Pearson–based Analysis of Positioning Error Distribution in Wireless Sensor Networks

Stefano Tennina\(^1\), Marco Di Renzo\(^1\), Fabio Graziosi\(^1\), Fortunato Santucci\(^1\)

\(^1\) Dept. of Electrical and Information Engineering, Center of Excellence DEWS, Univ. of L’Aquila, 67040 L’Aquila, Italy
\(^2\) Telecommunications Technological Center of Catalonia, Av. Canal Olímpic s/n, 08860 Castelldefels, Barcelona, Spain
E–Mail: \{stennina, graziosi, santucci\}@ing.univaq.it, marco.di.renzo@cttc.es

Abstract—In two recent contributions [1], [2], we have provided a comparative analysis of various optimization algorithms, which can be used for atomic location estimation, and suggested an enhanced version of the Steepest Descent (ESD) algorithm, which we have shown to be competitive with other distributed localization algorithms in terms of estimation accuracy and numerical complexity. Moreover, therein we have conducted a preliminary statistical characterization of the positioning error distribution, by showing that it can be well approximated by the family of Pearson distributions, as well as pointed out that its knowledge may be efficiently used to speed–up the analysis of iterative–based positioning algorithms by avoiding the need of simulating the whole location discovery algorithm, and allowing simulation at the atomic level only. In this contribution, based on the preliminary results shown in [1], we propose a comprehensive statistical analysis of the positioning error distribution for the ESD algorithm, by providing the parameters of the Pearson fitting distribution with respect to two important design factors for Wireless Sensor Networks (WSNs): i) the ranging error standard deviation, which represents the input parameter for every localization algorithm, and ii) the geometric dilution of precision factor, which provides a simple parameter to account for different network topologies. In particular, we report an extensive number of simulation results that may provide important insights to the family of Pearson distributions, as well as pointed out that its knowledge may be efficiently used to speed–up the analysis of iterative–based positioning algorithms by avoiding the need of simulating the whole location discovery algorithm, and allowing simulation at the atomic level only. In this contribution, based on the preliminary results shown in [1], we propose a comprehensive statistical analysis of the positioning error distribution for the ESD algorithm, by providing the parameters of the Pearson fitting distribution with respect to two important design factors for Wireless Sensor Networks (WSNs): i) the ranging error standard deviation, which represents the input parameter for every localization algorithm, and ii) the geometric dilution of precision factor, which provides a simple parameter to account for different network topologies. In particular, we report an extensive number of simulation results that may provide important insights to the system designer: i) allow a parametric analysis to figure out the joint effect of ranging errors and network topology on the performance of the localization algorithms, and ii) give a general framework for modeling the statistics of the positioning error, which may be used for network planning, as well as for the analysis and design of the upper layers of the protocol stack.

I. INTRODUCTION

Distributed Wireless Sensor Networks (WSNs) [3] have become increasingly popular in military and civilian applications, and have been proposed for a wide range of applications. The main purpose of a sensor network is to monitor an area, including detecting, identifying, localizing, and tracking one or more objects of interest. These networks may be used for i) military and homeland security purposes, as in surveillance, reconnaissance and combat scenarios or around the perimeter of a manufacturing plant for intrusion detection, and ii) civilian intentions, as in environmental and biometrics monitoring, disaster area monitoring and recovery, home automation, inventory management, logistic, and many others. The individual nodes incorporate wireless transceivers so that communication and networking are enabled. Moreover, these nodes should have small form factor, light weight, provide long time service with a limited energy source and be inexpensive, so that they can be deployed in large numbers. Additionally, the network is required to possess, in general, self–organizing capabilities, so that little or no human intervention for network deployment and setup is required.

In the frame of distributed WSNs, a fundamental component of self–organization is the ability of sensor nodes to “sense” their location in space [4]–[7]. In particular, the automatic location of sensors in WSNs represents a key enabling technology to allow and support a rich set of geographically aware protocols [8], and to accurately report the position of either detected targets or events [9], [10]. Over the last years, many distributed algorithms have been proposed and analyzed to address the problem of network location discovery (see, e.g., [7], [11]–[13]). However, in [7] and [13] the authors conclude that among the existing algorithms no one seems to perform better than the others, and they claim that the definition of location algorithms with accurate positioning capabilities and low communication and computation costs is still an ongoing research area.

Despite the positioning algorithms proposed to date exhibit advantages and disadvantages, a promising approach analyzed by many authors is the so–called “recursive positioning methods” [7], [12], [14]. Loosely speaking, recursive algorithms are often employed to overcome the limits related to the short–range communication capabilities of sensor nodes, by enabling the position estimation process to be composed by many subsequent steps/phases through which all the sensors in the network localize themselves in a distributed fashion. In particular, first of all, sensor nodes that have at least four (we consider the general 3D scenario) anchor nodes (i.e., nodes that are aware of their exact location, also called “startup anchors”) in their neighborhood (i.e., nodes within the radio range) activate a location discovery procedure to estimate their position; then, these localized nodes are converted (i.e., “converted anchors”) to the level of anchor nodes (with a certain degree of reliability [11], [12]), thus possibly enabling the activation of the location discovery procedure for sensor nodes not involved in this first phase. So, by applying the whole process iteratively, the converted anchor nodes are propagated through the network in the regions with a low

Work partially done in the framework of the HYCON Network of Excellence, contract number FP6–IST–511368.
Recursive positioning methods have several positive features, e.g., i) they appear to be a good solution for sensor nodes with limited range capabilities, ii) they may efficiently counteract the sparse anchor node problem, and iii) they are distributed by nature. However, they still present several critical design issues, e.g., i) in recursive approaches the positioning error may accumulate along the iterative process, thus severely corrupting the final estimates of sensor nodes located in remote areas (i.e., regions of the network where startup anchors are sparse) [14], and ii) some bad network topologies may introduce significant errors even with accurate distance estimates [11]. Moreover, iterative techniques typically require intensive network simulations to estimate the nodes’ position: in general, in order to estimate the position of a node in a remote area, the overall location discovery algorithm has to be simulated, such that error position accumulation can be taken into account. Of course, this procedure may be computational demanding when large networks have to be studied, and when various network topologies and algorithms have to be compared.

In order to cope with the first above–mentioned design issue, in [1] we have deeply investigated the problem of error accumulation, and proposed a comparative study of various optimization algorithms that can be used for distributed position estimation. In addition to the well–known Triangulation, Steepest Descent, Non–Linear Least Squares, and Conjugate Gradient algorithms [15], we have also proposed an enhanced version of the Steepest Descent (ESD), which we have shown to be superior to the other algorithms in terms of error accuracy, computational complexity (i.e., time required to estimate the final position), and algorithm initialization. Moreover, in [2] we have addressed the second important issue (i.e., intensive network simulations), and proposed a new methodology to avoid the need to simulate the whole location discovery algorithm for estimating the position of every node in the network, and account for error propagation as well. In particular, the approach proposed in [2] originates from the following remark: most network simulations rely on some priori statistical models to account for ranging errors [10], but only a few contributions [16], [17] have attempted to determine statistical models to characterize positioning errors and their propagation. However, if such a distribution (i.e., probability density function) was known, one could simulate the location discovery algorithm at the atomic level (i.e., by simulating only the atomic network composed by either four startup or converted anchors, and an un–localized sensor node) rather than running the whole localization algorithm, since the location error of the nodes involved in the atomic positioning procedure could be obtained from both its distribution and the network topology. In [2], we have also shown, with an example, that such a knowledge can be taken into account in the iterative running of the algorithm in a simple and effective way.

However, the new methodology proposed in [2] requires: i) to identify a family of distributions, which can fit the positioning error with good accuracy for most system setups, and ii) to estimate the parameters of this distribution with respect to the most important design parameters that affect the performance of the localization algorithm. Accordingly, the aim of the present paper is twofold: i) provide a parametric and comprehensive analysis of the positioning error with respect to two important design parameters for every localization algorithm, i.e., ranging error and network topology, and ii) determine the parameters of the fitting distribution for every combination of the latter two parameters. Using the results shown in this paper, the system designer may easily identify the statistics of the positioning error for various system setups, without the need to actually simulate the localization algorithm, and use this knowledge for resource management, routing, and application related issues [16], as well as define error control mechanisms, which may improve the accuracy and robustness of the network localization process on its own [17].

The reminder of the paper is organized as follows. Section II introduces the system model, the network scenario, and the main assumptions of the analysis. In Section III, the proposed Enhanced Steepest Descent (ESD) algorithm is described. In Section IV, the statistical distribution of the positioning error for the ESD algorithm is analyzed with respect to network topology and ranging error distribution. Finally, Section V concludes the paper.

II. SYSTEM DESCRIPTION

Let us consider $N_A$ wireless nodes $\{A_i\}_{i=1}^{N_A}$ distributed in the region of interest, whose exact locations in the considered scenario are known. These nodes will be denoted in the text as “startup anchors”. We also assume that in the same area $N_U$ wireless nodes $\{U_i\}_{i=1}^{N_U}$ with unknown location have been deployed. We will refer to these nodes as “unknown nodes”. These wireless nodes have a relatively simple radio interface to communicate among them, which allows not only data exchange but also distance measurements. The main goal of localization algorithms is to use the anchor nodes to somehow estimate the position of the unknown nodes in the specified coordinate system. In particular, position estimation algorithms require a minimum of either three or four anchor nodes in a two– and three–dimensional coordinate system, respectively [4].

A. Recursive Positioning Method

As discussed in Section I, we will consider a recursive positioning method for network location discovery. In particular, the well–known recursive and hierarchical method proposed in [12], [14] is analyzed for sake of illustration. The basic version of the algorithm in a 3D scenario involves the following steps:

1) Phase 1: Unknown nodes that are connected (i.e., they are in the neighborhood) to at least four “startup anchors” compute their position.

2) Phase 2: Once an unknown node has estimated its position, it becomes a “converted anchor” and broadcasts
its estimated position to other nearby unknown nodes, thus enabling them to estimate their positions.

3) Phase 3: This process is repeated until the positions of all the nodes that eventually can have either four “startup anchors” or “converted anchors” are estimated.

As a consequence, depending on the current step of the algorithm, the four reference nodes with known position may be either “startup anchors” or “converted anchors”, where the latter are nodes with unknown position at the beginning of the location discovery procedure, but which have localized themselves during one of the previous steps of the iterative algorithm. Of course, differently from the “startup anchors”, the position of the “converted anchors” is affected by a certain error. In what follows, we will denote with “reference nodes” both “startup” and “converted” anchors.

B. Position Computation

The recursive positioning method described in Section II-A requires a technique to compute the location of an unknown node from the position of four anchor nodes, which may be in part “startup anchors” and in part “converted anchors”. In general, the computation of the position of the unknown node involves two basic steps: i) measuring the distances between pairs of sensors, and ii) estimating the node’s position via some mathematical models.

With regard to distance estimation between pairs of nodes, we consider the time–of–flight technique [9]. In this method, the time–of–flight of a RF signal between the unknown node and a reference node is used to estimate the distance between them. In particular, after measuring the time–of–flight through ranging packet exchange, the distance can be obtained using the well–known speed–distance relationship [18]. Accurate ranging estimates can be obtained, e.g., using either Spread Spectrum or Ultra Wide Band technologies [10]. In particular, as shown in [10], the measured distances are noisy due to channel impairments, and to errors in distance estimation. Throughout the paper, the ranging error will be modeled as a Gaussian random variable with mean value given by the actual distance, and standard deviation denoted by $\sigma_R$.

With regard to position computation from range estimates, in the literature two basic mathematical models are often considered: i) triangulation, which foresees to estimate the position of the unknown node by finding the intersection of four spheres in a 3D environment, and ii) multilateration, according to which the estimated position is obtained via the minimization of an error cost function, depending on the difference between the actual measured distances and the estimated Euclidean distances between unknown and reference nodes. The main difference between the two approaches is that multilateration algorithms are more robust to noisy range measurements [7], even though they may suffer for the local minima problem. In Section III, we introduce an improved algorithm for position estimation, which is based on the multilateration principle.

C. Assumptions and Objectives of the Paper

As mentioned in Section I, the main objective of the present paper is to analyze the positioning error distribution of location algorithms, which can be efficiently used to evaluate the performance of position–aware protocols and applications, without requiring the simulation of the whole location discovery process [8]. Accordingly, in this contribution we mainly focus on the statistical characterization of positioning errors when considering the iterative location discovery algorithm at the atomic level rather than dealing with the whole network level algorithm. As a consequence, we will focus on the atomic step of the algorithm, which involves four reference nodes (composed by either “startup anchors” or “converted anchors”), and one unknown node. Of course, this assumption is not restrictive in our scenario since we have shown in [2] that the obtained results can be easily integrated into the whole recursive and hierarchical algorithm. Finally, the description of the simulation scenario, and the discussion of its generality to account for various network topologies can be found in Section IV-A.

III. AN IMPROVED OPTIMIZATION ALGORITHM FOR POSITIONING: THE ENHANCED STEEPEST DESCENT

In this section we introduce the proposed ESD algorithm. In particular, since the proposed method represents an improved version of the well–known Steepest Descent (SD), the latter one is briefly summarized first, and then compared with the proposed one. The following notation will be used: i) bold symbols are used to denote vectors and matrices, ii) $(\cdot)^T$ denotes transpose operation, iii) $\nabla(\cdot)$ is the gradient operator, iv) $\|\cdot\|$ is the Euclidean distance, v) $\angle(\cdot,\cdot)$ is the phase angle between two vectors, vi) $\hat{\mathbf{u}}_j = [\hat{u}_{j,x}, \hat{u}_{j,y}, \hat{u}_{j,z}]^T$ denotes the estimated position of the unknown node $U_j$, vii) $\mathbf{u}_j = [u_{j,x}, u_{j,y}, u_{j,z}]^T$ is the trial solution of the optimization algorithm for the unknown node $U_j$, viii) $\mathbf{u}_i = [x_i, y_i, z_i]^T$ are the positions of the reference nodes $\{A_i\}_{i=1}^4$, which are exact when the reference node is a “startup anchor”, and noisy when the reference node is a “converted anchor”, and ix) $\{d_{j,i}\}_{i=1}^4$ denotes the estimated (via ranging measurements) distance between reference node $\{A_i\}_{i=1}^4$ and the unknown node $U_j$.

A. Multilateration Methods

Both SD and ESD algorithms belong to the family of multilateration methods. In particular, in such a methods the position of an unknown node $U_j$ is obtained by minimizing the error cost function $F(\cdot)$ defined as follows:

$$F(\mathbf{u}_j) = \sum_{i=1}^4 \left( d_{j,i} - \|\mathbf{u}_j - \hat{\mathbf{u}}_i\| \right)^2$$

such that $\hat{\mathbf{u}}_j = \arg \min \{ F(\mathbf{u}_j) \}$. The minimization of (1) can be done using a variety of numerical optimization techniques, each one having its advantages and disadvantages in terms of accuracy, robustness, speed, complexity, and storage.
requirements [15]. Since optimization methods are iterative by nature, we will denote by the index \( k \) the \( k \)-th iteration of the algorithm, and with \( F(\mathbf{u}_j(k)) \) and \( \hat{\mathbf{u}}_j(k) \) the error cost function and the estimated position at the \( k \)-th iteration, respectively.

1) Steepest Descent (SD): The SD is an iterative line search method that allows to find the (local) minimum of the cost function in (1) at step \( k + 1 \) as follows [15, pp. 22, sec. 2.2]:

\[
\hat{\mathbf{u}}_j(k+1) = \hat{\mathbf{u}}_j(k) + \alpha_k \mathbf{p}(k)
\]

where \( \alpha_k \) is a step length factor, which can be chosen as described in [15, pp. 36, ch. 3], and \( \mathbf{p}(k) = -\nabla F(\mathbf{u}_j(k)) \) is the search direction of the algorithm. In particular, when the optimization problem is linear, some expressions exist to compute the optimal step length in order to improve the convergence speed of the algorithm. On the other hand, when the optimization problem is non-linear, as considered in this contribution, a fixed and small step value is in general preferred in order to reduce the oscillatory effect when the algorithm approaches a solution. In such a case, we have \( \alpha_k = 0.5\mu \) [12], where \( \mu \) is the learning speed.

2) Enhanced Steepest Descent (ESD): The SD method provides, in general, a good accuracy in estimating the final solution. However, it often requires a large number of iterations, which may result in a too slow convergence speed for mobile ad-hoc wireless sensor networks. The proposed ESD algorithm aims at improving the convergence speed of the SD algorithm, while trying to maintain its good accuracy for position estimation. The basic idea behind the ESD algorithm is to adjust the step length value \( \alpha_k \) as a function of the current and previous search directions \( \mathbf{p}(k) \) and \( \mathbf{p}(k-1) \), respectively. In particular, \( \alpha_k \) is adjusted as follows:

\[
\begin{align*}
\alpha_k &= \alpha_{k-1} + \gamma & \text{if } & \theta_k < \theta_{\min} \\
\alpha_k &= \frac{\alpha_{k-1}}{\delta} & \text{if } & \theta_k > \theta_{\max} \\
\alpha_k &= \alpha_{k-1} & \text{otherwise}
\end{align*}
\]

where \( \theta_k = \angle(\mathbf{p}(k), \mathbf{p}(k-1)) \), \( \gamma < 1 \) is a linear increment factor, \( \delta > 1 \) is a multiplicative decrement factor, and \( \theta_{\min} \) and \( \theta_{\max} \) are two threshold values which control the step length update.

By using the four degrees of freedom \( \gamma, \delta, \theta_{\min} \) and \( \theta_{\max} \), we can simultaneously control the convergence rate of the algorithm, and the oscillatory phenomenon when approaching the final solution in a simple way, and without appreciably increasing the complexity of the algorithm when compared to the SD method. Basically, the main advantage of the ESD algorithm is the adaptive optimization of the step length factor \( \alpha_k \) at run time, which allows to dynamically either accelerate or decelerate the convergence speed of the algorithm as a function of the actual value of the function to be optimized.

In Section IV, we will show the improvement introduced by this algorithm via numerical simulation.

IV. NUMERICAL RESULTS

In this section, we show some MATLAB simulation results in order to i) substantiate the improvement provided by the proposed ESD algorithm and ii) perform a statistical analysis of its error distribution. For space limitations, we report here just a subset of the obtained numerical results (see also [19]).

A. Simulation Scenario

We consider a simulation scenario, shown in Fig. 1, which is used to have a common reference environment to analyze the improvement provided by the proposed ESD algorithm, and to analyze statistically the positioning error distribution.

In Fig. 1, we have four anchor nodes \( A_1, A_2, A_3, A_4 \) and an unknown node \( U_1 \), which may be located in one of the positions \( T_h \) with \( h = 1,2,\ldots,9 \). In order to analyze the impact of the network geometry on the performance of the optimization algorithms, we have introduced a parameter similar to the so-called geometric dilution of precision factor [20]. In particular, in every \( T_h \) position the unknown node sees the reference nodes with an increasing angle when moving from \( T_1 \) to \( T_9 \); this corresponds to moving from a scenario \( (T_1) \) with a bad geometry where ambiguities may arise during position estimation, towards a scenario \( (T_9) \) where the unknown node is surrounded by reference nodes, thus giving...

![Fig. 1. Reference scenario and network topology for algorithms’ comparison.](image-url)
an ideally optimal network topology for position estimation regardless of the specific algorithm [7].

The main parameters used in the above simulation setup are as follows: i) the reference nodes’ positions are $\mathbf{u}_1 = [0, 0, 0]^T$, $\mathbf{u}_2 = [6, 0, 0]^T$, $\mathbf{u}_3 = [3, 6, 0]^T$, and $\mathbf{u}_4 = [3, 3, 1]^T$; ii) the only unknown node in Fig. 1 may occupy 9 positions, e.g., $\mathbf{u}_1 = [40, 4, 0]^T$ m in $T_1 (9^\circ)$ and $\mathbf{u}_1 = [3, 4, 0]^T$ m in $T_9 (216^\circ)$; iii) the ranging error model is the one described in Section II-B, with $\sigma_R$ in the range [0.1, 1.6] m; iv) the position error statistics are obtained by averaging over 5000 realizations of each ranging error for every position of the unknown node; v) the unknown node determines as estimated position the one corresponding to the minimum value assumed by the error function defined in (1), when the algorithm is initialized with two randomly distributed initial starting points [1]; vi) the maximum number of iterations for the localization algorithms is 5000; vii) the initial learning speed for ESD is $\mu = 0.1$; and viii) the degrees of freedom for the ESD algorithm are: $\gamma = 0.1$, $\delta = 1.75$, $\theta_{\min} = 5^\circ$ and $\theta_{\max} = 30^\circ$.

### B. Results

In Table I, we have provided a brief summary and comparison of some optimization algorithms in terms of computational time, mean and standard deviation of the positioning error\(^1\). We observe that: i) the positioning error increases when moving the unknown node from $T_1$ to $T_9$ (see Fig. 1) due to the network topology, as expected, ii) the triangulation algorithm (INV) provides the worst performance in terms of error accuracy, iii) the ESD algorithm provides the same accuracy as the SD and NLS algorithms, but reaches the final solution faster (this is an important result for, e.g., mobile networks), iv) in terms of global computational time the ESD performs as well as Conjugate Gradient (CG) algorithms [1], [21] in most scenarios, but outperforms CG algorithms in those network topologies that are more prone to ambiguities (e.g., when the unknown node is located in $T_1$–$T_9$ positions in Fig. 1).

In Figs. 2–7, we have analyzed statistically the error distribution of the position error for the proposed ESD algorithm with respect to network topology and ranging error. In particular, by relying on the preliminary results shown in [1], [2], where we have shown that the Pearson system of distributions can fit the positioning error very well, we report here a parametric estimation of the four parameters, i.e., mean, standard deviation, skewness, and kurtosis, that characterize this family of distributions [22]. The results shown in the above figures assume that the reference nodes are “startup anchors”, even though the analysis can be extend to “converted anchors” as well [19]. Since the analysis is performed in spherical coordinates, in Figs. 2–5 we have provided the parameters for the absolute value of the error position, i.e., the distance from the true position, and, as as example, in Figs. 6, 7, we have reported the first Pearson parameter for the azimuthal and elevation phase angles, respectively. In particular, Figs. 2, 3 well show that both the mean value and the standard deviation of the positioning error get smaller when either the distance between sensor nodes is estimated better, i.e., the standard deviation of the ranging error is smaller, or the angle under which the unknown node sees the reference nodes increases. So, as expected, the performance of the algorithm gets better when ranging distances are better estimated and when the network topology is less prone to ambiguities.

### V. Conclusions

In this paper, we have proposed a statistical model for estimating the positioning error distribution of a new atomic localization algorithm called Enhanced Steepest Descent. The statistical analysis relies on modeling the error distribution via the Pearson system of distributions, which is shown to fit very well simulated data. The paper proposes a comprehensive parametric analysis of the parameters of the Pearson fitting distribution with respect to main design issues for Wireless Sensor Networks: i) ranging error, and ii) network topology. The proposed results may be used for the analysis and design of the upper layers of the protocol stack of Wireless Sensor Networks, such as resource management, routing, and application related issues, as well as for defining error control mechanisms, which may improve the accuracy and robustness of the network localization process on its own.

### REFERENCES

Fig. 2. First parameter (i.e., mean error) of the Pearson fitting distribution of the absolute positioning error versus the ranging error and the angle under which the unknown node sees the references nodes (see Fig. 1).

Fig. 3. Second parameter (i.e., error standard deviation) of the Pearson fitting distribution of the absolute positioning error versus the ranging error and the angle under which the unknown node sees the references nodes (see Fig. 1).


Fig. 4. Third parameter (i.e., error skewness) of the Pearson fitting distribution of the absolute positioning error versus the ranging error and the angle under which the unknown node sees the references nodes (see Fig. 1).

Fig. 5. Fourth parameter (i.e., error kurtosis) of the Pearson fitting distribution of the absolute positioning error versus the ranging error and the angle under which the unknown node sees the references nodes (see Fig. 1).


Fig. 6. First parameter (i.e., mean error) of the Pearson fitting distribution of the azimuthal phase angle positioning error versus the ranging error and the angle under which the unknown node sees the references nodes (see Fig. 1).

Fig. 7. First parameter (i.e., mean error) of the Pearson fitting distribution of the elevation phase angle positioning error versus the ranging error and the angle under which the unknown node sees the references nodes (see Fig. 1).


