov. 2008.
[57] A. A. Kannan, G. Mao, and B. Vucetic, "Simulated annealing based wireless sensor network localization with flip ambiguity mitigation," in *Proc. IEEE 63rd Veh. Technol. Conf.*, Sep. 2006, vol. 2, pp. 1022–1026.
[58] S. Severi, G. T. F. de Abreu, G. Destino, and D. Dardari, "Understanding and solving flip-ambiguity in network localization via semidefinite programming," in *Proc. IEEE Ann. Globecom Conf. (GLOBECOM'09)*, Nov. 4, 2009.
[59] A. Bhattacharyya, "On some analogues of the amount of information and their use in statistical estimation," *Indian J. Statist.—Sankhya*, vol. 8, no. 3, pp. 201–214, 1947.
[60] X. L. Y. B-Shalom and T. Kirubarajan, *Estimation with Application to Tracking and Navigation*. New York: Wiley-Intersci., 2001.
[61] H. Wymeersch, U. Ferner, and M. Z. Win, "Cooperative bayesian self-tracking for wireless networks," *IEEE Commun. Lett.*, vol. 12, no. 7, 2008.
[62] D. Macagnano and G. T. F. de Abreu, "Gershgorin analysis of random Gramian matrices with application to MDS tracking," *IEEE Trans. Signal Process.*, vol. 59, no. 4, 2011.
[63] H. Wang, K. Yao, and D. Estrin, Information-Theoretic Approaches for Sensor Selection and Placement in Sensor Networks for Target Localization and Tracking Center for Embedded Network Sensing—UCLA, 2005.
[64] D. J. Torrieri, "Statistical theory of passive location systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 20, no. 2, pp. 183–198, Mar. 1984.
[65] D. Macagnano and G. Abreu, "Gershgorin analysis of random Gramian matrices with application to MDS tracking," *IEEE Trans. Signal Process.*, Apr. 2010.
[66] C. Adams and R. Franzosa, *Introduction to Topology Pure and Applied*. New York: Pearson/Prentice-Hall, 2008.
[67] T. F. Cox and M. A. A. Cox, *Multidimensional Scaling*, 2nd ed. New York: Chapman and Hall/CRC, 2000.

Giuseppe Destino (S'07) received two M.Sc. degrees simultaneously from the Politecnico di Torino, Italy, and the University of Nice, France, in 2005.

From February 2003 to February 2004, he was a student researcher in wireless communications at the Research Institute Eurécom, Sophia-Antipolis, France. From March 2004 to March 2005, he was a research assistant with the Centre for Wireless Communications, University of Oulu, Oulu, Finland, where he is now pursuing the Ph.D. degree. His research interests are localization (positioning) algorithms and the application of convex optimization and graph theory to wireless communications.

Giuseppe Abreu (S'99–M'04–SM'09) received the B.Eng. degree in electrical engineering and a specialization (Latu Sensu) degree in telecommunications engineering from the Universidade Federal da Bahia (UFBa), Salvador, Bahia, Brazil, in 1996 and 1997, respectively. After some time in industry, he joined the Faculty of Engineering of the Yokohama National University, Japan, where he received the M.Eng. and Ph.D. degrees in physics, electrical, and computer engineering in March 2001, and March 2004, respectively.

In April 2004, he joined the Centre for Wireless Communications at the University of Oulu, Finland, and became an Adjunct Professor on Statistical Signal Processing and Communications Theory for Multi-dimensional and Distributed Wireless Systems with the Department of Electrical and Information Engineering in May 2006. He has published more than one hundred scientific articles in reputable journals and international conferences. His research interests include localization and tracking algorithms, communications theory, statistical modeling, signal processing, waveform design, space-time coding/decoding, random matrices, parameter estimation, beam pattern synthesis, ultra-wideband communications, and wireless secrecy/security.

Prof. Abreu is an Editor of the IEEE TRANSACTIONS OF WIRELESS COMMUNICATIONS, permanent General Co-chair of the IEEE Workshop on Positioning, Navigation and Communications, and a Member of the IEEE Communications Theory Technical Committee for his role as Technical Program Co-chair of the Communication Theory Workshop in 2011.