

# RECEIVED SIGNAL STRENGTH-BASED JOINT PARAMETER ESTIMATION ALGORITHM FOR ROBUST GEOLOCATION IN LOS/NLOS ENVIRONMENTS

Feng Yin<sup>\*</sup> Carsten Fritsche<sup>\*</sup> Fredrik Gustafsson<sup>†</sup> Abdelhak M. Zoubir<sup>\*</sup>

<sup>\*</sup> Signal Processing Group, Technische Universität Darmstadt, Germany.

Email: {fyin, zoubir}@spg.tu-darmstadt.de

<sup>\*</sup> IFEN GmbH, Germany. Email: carsten@isy.liu.se

<sup>†</sup> Division of Automatic Control, Linköping University, Sweden. Email: fredrik@isy.liu.se

## ABSTRACT

We consider received-signal-strength-based robust geolocation in mixed line-of-sight/non-line-of-sight propagation environments. Herein, we assume a mode-dependent propagation model with unknown parameters. We propose to jointly estimate the geographical coordinates and propagation model parameters. In order to approximate the maximum-likelihood estimator (MLE), we develop an iterative algorithm based on the well-known expectation and maximization criterion. As compared to the standard ML implementation, the proposed algorithm is simpler to implement and capable of reproducing the MLE. Simulation results show that the proposed algorithm attains the best geolocation accuracy as the number of measurements increases.

**Index Terms**— Expectation-maximization (EM) criterion, geolocation, line-of-sight (LOS)/non-line-of-sight (NLOS), received-signal-strength (RSS).

## 1. INTRODUCTION

Wireless geolocation has been of significant interest during the past two decades due to the expanding location-aware services [1]. In order to give a position estimate, we prefer to use a two-step geolocation technique which comprises two phases: measurements collection phase and position estimation phase. Based on different types of measurements, such as time-of-arrival (TOA), received-signal-strength (RSS), and angle of arrival (AOA), a plethora of algorithms have been proposed, see for instance [2] and references therein.

We focus on the position estimation using RSS measurements. RSS-based algorithms are attractive for wireless geolocation due to two main reasons. First, the RSS measurements are readily available in almost any wireless infrastructure, and therefore expensive system update can be avoided. Second, RSS-based algorithms are in general of low-complexity, and thus can be easily integrated into other geolocation systems [2, Chapter 11].

We consider a specific geolocation problem where the propagation condition may switch between line-of-sight (LOS) and non-LOS (NLOS) in wireless networks. In contrast to the vast majority of papers in the literature, we assume a mode-dependent propagation model in order to well capture the switching between LOS and NLOS. Rather than assuming that the propagation model parameters are known, we estimate them jointly with the unknown geographical coordinates. Herein, we pursue the maximum-likelihood estimator (MLE) due to its notable asymptotic performance, known from classical estimation theory [3]. Unfortunately, the standard ML implementation is too cumbersome. As an alternative, we propose

to design an iterative algorithm based on the expectation and maximization (EM) criterion [4], which is famous for its simple implementation and ability to reproduce the MLE.

The remainder of this paper is organized as follows. In Section 2, we introduce the signal model and state the problem at hand. In Section 3, we briefly introduce the standard way of solving our estimation problem. Section 4 describes our newly proposed algorithm, which is developed in light of the EM criterion, followed by simulation results in Section 5. Finally, we relate our original contributions to prior work in Section 6.

## 2. SIGNAL MODEL

It is assumed that  $M$  anchors surround a stationary agent to be located in a wireless network. Let  $\mathbf{x}_m = [x_m, y_m]^T$  be the known geographical coordinates of the  $m$ th anchor and let  $\mathbf{x} = [x, y]^T$  be the unknown geographical coordinates of the agent. Here, we consider a centralized geolocation process, in which we collect  $K$  averaged (over a time interval  $T$ ) received-signal-strength (RSS) measurements at each anchor. The  $k$ th averaged RSS measurement collected at the  $m$ th anchor,  $z_k^{(m)}$  (dBm), is modeled by

$$z_k^{(m)} = \underbrace{P_T - L_k^{(m)}(r_k^{(m)})}_{h_k^{(m)}(r_k^{(m)})} + v_k^{(m)}(r_k^{(m)}), \quad (1)$$

for  $k = 1, 2, \dots, K$  and  $m = 1, 2, \dots, M$ , where  $P_T$  (dBm) is a known transmit power of the agent;  $r_k^{(m)}$  is a mode variable, indicating the current propagation condition;  $L_k^{(m)}(r_k^{(m)})$  denotes the path loss and  $v_k^{(m)}(r_k^{(m)})$  is an error term accounting for the shadowing effect (both in dB scale). Here, we set  $r_k^{(m)} = 1$  for LOS condition or  $r_k^{(m)} = 2$  for NLOS condition and assume both the path loss and the error term are dependent on the current mode  $r_k^{(m)}$ . We stress, however, that  $r_k^{(m)}$  is unknown. Besides, we assume multipath fading effect in  $z_k^{(m)}$  has been effectively eliminated by an *a priori* time-averaging [5, Chapter 2].

As the propagation condition may switch between LOS and NLOS, we follow the mode-dependent log-distance path loss model (in dB scale) as defined in [6, 7, 8]. That is,

$$L_k^{(m)}(r_k^{(m)}) = \begin{cases} A_{\text{LOS}} + 10B_{\text{LOS}} \log_{10} \left( \frac{d_m(\mathbf{x})}{d_0} \right), & r_k^{(m)} = 1 \\ A_{\text{NLOS}} + 10B_{\text{NLOS}} \log_{10} \left( \frac{d_m(\mathbf{x})}{d_0} \right), & r_k^{(m)} = 2 \end{cases} \quad (2)$$

where the column vector  $\beta_1$  includes parameters  $A_{\text{LOS}}$  and  $B_{\text{LOS}}$  denoting a reference path loss value at  $d_0 = 1$  kilometer (km) and a path loss exponent value, respectively; analogously, column vector  $\beta_2$  includes parameters  $A_{\text{NLOS}}$  and  $B_{\text{NLOS}}$  representing the counterparts under the NLOS condition; and  $d_m(\mathbf{x})$  denotes the distance between the agent and the  $m$ th anchor, i.e.,

$$d_m(\mathbf{x}) = \sqrt{(x - x_m)^2 + (y - y_m)^2}.$$

The shadowing effect is also mode-dependent. Hence, we assume  $v_k^{(m)}(r_k^{(m)} = 1) = v_{\text{LOS},k}^{(m)} \sim \mathcal{N}(v; 0, \sigma_1^2)$  under the LOS condition, whereas  $v_k^{(m)}(r_k^{(m)} = 2) = v_{\text{NLOS},k}^{(m)} \sim \mathcal{N}(v; 0, \sigma_2^2)$  under the NLOS condition. The standard deviation  $\sigma_2$  (dB) is usually much larger than  $\sigma_1$  (dB) reflecting the fact that the shadowing effect is more prominent under the NLOS condition.

In general, the switching between LOS and NLOS is time-dependent and favorable to be modeled by a 2-state Markov chain [9] with mode variable  $r_k^{(m)}$ . Herein, we assume the Markov chain is time-homogeneous and regular, from which it follows that the Markov chain will converge to a unique stationary distribution as  $k \rightarrow \infty$  [10], namely,

$$\lim_{k \rightarrow \infty} \Pr\{r_k^{(m)} = l\} = \alpha_l^{(m)}, \quad l = 1, 2,$$

where  $\alpha_l^{(m)}$  denotes the probability of the  $m$ th anchor being in the  $l$ th mode ( $l = 1$  for LOS or  $l = 2$  for NLOS) and  $\alpha_1^{(m)} + \alpha_2^{(m)} = 1$ . Since we consider here a homogeneous environment, we assume  $\alpha_l^{(1)} = \alpha_l^{(2)} = \dots = \alpha_l^{(M)} = \alpha_l$ ,  $l = 1, 2$ .

In the sequel, we focus on snapshot-based estimator design using RSS measurements  $z_k^{(m)}$ ,  $k = 1, 2, \dots, K$  and  $m = 1, 2, \dots, M$ . In order to facilitate the algorithm design as well as the measurement generation, we assume the shadowing errors  $v_k^{(m)}$ ,  $k = 1, 2, \dots, K$  and  $m = 1, 2, \dots, M$  are independent and identically distributed (i.i.d.) random variables with a two-mode Gaussian mixture distribution, to which the Markov chain will converge as  $k$  increases. Hence, we have

$$p(z_k^{(m)}) = \sum_{l=1}^2 \alpha_l \mathcal{N}(z_k^{(m)}; h_k^{(m)}(r_k^{(m)} = l), \sigma_l^2).$$

For ease of subsequent interpretation, we introduce a vectorized signal model as follows:

$$\mathbf{z} = \mathbf{h}(\mathbf{r}) + \mathbf{v}(\mathbf{r})$$

where the vector notations are of the form:

$$\begin{aligned} \mathbf{r} &= [r_1^{(1)}, \dots, r_K^{(1)}, \dots, r_1^{(m)}, \dots, r_K^{(m)}, \dots, r_1^{(M)}, \dots, r_K^{(M)}]^T \\ \mathbf{z} &= [z_1^{(1)}, \dots, z_K^{(1)}, \dots, z_1^{(m)}, \dots, z_K^{(m)}, \dots, z_1^{(M)}, \dots, z_K^{(M)}]^T \\ \mathbf{h}(\mathbf{r}) &= [h_1^{(1)}(r_1^{(1)}), \dots, h_K^{(1)}(r_K^{(1)}), \dots, h_1^{(M)}(r_1^{(M)}), \dots, h_K^{(M)}(r_K^{(M)})]^T \\ \mathbf{v}(\mathbf{r}) &= [v_1^{(1)}(r_1^{(1)}), \dots, v_K^{(1)}(r_K^{(1)}), \dots, v_1^{(M)}(r_1^{(M)}), \dots, v_K^{(M)}(r_K^{(M)})]^T. \end{aligned}$$

Column vectors  $\mathbf{r}$ ,  $\mathbf{z}$ ,  $\mathbf{h}$ , and  $\mathbf{v}$  are all of dimension  $MK \times 1$ . In the sequel, we introduce  $N = MK$  to indicate the total number of RSS measurements and use  $r^{(i)}$ ,  $z^{(i)}$ ,  $h^{(i)}(r^{(i)})$ , and  $v^{(i)}(r^{(i)})$  to represent the  $i$ th entry of the corresponding vectors defined above.

Throughout this paper we refer to  $\alpha_1$ ,  $\alpha_2$ ,  $A_{\text{LOS}}$ ,  $B_{\text{LOS}}$ ,  $A_{\text{NLOS}}$ ,  $B_{\text{NLOS}}$ ,  $\sigma_1^2$ , and  $\sigma_2^2$  as mode-dependent propagation model parameters. Rather than assuming that they are readily known from some foregoing off-line calibrations (e.g., [11]) or sophisticated estimation procedures with auxiliary information (e.g., [12]), we propose to estimate them jointly with the unknown geographical coordinates.

### 3. STANDARD MAXIMUM-LIKELIHOOD IMPLEMENTATION

Defining a vector parameter  $\boldsymbol{\theta} = [\alpha_1, \alpha_2, \beta_1^T, \beta_2^T, \sigma_1^2, \sigma_2^2, \mathbf{x}]^T$ , we can easily express the log-likelihood function of a realization of  $\mathbf{z}$  conditioned on  $\boldsymbol{\theta}$ , namely,

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{z}) = \ln \left( \prod_{i=1}^N p(z^{(i)}; \boldsymbol{\theta}) \right). \quad (3)$$

By introducing  $\tilde{z}_l^{(i)}(\mathbf{x}, \beta_l) = z^{(i)} - h^{(i)}(r^{(i)} = l)$  for  $l = 1, 2$ , we then re-write (3) as

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{z}) = \sum_{i=1}^N \ln \left( \sum_{l=1}^2 \alpha_l \mathcal{N}(\tilde{z}_l^{(i)}(\mathbf{x}, \beta_l); 0, \sigma_l^2) \right). \quad (4)$$

The maximum-likelihood estimate  $\hat{\boldsymbol{\theta}}_{\text{ML}}$  is the solution of the following optimization problem:

$$\begin{aligned} & \underset{\boldsymbol{\theta}}{\text{maximize}} \quad \mathcal{L}(\boldsymbol{\theta}; \mathbf{z}) \\ & \text{subject to} \quad 0 < \alpha_1, \alpha_2 < 1, \alpha_1 + \alpha_2 = 1, \\ & \quad \sigma_1^2 > 0, \sigma_2^2 > 0. \end{aligned} \quad (5)$$

The cost function, cf.(4), contains “the logarithm of the sum”, implying difficulties in obtaining the MLE. Instead of solving (5) with tremendous efforts, we take a detour to approximate the MLE by using the EM criterion.

### 4. EXPECTATION-MAXIMIZATION ALGORITHM

Before we apply the EM criterion to our estimation problem, let us first introduce a complete data set  $\boldsymbol{\varpi} = \{\mathbf{y}, \mathbf{z}\}$  with  $\mathbf{y} = [y^{(1)}, y^{(2)}, \dots, y^{(N)}]^T$  denoting a vector of  $N$  random variables (also called latent variables) whose values tell us which mixture component has generated (or which propagation condition has led to) the corresponding path loss and error term. Since  $z^{(i)}$ ,  $i = 1, 2, \dots, N$  are independent and  $y^{(i)}$ ,  $i = 1, 2, \dots, N$  are independent as a consequence, the complete data log-likelihood function can be easily expressed by

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{y}, \mathbf{z}) = \sum_{i=1}^N \ln \left( \alpha_{y^{(i)}} \mathcal{N}(\tilde{z}_{y^{(i)}}^{(i)}(\mathbf{x}, \beta_{y^{(i)}}); 0, \sigma_{y^{(i)}}^2) \right).$$

In the sequel, we apply the EM criterion to the complete data model to approximate the MLE. The designed iterative algorithm is called “EM algorithm” in the remaining parts of this paper. Given a realization of  $\mathbf{z}$  and the *a priori* parameter estimate  $\boldsymbol{\theta}^{(\eta)}$ , the proposed EM algorithm alternates between the following two steps on the  $(\eta + 1)$ th iteration.

**Expectation Step:** As the first step, we perform expectation of the complete data log-likelihood in terms of  $\mathbf{y}$ . Since  $\mathbf{y}$  is discrete-valued, we have

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(\eta)}) = \sum_{\mathbf{y} \in \mathfrak{Y}} \ln(p(\mathbf{y}, \mathbf{z}; \boldsymbol{\theta})) \Pr\{\mathbf{y} | \mathbf{z}; \boldsymbol{\theta}^{(\eta)}\}$$

where  $\mathfrak{Y}$  is the parameter space of  $\mathbf{y}$  and  $\Pr\{\mathbf{y} | \mathbf{z}; \boldsymbol{\theta}^{(\eta)}\}$  is a conditional probability of the latent variables  $\mathbf{y}$ . Sticking to the procedure used for simplifying  $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(\eta)})$  in [13], we finally have

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(\eta)}) = \sum_{i=1}^N \sum_{l=1}^2 \ln \left( \alpha_l \mathcal{N}(\tilde{z}_l^{(i)}(\mathbf{x}, \beta_l); 0, \sigma_l^2) \right) \tilde{P}_{l,i}^{(\eta)} \quad (6)$$

where  $\tilde{P}_{l,i}^{(\eta)}$  is a short-hand notation of the conditional probability  $\Pr\{l|z^{(i)}; \theta^{(\eta)}\}$  due to Bayes' rule:

$$\Pr\{l|z^{(i)}; \theta^{(\eta)}\} = \frac{\alpha_l^{(\eta)} \mathcal{N}(\tilde{z}_l^{(i)}(\mathbf{x}^{(\eta)}, \beta_l^{(\eta)}); 0, \sigma_l^{2,(\eta)})}{p(z^{(i)}; \theta^{(\eta)})}. \quad (7)$$

**Maximization Step:** In the second step, we update the parameter estimate  $\theta^{(\eta+1)}$  through maximizing  $Q(\theta; \theta^{(\eta)})$ , derived in (6), with respect to  $\theta$ , that is,

$$\theta^{(\eta+1)} = \arg \max_{\theta} Q(\theta; \theta^{(\eta)}).$$

For convenience, we reformulate  $Q(\theta; \theta^{(\eta)})$  as follows:

$$Q(\theta; \theta^{(\eta)}) = Q_0(\alpha_1, \alpha_2; \theta^{(\eta)}) + \sum_{l=1}^2 Q_l(\beta_l, \sigma_l^2; \mathbf{x}; \theta^{(\eta)})$$

where

$$Q_0(\alpha_1, \alpha_2; \theta^{(\eta)}) \triangleq \sum_{i=1}^N \ln(\alpha_1) \tilde{P}_{1,i}^{(\eta)} + \ln(\alpha_2) \tilde{P}_{2,i}^{(\eta)},$$

$$Q_l(\beta_l, \sigma_l^2; \mathbf{x}; \theta^{(\eta)}) \triangleq \sum_{i=1}^N \ln(\mathcal{N}(\tilde{z}_l^{(i)}(\mathbf{x}, \beta_l); 0, \sigma_l^2)) \tilde{P}_{l,i}^{(\eta)}.$$

In order to obtain  $\alpha_l^{(\eta+1)}$  that maximizes  $Q(\theta; \theta^{(\eta)})$ , we solve

$$\frac{\partial}{\partial \alpha_l} [Q_0(\alpha_1, \alpha_2; \theta^{(\eta)}) + \lambda(\alpha_1 + \alpha_2 - 1)] = 0$$

where  $\lambda$  is the Lagrange multiplier. Consequently, we get

$$\alpha_l^{(\eta+1)} = \frac{1}{N} \sum_{i=1}^N \tilde{P}_{l,i}^{(\eta)}, \quad l = 1, 2. \quad (8)$$

Next, we update the estimate of  $\beta_l$ ,  $l = 1, 2$  as follows:

$$\beta_l^{(\eta+1)} = \arg \max_{\beta_l} Q_l(\beta_l, \sigma_l^2; \mathbf{x}; \theta^{(\eta)}),$$

which is equivalent to, as per observation,

$$\beta_l^{(\eta+1)} = \arg \min_{\beta_l} \sum_{i=1}^N \left( \tilde{z}_l^{(i)}(\mathbf{x}, \beta_l) \right)^2 \tilde{P}_{l,i}^{(\eta)}. \quad (9)$$

The cost function in (9) can be written as

$$\sum_{i=1}^N \left( \tilde{z}_l^{(i)}(\mathbf{x}, \beta_l) \right)^2 \tilde{P}_{l,i}^{(\eta)} = (\boldsymbol{\xi} - \boldsymbol{\Sigma} \beta_l)^T \mathbf{W}_l (\boldsymbol{\xi} - \boldsymbol{\Sigma} \beta_l)$$

where  $\boldsymbol{\xi} = [z^{(1)} - P_T, z^{(2)} - P_T, \dots, z^{(N)} - P_T]^T$ ,

$$\boldsymbol{\Sigma} = \begin{bmatrix} -1, & -10 \log_{10}(d^{(1)}(\mathbf{x})/d_0) \\ -1, & -10 \log_{10}(d^{(2)}(\mathbf{x})/d_0) \\ \vdots & \vdots \\ -1, & -10 \log_{10}(d^{(N)}(\mathbf{x})/d_0) \end{bmatrix}, \quad (10)$$

and  $\mathbf{W}_l = \text{diag}(\tilde{P}_{l,1}^{(\eta)}, \tilde{P}_{l,2}^{(\eta)}, \dots, \tilde{P}_{l,N}^{(\eta)})$  is a diagonal matrix. In (10),  $d^{(i)}(\mathbf{x})$ ,  $i = 1, 2, \dots, N$ , is the  $i$ th entry of

$$\mathbf{d}(\mathbf{x}) = [\underbrace{d_1(\mathbf{x}), \dots, d_1(\mathbf{x})}_{K \text{ repetition}}, \dots, \underbrace{d_M(\mathbf{x}), \dots, d_M(\mathbf{x})}_{K \text{ repetition}}]^T.$$

It is clear that  $\beta_l^{(\eta+1)}$  is a weighted least squares estimate given by

$$\beta_l^{(\eta+1)} = (\boldsymbol{\Sigma}^T \mathbf{W}_l \boldsymbol{\Sigma})^{-1} \boldsymbol{\Sigma}^T \mathbf{W}_l \boldsymbol{\xi}. \quad (11)$$

Since  $\mathbf{x}$  in (10) is unknown, we replace it with  $\mathbf{x}^{(\eta)}$ .

In order to obtain  $\sigma_l^{2,(\eta+1)}$  that maximizes  $Q(\theta; \theta^{(\eta)})$ , we take the derivative of  $Q_l(\beta_l, \sigma_l^2; \mathbf{x}; \theta^{(\eta)})$  with respect to  $\sigma_l^2$  and set it equal to zero. The results are given by

$$\sigma_l^{2,(\eta+1)} = \frac{\sum_{i=1}^N \left( \tilde{z}_l^{(i)}(\mathbf{x}^{(\eta)}, \beta_l^{(\eta+1)}) \right)^2 \tilde{P}_{l,i}^{(\eta)}}{\sum_{i=1}^N \tilde{P}_{l,i}^{(\eta)}}, \quad l = 1, 2, \quad (12)$$

where  $\mathbf{x}^{(\eta)}$  and  $\beta_l^{(\eta+1)}$  replace  $\mathbf{x}$  and  $\beta_l$ , respectively.

Next, we find  $\mathbf{x}^{(\eta+1)}$  through maximizing  $Q(\theta; \theta^{(\eta)})$  in terms of  $\mathbf{x}$ , which is easily shown to be equivalent to

$$\mathbf{x}^{(\eta+1)} = \arg \min_{\mathbf{x}} \left\{ g(\mathbf{x}) \triangleq \sum_{i=1}^N \sum_{l=1}^2 \left( \tilde{z}_l^{(i)}(\mathbf{x}, \beta_l) \right)^2 \frac{\tilde{P}_{l,i}^{(\eta)}}{\sigma_l^2} \right\}. \quad (13)$$

Before performing the minimization, we replace  $\beta_1, \beta_2, \sigma_1^2$  and  $\sigma_2^2$  in (13) with their latest updates respectively. Since  $\tilde{z}_l^{(i)}(\mathbf{x}, \beta_l)$  contains non-linear terms of  $\mathbf{x}$ , we resort to numerical methods to find  $\mathbf{x}^{(\eta+1)}$ . Here, we adopt a BFGS quasi-Newton method as described in [14] because it guarantees downhill progress towards the local minimum in each Newton step, as reported in [15].

We repeat the E-step and M-step until some convergence condition is met. The EM estimate obtained at the convergence is denoted by  $\hat{\theta}$ . We summarize the new algorithm in Algorithm 1 below. Clearly, our EM algorithm is much simpler to implement as compared to directly solving the original problem in (5), because only the geographical coordinates have to be solved numerically.

---

#### Algorithm 1 EM Algorithm for Joint Parameter Estimation

---

##### Step 1—Initialization:

Define the convergence tolerance  $\Delta$ ; Set  $\eta = 0$ ; Choose an initial guess  $\theta^{(\eta=0)} = [\alpha_1^{(0)}, \alpha_2^{(0)}, \beta_1^{T,(0)}, \beta_2^{T,(0)}, \sigma_1^{2,(0)}, \sigma_2^{2,(0)}, \mathbf{x}^{T,(0)}]^T$ .

##### Step 2—Perform Expectation and Maximization:

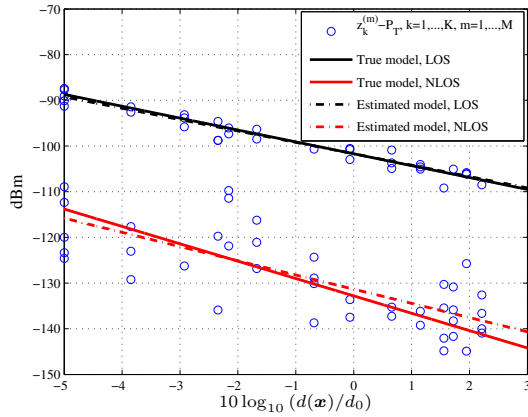
In the  $(\eta + 1)$ th iteration ( $\eta \in \mathbb{Z}, \eta \geq 0$ ), do:

- Compute  $\tilde{P}_{l,i}^{(\eta)}$  according to (7).
- Find  $\alpha_l^{(\eta+1)}, \beta_l^{(\eta+1)}, \sigma_l^{2,(\eta+1)}$ ,  $l = 1, 2$  according to (8), (11), and (12), respectively.
- Find  $\mathbf{x}^{(\eta+1)}$  through minimizing  $g(\mathbf{x})$  (cf.(13)) in terms of  $\mathbf{x}$  via the BFGS quasi-Newton method initialized with  $\mathbf{x}^{(\eta)}$ .

##### Step 3—Convergence Check:

If  $|\mathcal{L}(\theta^{(\eta+1)}; \mathbf{z}) - \mathcal{L}(\theta^{(\eta)}; \mathbf{z})| \leq \Delta$ , then stop; otherwise reset  $\eta \leftarrow \eta + 1$  and return to **Step 2**.

---



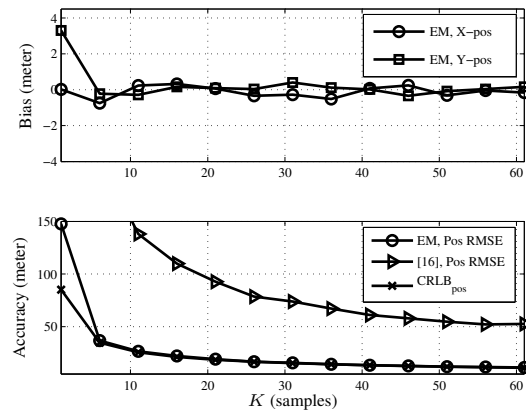
**Fig. 1.** True log-distance path loss model (in dB scale) versus its estimate in one particular Monte Carlo trial. Note that we introduced a negative sign in the original model (cf. (2)) for better illustration. Herein, the estimated parameters are  $\hat{A}_{\text{LOS}} = 101.70$ ,  $\hat{B}_{\text{LOS}} = 2.49$ ,  $\hat{A}_{\text{NLOS}} = 131.33$ ,  $\hat{B}_{\text{NLOS}} = 3.11$ , all in dB.

## 5. SIMULATIONS

We consider geolocation in a cellular network, which comprises  $M = 15$  base stations (anchors) and one mobile station (agent). The base stations are located in a two dimensional (2-D) plane with fixed coordinates in km scale as follows:  $(x_1 = -1.1, y_1 = 0.8)$ ,  $(x_2 = -1.1, y_2 = 0)$ ,  $(x_3 = -0.75, y_3 = 1)$ ,  $(x_4 = -0.6, y_4 = 0.1)$ ,  $(x_5 = -0.6, y_5 = -0.9)$ ,  $(x_6 = -0.5, y_6 = 0.2)$ ,  $(x_7 = -0.35, y_7 = 0.3)$ ,  $(x_8 = 0, y_8 = 0.4)$ ,  $(x_9 = 0, y_9 = 0)$ ,  $(x_{10} = 0.2, y_{10} = 0.1)$ ,  $(x_{11} = 0.4, y_{11} = 0.8)$ ,  $(x_{12} = 0.4, y_{12} = 0)$ ,  $(x_{13} = 0.75, y_{13} = -1)$ ,  $(x_{14} = 0.9, y_{14} = 0.6)$ ,  $(x_{15} = 1, y_{15} = -0.6)$ . The mobile station is located at  $(x = 0.3, y = 0.5)$  km. The transmit power is  $P_T = 40$  dBm. The mode-dependent propagation model parameters are  $\alpha_1 = 0.5$ ,  $\alpha_2 = 0.5$ ,  $A_{\text{LOS}} = 101.7$ ,  $B_{\text{LOS}} = 2.6$ ,  $A_{\text{NLOS}} = 132.8$ ,  $B_{\text{NLOS}} = 3.8$ ,  $\sigma_1 = 2$ , and  $\sigma_2 = 8$ , all in dB except for the first two terms.

In the first simulation, we study the estimation performance of the EM algorithm in a Monte Carlo experiment with 1500 independent trials. Here, we assume  $K = 5$  and start the EM algorithm with an initial guess  $\theta^{(0)} = [0.3, 0.7, 95, 2, 150, 4.5, 8, 80, 350, 400]^T$ . In each Monte Carlo trial, we record the EM estimate of  $\theta$  and ultimately, we calculate the mean of the parameter estimates over all trials. Due to space limitation, we only show the performance of the path loss model. Figure 1 lends some insight into how well the true path loss model (cf. (2)) can be estimated by the EM algorithm. It is obvious from the figure that the conventional single-mode propagation model is insufficient to be adopted in the given scenario. Moreover, the mean value of the path loss model parameters turn out to be  $\hat{A}_{\text{LOS}} = 101.70$ ,  $\hat{B}_{\text{LOS}} = 2.59$ ,  $\hat{A}_{\text{NLOS}} = 132.80$ ,  $\hat{B}_{\text{NLOS}} = 3.82$  (all in dB), which almost coincide with the true values.

In the second simulation, we study the geolocation accuracy by numerically evaluating the bias and root mean square error (RMSE) of the EM position estimator  $\hat{x}$ . The simulation parameters remain unaltered except for  $K$ . For the sake of comparison, we also demonstrate the performance achieved by matching a conventional single-mode propagation model (assuming  $r_k^{(m)}$  is always equal to 1 or 2



**Fig. 2.** Upper: Bias of the EM position estimator  $\hat{x} = [\hat{x}, \hat{y}]^T$  as a function of  $K$ . Bottom: RMSE of  $\hat{x}$  versus the best achievable geolocation accuracy,  $\text{CRLB}_{\text{pos}}$ , for different  $K$ .

in our signal model) with our mixed propagation environment. The model parameters  $A$ ,  $B$ , and  $\sigma^2$  are calculated by following a similar procedure as the one developed in [16], which can be treated as a special case of Algorithm 1. Besides, we compute the best achievable position RMSE, denoted by  $\text{CRLB}_{\text{pos}}$ , as follows:

$$\text{CRLB}_{\text{pos}} \triangleq \sqrt{([\mathcal{F}^{-1}(\theta)]_{9,9} + [\mathcal{F}^{-1}(\theta)]_{10,10})}$$

where Fisher information matrix  $\mathcal{F}(\theta)$  is numerically calculated using Monte Carlo integration techniques [17]. The results of the bias, RMSE, and  $\text{CRLB}_{\text{pos}}$  are shown in Fig. 2, where we observe the following. First, assuming a mode-dependent propagation model yields considerably improved geolocation accuracy. Second, the EM position estimator approaches the MLE as  $K$  increases. Furthermore, our algorithm only consumes circa 0.15 seconds on average (on a PC with Intel®Core™i5-760 processor and 8GB RAM) to generate a position estimate in this simulation, which makes it a good candidate for many real-time location-aware services.

## 6. RELATION TO PRIOR WORK AND CONCLUSIONS

In the literature, there exist many geolocation algorithms built on RSS measurements. Here, we narrow down our scope to the class of probabilistic algorithms where the RSS is assumed to be stochastic and influenced by factors such as path loss, multipath fading, and shadowing [2, Chapter 12]. The vast majority of studies consider geolocation based on a single-mode propagation model. For instance, [11, 12] estimate the model parameters *a priori* to the position estimation; while [16, 18, 19] assume some or all of the model parameters are unknown and propose to estimate them together with the position. In mixed LOS/NLOS environments, field-trial results motivate us to consider a mode-dependent propagation model. In order to avoid extra operational cost, we assume the mode-dependent model parameters are all unknown and conduct joint estimation based on the EM criterion. The designed position estimator well approximates the MLE and largely outperforms the one proposed in [16], where only a single-mode propagation model is considered. Moreover, it attains the best geolocation accuracy for large data records.

## 7. REFERENCES

- [1] F. Gustafsson and F. Gunnarsson, "Mobile positioning using wireless networks: possibilities and fundamental limitations based on available wireless network measurements," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 41–53, Jul. 2005.
- [2] R. Zekavat and R. M. Buehrer, *Handbook of Position Location*, John Wiley & Sons, Inc., Hoboken, NJ, 2011.
- [3] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1993.
- [4] A. P. Dempster, N. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. B*, vol. 39, pp. 1–38, 1977.
- [5] Y. Qi, *Wireless geolocation in a non-line-of-sight environment*, Ph.D. thesis, Princeton University, Princeton, NJ, USA, Nov. 2003.
- [6] E. Damasso and L. M. Correia, *COST Action 231: Digital Mobile Radio Towards Future Generation Systems : Final Report*, European Commission, Brussels, Belgium, 1999.
- [7] K. Yu and Y.-J. Guo, "Statistical NLOS identification based on AOA, TOA, and signal strength," *IEEE Trans. Veh. Technol.*, vol. 58, no. 1, pp. 274–286, Jan. 2009.
- [8] C. Fritsche, *Statistical Data Fusion for Hybrid Localization of Mobile Terminals*, Ph.D. thesis, Technische Universität Darmstadt, Darmstadt, Germany, Mar. 2011.
- [9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, New York, NY, USA, 1991.
- [10] S. Lipschutz, *Schaum's Outline of Theory and Problems of Probability*, Mc-Graw Hill, New York, NY, USA, 1974.
- [11] T. Roos, P. Myllymaki, and H. Tirri, "A statistical modeling approach to location estimation," *IEEE Trans. Mobile Computing*, vol. 1, no. 1, pp. 59–69, Jan.-Mar. 2002.
- [12] A. Coluccia and F. Ricciato, "On ML estimation for automatic RSS-based indoor localization," in *Proc. IEEE Int. Symp. Wireless Pervasive Comput.*, Modena, Italy, May 2010, pp. 495–502.
- [13] J. A. Bilmes, "A gentle tutorial on the EM algorithm and its application to parameter estimation for gaussian mixture and hidden markov models," Tech. Rep., University of California, Berkeley, Berkeley, CA, USA, 1998.
- [14] F. Yin and A. M. Zoubir, "Robust positioning in NLOS environments using nonparametric adaptive kernel density estimation," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 3517–3520.
- [15] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*, John Wiley & Sons, Inc., New York, NY, 2001.
- [16] R. Zemek, S. Hara, K. Yanagihara, and K.-I. Kitayama, "A joint estimation of target location and channel model parameters in an IEEE 802.15.4-based wireless sensor network," in *Proc. IEEE Int. Symp. Personal, Indoor and Mobile Radio Communications (PIMRC)*, Athens, Greece, Sep. 2007, pp. 1–5.
- [17] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer-Verlag, New York, NY, 1999.
- [18] X.-R. Li, "RSS-based location estimation with unknown pathloss model," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3626–3633, Dec. 2006.
- [19] H. C. So and L.-X. Lin, "Linear least squares approach for accurate received signal strength based source localization," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 4035–4040, Aug. 2011.