RSS-Based Sensor Network Localization in Contaminated Gaussian Measurement Noise

Feng Yin and Ang Li and Abdelhak M. Zoubir
Signal Processing Group
Technische Universität Darmstadt, Germany
Email: {fyin, zoubir}@spg.tu-darmstadt.de

Carsten Fritsche
IFEN GmbH
Poing, Germany
Email: carsten@isy.liu.se

Fredrik Gustafsson
Division of Automatic Control
Linköping University, Sweden
Email: fredrik@isy.liu.se

Abstract—We study received signal strength-based cooperative localization in wireless sensor networks. We assume that the measurement noise fits a contaminated Gaussian model so as to take into account some outlier conditions. In addition, some environment-dependent parameters are assumed to be unknown. We propose an expectation-maximization based algorithm for robust centralized network localization without offline training. As benchmark for comparison, we express the best achievable localization accuracy in terms of the Cramér-Rao bound. Experimental results demonstrate the advantages of the proposed algorithm as compared to some representative algorithms.

Keywords—Cooperative localization, Cramér-Rao bound (CRB), expectation-maximization (EM), non-Gaussian noise, received signal strength (RSS).

I. INTRODUCTION

Location information is crucial to numerous wireless sensor network applications. Cooperative localization enables robust, accurate and concurrent inference of a large number of agents (wireless sensors with unknown positions), given a few true anchor positions and a set of noisy measurements. Among different types of measurements, received signal strength (RSS) is easier to measure and the resulting localization system requires less cost.

Representative non-Bayesian RSS-based network localization algorithms include the maximum-likelihood estimation (MLE) algorithm [1], the multi-dimensional scaling (MDS) algorithm [2], the MDS-MLE algorithm [3], and the semidefinite programming (SDP) based algorithm [4], to mention a few. To the best of our knowledge, the performance of the existing algorithms highly relies on the Gaussian measurement noise assumption and/or the accurate knowledge about the environment-dependent model parameters (e.g., path loss exponent, shadowing noise variance, etc.), calibrated a priori in an offline training phase.

It has been validated in various measurement campaigns that noisy measurements often include outliers that are far from the bulk of the data and consequently violate the Gaussian model assumption. Besides, the environment-dependent model parameters may vary with time. These motivate us to design a new algorithm that is able to achieve more robustness against outliers and meanwhile get rid of the expensive offline training phase. Towards that end, we adopt a contaminated Gaussian (CG) model to represent the measurement noise and propose to jointly estimate the positions and auxiliary environment-dependent model parameters via the expectation-maximization (EM) criterion [5]. It is noteworthy that outlier compensation using the EM algorithm has been tested in [6] for round trip time-of-arrival (RTOA)-based cooperative localization.

Our contributions of this paper are in order. First, we propose a centralized EM algorithm using RSS measurements. Secondly, we derive the Cramér-Rao bound (CRB) for position estimation in non-Gaussian noise (with possibly any distribution), which generalizes the results (valid merely for the Gaussian model) in [1]. Thirdly, we test our algorithm under real experimental settings.

The remainder of this paper is organized as follows. Section II presents the signal model. Section III describes our proposed EM algorithm, followed by the CRB computation in Section IV. Simulation results and algorithm comparisons are shown in Section V. Finally, conclusions are drawn in Section VI.

II. SIGNAL MODEL

We consider a connected network of \textit{N} sensors in a two-dimensional (2-D) space. We assume, without loss of generality, that the first \textit{N}_u agents are agents with unknown positions \( \mathbf{x}_i = [x_i, y_i]^T \), \( i = 1, 2, ..., N_u \) and the remaining sensors are anchors with known positions. For many reasons, for instance, limited link capacity or short communication range, full communication among sensors is impossible. Hence, we define \( \mathcal{H}(i) \) to indicate the set of neighboring sensors with whom the \textit{i}th sensor communicates. Moreover, we assume: (1) \( i \notin \mathcal{H}(i) \); (2) \( j \in \mathcal{H}(i) \) if and only if \( i \in \mathcal{H}(j) \).

A time-averaged RSS measurement \( r_{i,j} \) is stored on the \textit{i}th agent if and only if the signal is broadcast by the \textit{j}th sensor with index \( j > i \) and \( r_{i,j} \geq P_{thr} \) (if this holds, sensor \( i \) will simultaneously include sensor \( j \) in its neighbor list). We assume reciprocal communication channel for any neighboring sensor pair \((i,j)\), implying that \( r_{i,j} = r_{j,i} \). In this work, position inference will be carried out based on the statistical RSS measurement model:

\[
\begin{align*}
   r_{i,j} = P_T - A - 10B \log_{10} \left( \frac{d_{i,j}}{d_0} \right) + \nu_{i,j}
\end{align*}
\]

where \( P_T \) (dBm) is a known transmit power and identical for all sensors (extension to unknown and unequal transmit power over sensors, like in [4], might be possible); \( A \) (dB) is a reference path loss value at \( d_0 = 1 \) meter (m); \( B \) (dB) is a path loss exponent; \( d_{i,j} = ||x_i - x_j|| \) is the true Euclidean distance (m) between sensor \( i \) and sensor \( j \); and \( \nu_{i,j} \) (dB) is measurement noise.
Typically, $r_{i,j}$ only accounts for the large-scale shadowing effect as the small-scale multipath fading is expected to be eliminated by the time averaging and the sensor noise is negligible. In this case, a single zero-mean Gaussian distribution suffices the modeling. However, under atypical condition that the small-scale multipath fading is hard to remove [3] and/or a man made interference has been triggered near the sensor and/or a non-line-of-sight (NLOS) side-path dominates the largely attenuated line-of-sight (LOS) path, uncertainty about the actual position will be increased. For simplicity we assume another zero mean Gaussian distribution but with higher variance for the modeling. Without any prior knowledge about the occurrence of these two conditions, we assume that the noise terms observed for different sensor pairs are independent and identically distributed (i.i.d.) according to a contaminated Gaussian distribution:

$$p_Y(v) = \alpha_1 N(v; 0, \sigma_1^2) + \alpha_2 N(v; 0, \sigma_2^2)$$  \hspace{1cm} (2)

where $\alpha_1$ is the prior probability for typical (non-outlier) condition and $\alpha_2$ for atypical (outlier) condition.

Rather than assuming the auxiliary parameters $A$, $B$, $\alpha_1$, $\alpha_2$, $\sigma_1^2$, $\sigma_2^2$, $l = 1, 2$ to be known from foregoing offline training, we estimate them jointly with the unknown positions. For ease of interpretation in the subsequent sections, we let $\theta = [\theta_1^T, \theta_2^T]^T$ with $\theta_a = [A, B, \alpha_1, \alpha_2, \sigma_1^2, \sigma_2^2]^T$ and $\theta_p = [x_1, x_2, \ldots, x_N, y_1, y_2, \ldots, y_N]^T$. We further let $S$ be the set of all feasible sensor pairs $(i, j)$ for which RSS measurements $r_{i,j}$ are obtained and stacked in a vector $r$.

III. CENTRALIZED EM ALGORITHM

In this paper, we adopt the EM criterion to find the ML estimator of $\theta$. As in [6], we introduce a complete data set $\{z, r\}$ where the column vector $z$ encompasses $|S|$ latent variables $z_{i,j}$ whose value indicates under which condition (typical or atypical) the corresponding RSS measurement $r_{i,j}$ has been generated. The proposed EM algorithm works with the complete data log-likelihood function:

$$L_C(\theta; z, r) = \sum_{(i,j) \in S} \ln \left( \alpha_{z_{i,j}} N(r_{i,j}; P_T + PL_{i,j}, \sigma_{z_{i,j}}^2) \right)$$

A. The Algorithm

**Expectation Step:** We perform expectation of the complete data log-likelihood in terms of $z$, namely,

$$Q(\theta; \theta^{(n)}) = \sum_{z \in Y} L_C(\theta; z, r) \Pr(z \mid r; \theta^{(n)})$$

where $Y$ is the parameter space of $z$. After some manipulations, we obtain

$$Q(\theta; \theta^{(n)}) = \sum_{z_{i,j} \in S} \sum_{l=1}^2 \ln \left( \alpha_{z_{i,j}} N(r_{i,j}; P_T + PL_{i,j}, \sigma_l^2) \right) \hat{P}_{l,i,j}^{(n)}$$

where $\hat{P}_{l,i,j}^{(n)}$ is a short-hand notation of the conditional probability $\Pr(z_{i,j} = l | r_{i,j}; \theta^{(n)})$ and is calculated in light of Bayes’ rule as

$$\hat{P}_{l,i,j}^{(n)} = \frac{\alpha_{z_{i,j}}^{(n)} N(r_{i,j}; P_T + PL_{l,i,j}^{(n)}, \sigma_l^2^{(n)})}{\sum_{l'=1}^2 \alpha_{z_{i,j}}^{(n)} N(r_{i,j}; P_T + PL_{l',i,j}^{(n)}, \sigma_{l'}^2^{(n)})} \hspace{1cm} (4)$$

Note that $PL_{l,i,j}^{(n)}$ defined in (4) denotes the evaluation of $PL_{i,j}$ at $[A^{(n)}, B^{(n)}, x_{i,j}^{(n)}, y_{i,j}^{(n)}]$.

**Maximization Step:** We search for an updated parameter estimate $\theta^{(n+1)}$ such that

$$Q(\theta^{(n+1)}; \theta^{(n)}) \geq Q(\theta^{(n)}; \theta^{(n)}) \hspace{1cm} (5)$$

We follow a similar methodology as introduced in [6] for this purpose, which ultimately leads to our proposed EM algorithm shown in Algorithm 1.

**Algorithm 1 The Proposed Centralized EM Algorithm**

**Initialization:** Choose a convergence tolerance $\Delta$, an initial guess $\theta^{(0)}$, and a maximum number of iterations $N_{itr}$.

**Expectation and Maximization (EM stage):** In the $(n + 1)$th iteration $(n \in \mathbb{Z}, n \geq 0)$, do:

1) Compute $\hat{P}_{l,i,j}^{(n)}$ according to (4).
2) Update $\theta^{(n+1)}$ analytically according to

$$\alpha_{l}^{(n+1)} = \frac{1}{|S|} \sum_{(i,j) \in S} \hat{P}_{l,i,j}^{(n)}$$

$$\left[A^{(n+1)}, B^{(n+1)}\right]^T = \left(\Sigma^T W \Sigma\right)^{-1} \Sigma^T W (r - P_T), \hspace{1cm} (7)$$

$$\sigma_l^{2(n+1)} = \frac{\sum_{(i,j) \in S} \hat{P}_{l,i,j}^{(n)} (r_{i,j} - P_T - PL_{l,i,j}^{(n+1)})^2}{|S| \alpha_l^{(n+1)}}$$

with

$$PL_{l,i,j}^{(n+1)} = -A^{(n+1)} - 10B^{(n+1)} \log_{10} \left( \frac{\|x_{i,j}^{(n)} - x_{j}^{(n)}\|}{d_0} \right)$$

$$\Sigma = \begin{bmatrix} -1 & 0 \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & -10 \log_{10} (\|x_{i,j}^{(n)} - x_{j}^{(n)}\|/d_0) \end{bmatrix}_{|S| \times 2}$$

$$W = \text{diag} \left( \frac{\hat{P}_{l,i,j}^{(n)}}{\sigma_1^{2(n)}}, \frac{\hat{P}_{l,i,j}^{(n)}}{\sigma_2^{2(n)}}, \ldots \right)_{|S| \times |S|}$$

Note that the data structures of $r$, $\Sigma$, and $W$ follow that of $S$ and $x_{i,j}^{(n)} = x_j$ in (9) if $j \in \{N_u + 1, \ldots, N\}$.

3) Update $\theta^{(n+1)}$ through numerically minimizing

$$f(\theta_p) = \sum_{(i,j) \in S} \sum_{l=1}^2 \frac{\hat{P}_{l,i,j}^{(n)} (r_{i,j} - P_T - PL_{l,i,j})^2}{\sigma_l^2} \hspace{1cm} (10)$$

via the BFGS quasi-Newton method with initial guess $\theta_p^{(n)}$.

Note that $[A, B, \sigma_1, \sigma_2]$ have to be replaced with their latest updates in (10) prior to the minimization.

**Convergence Check:**

If $|\ln p(r; \theta^{(n+1)}) - \ln p(r; \theta^{(n)})| \leq \Delta$ or $N_{itr}$ has been reached, then stop; otherwise reset $n \leftarrow n + 1$ and return to the EM stage.
B. Selection of Initial Guess

It is well known that the performance of the EM-type algorithms depends on the initial guess. In order to give a good starting point with low computational complexity, we propose one strategy in Algorithm 2.

Algorithm 2 Initialization Strategy

1) Choose \( A(0) \) and \( B(0) \) empirically, for instance, \( A(0) \) from the device specifications and \( B(0) \) from the classical value (2 to 5), given a specific localization environment.
2) Compute a set of distance estimates, for instance, the ML estimates under the single Gaussian noise assumption [1]:

\[
\hat{d}_{i,j} = \frac{d_0}{C} 10^{\frac{p_{r_i}-A(0)-r_{i,j}}{10n(0)}}, (C \approx 1.2), \quad \forall(i,j) \in \mathcal{S}.
\]

3) Initialize the agents’ positions \( x_i(0), i = 1, 2, \ldots, N_a \) in the classical MDS algorithm using the obtained \( \hat{d}_{i,j} \)’s. Advanced algorithms, like [3], can also be used.
4) Extract residuals, according to

\[
\hat{v}_{i,j} = r_{i,j} - P_T - P L_{i,j}(0), \quad \forall(i,j) \in \mathcal{S}.
\]

Compute the variance of the residuals and denote it by \( \sigma_r^2 \). From [7] we know that \( \sigma_r^2 \approx \alpha_1 \sigma_x^2 + \alpha_2 \sigma_y^2 \).
5) Set a coarse estimate of \( \alpha_1 \) and \( \alpha_2 \), namely \( \alpha_1(0) \) and \( \alpha_2(0) \). Choose a positive number \( K \) such that \( \sigma_2^2 = K \sigma_r^2 \). As a consequence, we have

\[
\sigma_1^2(0) = \frac{\sigma_r^2}{\alpha_1(0) + K \alpha_2(0)}, \quad \sigma_2^2(0) = K \sigma_1^2(0).
\]

In order to give a better starting point, we can try different combinations of \( \alpha_1(0), \alpha_2(0), \) and \( K \).

IV. CRAMÉR-RAO BOUND FORMATION

The Fisher information matrix (FIM) for our joint estimation problem is difficult to evaluate in closed form. Nevertheless, a numerical approximation can be obtained by following the same idea as described in [6], namely,

\[
\mathcal{F} (\theta) \approx \frac{1}{N_a} \sum_{n=1}^{N_M} \left\{ \nabla_\theta \ln p(r^{(n)}; \theta) \cdot \nabla_\theta^T \ln p(r^{(n)}; \theta) \right\},
\]

where \( r^{(n)} \), \( n = 1, 2, \ldots, N_M \) are mutually independent realizations of \( r \). Given a sufficiently large \( N_M \), the approximated FIM can be very close to the true one. However, this method is computationally expensive. As we are considering a localization problem, it is of more interest to spotlight the position estimation. Hence, we eliminate the auxiliary parameters \( \theta_a \) from the list of unknown parameters and derive a loose FIM but with a more compact form and reduced computational complexity as trade-off as follows:

\[
\mathcal{F} (\theta_p) = \mathbb{E}_{p(r; \theta_p)} \left\{ \nabla_{\theta_p} \ln p(r; \theta_p) \cdot \nabla_{\theta_p}^T \ln p(r; \theta_p) \right\}.
\]

Adopting the results given in [7] and further doing some manipulations, we obtain

\[
\mathcal{F} (\theta_p) \triangleq \begin{bmatrix} \mathcal{F}_{xx} & \mathcal{F}_{xy} & \mathcal{F}_{xz} \\ \mathcal{F}_{yx} & \mathcal{F}_{yy} & \mathcal{F}_{yz} \\ \mathcal{F}_{zx} & \mathcal{F}_{zy} & \mathcal{F}_{zz} \end{bmatrix}
\]

where \( \mathcal{F}_{mn}, m, n \in \{x, y\} \) are all \( N_u \times N_u \) square matrices whose entries are given respectively by

\[
[\mathcal{F}_{mn}]_{i,i'} = \begin{cases} \frac{b \cdot \mathcal{I}_o \cdot \sum_{j \in \mathcal{H}(i)} (m_j - m_i)(n_j - n_i)}{d_{i,j}^4}, & i = i' \\ \frac{b \cdot \mathcal{I}_o \cdot \delta_{i,i'} \cdot (m_i - m_j)(n_i - n_j)}{d_{i,j}^4}, & i \neq i' \end{cases}
\]

with \( b = (10B/\ln(10))^2 \), \( \delta_{i,i'} \) being an indicator whose value is one if \( i' \in \mathcal{H}(i) \) or zero otherwise, and

\[
\mathcal{I}_o = \mathbb{E}_{p(v)} \left\{ \nabla_{\mathbb{E}} p(v) \nabla_{\mathbb{E}}^T p(v) \right\}.
\]

1) For most of the noise distributions, \( \mathcal{I}_o \) has to be approximated using Monte Carlo integration [8], i.e.,

\[
\mathcal{I}_o \approx \frac{1}{N_M} \sum_{n=1}^{N_M} \left\{ \nabla_{\mathbb{E}} p(v^{(n)}) \nabla_{\mathbb{E}}^T p(v^{(n)}) \right\}.
\]

V. EXPERIMENTAL RESULTS

In the sequel, we use the real sensor network and RSS measurements mentioned in [1] (also published online as mat files) to test our proposed EM algorithm. Details about the measurement campaigns can be found easily in the original paper and thus is ignored here due to space limitations. We assume \( P_{th,r} = -90 \) dBm, below which data packets cannot be demodulated. The transmit power is set to \( P_T = 0 \) dBm so as to fit our signal model (cf. (1)) with the real data. We choose an initial guess for the EM algorithm according to Algorithm 2, more precisely, \( \theta_0 = [A(0) = 37.5, B(0) = 2.30, \alpha_1(0) = 0.90, \alpha_2(0) = 0.10, \sigma_1^2(0) = 11.84, \sigma_2^2(0) = 59.20]^T \).

We compare the scene localization RMSE of the EM algorithm with that of the centralized MLE [1], the classical MDS algorithm [2], the dwMDS algorithm [2], and the MDS-ML algorithm [3]. These competitors need rather accurate knowledge of the received power at \( d_0 \) and the path loss exponent \( B \), which are usually measured in an offline calibration. In contrast, the EM algorithm avoids the offline calibration. This readily shows the advantage of our algorithm. Due to space limitations, we merely show the EM position estimate versus the true sensor positions in Fig. 1. The RMSE values are given in Table I. Some remarks are in order. First, we use the ultimate EM estimate of \( \theta_p, \theta_p^{EM} = [A = 39.54, B = 2.10, \alpha_1 = 0.84, \alpha_2 = 0.16, \sigma_1^2 = 3.60, \sigma_2^2 = 33.10] \), as the true value (for lack of ground truth) to compute \( CRB_{pos} \) for the contaminated Gaussian model, cf. Section IV. Secondly, we only observe slight improvement. On the one hand, the algorithms are tested with only one realization of the RSS data rather than in a large scale Monte Carlo experiment. On the other hand, the proposed algorithm has more parameters...
to estimate. Apart from these, the CG model might be still insufficient to represent the underlying model. Thirdly, the fundamental limits of the RMSE discloses the sub-optimality of using the single Gaussian distribution (with \( \sigma^2 = 3.92 \)) to model the noise terms if they were in fact generated from the CG distribution, cf. (2).

One question naturally arises, i.e., when do we benefit from a CG distribution in the modeling (assuming no model mismatch problem)? To give the answer, we assume

\[
p_V(v) = (1 - \epsilon)N(v; 0, \sigma^2) + \epsilon N(v; 0, k\sigma^2)
\]

where \( \sigma^2 = 1/(1 + (k - 1)\epsilon), \ 0 \leq \epsilon \leq 1, \) and \( k \geq 1. \)

This setup allows a straightforward comparison with a single Gaussian distribution \( N(v; 0, 1) \) (an approximation of \( p_V(v) \)). We evaluate the ratio \( \Gamma = \text{CRB}^{\text{CG}}_{\text{pos}} / \text{CRB}^{\text{G}}_{\text{pos}} = 1/\mathcal{I}_v \) versus \( \epsilon \) and \( k \). Here, the same \( A \) and \( B \) are used for computing \( \text{CRB}^{\text{CG}}_{\text{pos}} \) and \( \text{CRB}^{\text{G}}_{\text{pos}} \). A contour plot of these results is shown in Fig. 2, which is similar to [9, Fig. 2.3a], plotted for a different purpose. The main messages conveyed by Fig. 2 are twofold. First, we can use a single Gaussian distribution to approximate a CG distribution with \( \Gamma \approx 1 \) when (1) \( \epsilon \) is close to zero or one but \( k \) can be large; (2) \( k \) is close to one but \( \epsilon \) can be any value between zero and one. Secondly, a single Gaussian approximation is most unfavorable with \( \Gamma \ll 1 \) when \( \epsilon \approx 0.35 \) and \( k \gg 1 \).

VI. Conclusion

We have proposed an EM based algorithm for sensor network localization using non-Gaussian noise contaminated RSS measurements. Simulation results based on real data show that our EM algorithm outperforms several salient competitors. Although not shown in the paper, the localization RMSE of the EM estimator tends to achieve the fundamental limits for large data records, given good starting point. Despite the scalability problem, its performance serves as a benchmark for evaluating more attractive distributed EM algorithms in our future work.

VII. Acknowledgement

Author Feng Yin would like to thank Prof. N. Patwari for his explanations on the heavy-tailed behaviour of the real data.

References


