A Novel Geometry-Based Model for Localization Based on Received Signal Strength

1st Zhihe Li
Dept. of Electronic Engineering
Tsinghua University
Beijing, China
lizhihe16@mails.tsinghua.edu.cn

2nd Xiaofeng Zhong
Dept. of Electronic Engineering
Tsinghua University
Beijing, China
zhongxf@tsinghua.edu.cn

3rd Jie Wei
School of Electronic and Information Engineering
Beijing Jiaotong University
Beijing, China
jwei@bjtu.edu.cn

***Abstract*—A novel geometry-based model (GBM) is proposed to locate the transmitter based on received signal strength (RSS), where there is no anchor point. Previous studies were mostly based on maximum likelihood (ML) estimator. The novel GBM estimator gives a new approach for localization based on the geometrical significance, which is to find a point with least sum of distance to the edge of circles. The corresponding Cramér-Rao Lower Bound (CRLB) of unbiased estimator and a time-saving Monte Carlo approach to the CRLB of GBM are derived. In addition, the confidence interval of Root-Mean-Square Error (RMSE) is also derived based on Central Limit Theorem in order to make sure the results of simulation are statistical significant. The simulation results demonstrate that GBM estimator can achieve much higher accuracy compared to ML. The proposed GBM algorithm is verified with experimental data set from commercial cellular networks, and shows better performances.**

***Index Terms*—geometry-based model (GBM), maximum likelihood (ML), Cramér-Rao Lower Bound (CRLB), confidence interval (CI), received signal strength (RSS), data set from commercial cellular networks.**

I. Introduction

Localization of a transmitter is of vital importance in civilian and military applications, such as localizing an emergency call or an illegal broadcasting station. In addition, the localization of sensors in a wireless sensor network has been the subject of great interest in many studies, which needs a location estimator with higher accuracy. Root-Mean-Square Error (RMSE) and Absolute-Mean Error (MAE) are usually used to evaluate estimators.

There are many features can be applied to implement localization, including received signal strength (RSS) [1] [2] [3], time-of-arrival [4], time-difference-of-arrival [5], and angle-of-arrival [6]. RSS based method is popular among them because of its low requirement for hardware. The only measurements needed are receivers' location and RSS.

The RSS based methods could be classified into two types, those with or without anchor points. Anchor point is a point whose distance to transmitter is known. The existence of anchor point is common for wireless sensor network but it is a strong assumption in some situation, such as localizing an illegal broadcasting station. RSS difference is a method applied in the situation without anchor point. Here we have assumed that the transmitter's power does not change and it is commonly satisfied. In the following part, we will only discuss methods without anchor point.

Maximum likelihood (ML) estimator is the most popular estimator and baseline. There are many other ML based estimators proposed, but neither semi-definite programming (SDP) nor linear least square (LLS) can achieve lower error than ML [2]. Since what we need here is baseline for accuracy, ML estimator will be utilized directly.

Two LLS approaches were proposed in the paper [7]. In the process of derivation, it demonstrated that every two receivers can form a circle where the transmitter located on. However, instead of researching deeply into this phenomenon, two LLS approaches were derived. We will locate the transmitter based on the geometrical significance of these circles, which is to find a point with least sum of distance to the edge of these circles. Another approach based on circular positioning algorithm to locate transmitter was also proposed [3]. But the application scenario of this paper is anchor point enabled and Cramér-Rao Lower Bound (CRLB) of this method was not analyzed. CRLB is the lower bound of error in theory of estimators.

In this paper, we will propose a geometry-based model (GBM), whose geometrical significance is explained above. Actually, we can see that the objective function can be explained as a joint distribution of independent identical Truncated Normal Distribution. In simulation part, we will see that GBM estimator is biased, so the unbiased CRLB is meaningless for GBM and method to obtain the CRLB of GBM is necessary. We proposed a time-saving Monte Carlo based method here. In addition, in order to judge whether the results of simulation are statistically significant, we also proposed a Central Limit Theorem based method to calculate the confidence interval of root-mean-square error (RMSE). The simulation results demonstrated that GBM estimator can approach the RMSE of ML for little noise but much higher accuracy for higher noise compared to ML and CRLB has the similar result. Moreover, we also utilized these estimators to experimental data set from commercial cellular networks so as to verify the results of simulation and test the practicability of estimators. The results of measured data demonstrated that GBM estimator has much higher accuracy, which told us GBM is more robust than ML.

The rest of the paper is organized as follows. Section II introduces path loss model and conventional maximum likelihood estimator, derives geometry-based model, and explains the relationship between bias, variance and mean square error. The formula of unbiased CRLB and the method based on Monte Carlo to calculate CRLB of GBM are derived in Section III. Section IV describes how to calculate the confidence interval of RMSE. The simulation results and estimator comparisons are discussed in Section V. Section VI includes introduction of measured data, preprocess method, results of estimators and the corresponding analysis. Section VII concludes this paper.

Notation. Throughout the paper, the following notations are used. Lowercase and uppercase letters denote scalar values. Bold uppercase letters and bold lowercase letters denote matrices and vectors. $\|\cdot\|_2$ is two norm which means the Euclidean distance.

II. Path Loss Model and Estimators

A. Path Loss Model

The RSS of receiver at location $\mathbf{x}_i$ in the 2-D plane can be formulated by the following formula [8],

$$P_i = C - 10\alpha\log_{10}(\|\boldsymbol{\theta} - \mathbf{x}_i\|_2) + n_i, i = 1, 2, ..., N, \quad (1)$$

where P_i is the received signal strength (RSS), $\boldsymbol{\theta}$ is the transmitter location, α is the path loss exponent, C is a constant which is independent with $\boldsymbol{\theta}$ and $\mathbf{x}_i$, n_i is Gaussian noise with mean of 0 and variance of ${\sigma_i}^2$, and i is the index of receivers. In order to simplify formulation, we have

$$d_i = \|\boldsymbol{\theta} - \mathbf{x}_i\|_2. \quad (2)$$

Since C is a constant which is independent with $\boldsymbol{\theta}$ and $\mathbf{x}_i$, it can be eliminated by subtracting another receiver's RSS, which means:

$$P_{ij} = 10\alpha(\log_{10}d_j - \log_{10}d_i) + n_{ij}, i \neq j, \quad (3)$$

where $P_{ij} = P_i - P_j$, n_{ij} is Gaussian noise with mean of 0 and variance of $\sigma_i^2 + \sigma_j^2$. Noted that we dont have anchor point, let j be the index of receiver with maximum RSS. And in the following part, we will denote $\mathbf{x}_i, d_i$ as $\mathbf{x}_M, d_M$, where M is the index of receiver with maximum RSS.

B. Conventional Maximum Likelihood (ML) Estimator

The conventional ML Estimator is to find a $\hat{\boldsymbol{\theta}}$ which satisfies the following optimization problem:

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{j=1}^{N} [P_{Mj} - 10\alpha(\log_{10}d_j - \log_{10}d_M)]^2, j \neq M. \quad (4)$$

Since it is a non-convex problem, many convex (such as semi-definite programming, SDP) or linear least square (LLS) estimators have been proposed, as mentioned in Introduction. As ML estimator is for comparison here, we need to take the algorithm that can achieve the lowest error as competitor, so we optimize the non-convex nonlinear least square loss function directly. Here we randomly choose initial point multiple times and the solution with the least loss is selected.

C. Geometry-Based Model (GBM) Estimator

A LLS esimator was proposed [7], where each two receivers forming a circle was derived in the process of derivation, where transmitter located on. However, this knowledge was not used for optimization directly. Instead, two circle equations were subtracted to obtain a linear equation, which is the LLS approach. In the following, we will utilize above knowledge to obtain a geometry-based model.

If we ignore noise in (3), we can get

$$P_{ij} - 5\alpha\log_{10}(\frac{d_j}{d_i})^2 = 0, i \neq j. \quad (5)$$

Let

$$\gamma = 10^{\frac{P_{ij}}{5\alpha}}, \quad (6)$$

then (5) can be rewritten as

$$\|\boldsymbol{\theta} - \mathbf{x}_j\|_2{}^2 = \gamma_{ij}\|\boldsymbol{\theta} - \mathbf{x}_i\|_2{}^2. \quad (7)$$

If $\gamma_{ij} \neq 1$, which means the RSS of two receivers are different, (7) can be rewritten as

$$\left\|\boldsymbol{\theta} - \frac{\mathbf{x}_j - \gamma_{ij}\mathbf{x}_i}{1 - \gamma_{ij}}\right\|_2^2 = \frac{\gamma_{ij}{d_{ij}}^2}{(1 - \gamma_{ij})^2}, \quad (8)$$

where $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2$. obviously it is a circle where $\boldsymbol{\theta}$ locates on, whose center and radius are

$$\mathbf{o}_j = \frac{\mathbf{x}_j - \gamma_{ij}\mathbf{x}_i}{1 - \gamma_{ij}}, {r_j}^2 = \frac{\gamma_{ij}{d_{ij}}^2}{(1 - \gamma_{ij})^2}. \quad (9)$$

Let i be the index of receiver with maximum RSS (If there are more than one, randomly choose one and delete the others) and replace i with M. Suppose that (8) is affected by the following additive noise:

$$\|\boldsymbol{\theta} - o_j\|_2{}^2 = d_j + n_j, j \neq M, \quad (10)$$

where n_j is noise obeying Truncated Normal Distribution with Gaussian coefficient $\mu = 0, \sigma = \sigma_j$. Note that the noise n_j here is independent with noise in (1) since these two equations are different independent models. Since $\boldsymbol{\theta}$ is real, the right side of (10) must be equal or more than 0, which means the value range of n_j is $[-d_j, +\infty]$. So we build a model with n_j obeying Truncated Normal Distribution.

With the model above, we can define a random variable $X_j = \|\boldsymbol{\theta} - \mathbf{o}_j\|_2$ obeying Truncated Normal Distribution, whose probability density function is

$$f(x_j) = \frac{1}{C(r_j)}\exp(-\frac{(x_j - r_j)^2}{2{\sigma_j}^2})I[0, +\infty), \quad (11)$$

where $C(r_j) = \int_0^{+\infty}\exp(-\frac{(x_j-r_j)^2}{2{\sigma_j}^2})dx_j$ is the normalization coefficient and $I[0, +\infty)$ is the indicator whose value is 1 in $[0, +\infty)$ else is 0.

Hence, the loss function and optimization problem can be written as

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \frac{1}{2}\sum_{j=1}^{n}\frac{1}{\sigma_j^2}(\|\boldsymbol{\theta}-\mathbf{o}_j\|_2 - r_j)^2, \tag{12}$$

where n is the number of left points. Since $C(r_j)$ is independent with $\boldsymbol{\theta}$, we can ignore them when optimizing.

When σ_j is the same, the optimization problem is to find a point whose sum of distance to the edge of these circles is minimum. So this optimization problem has a distinct geometry significance.

Since the convexity of GBM can not be proved, we will utilize the same method here as ML, which is randomly choosing initial point multiple times and the solution with the least loss is selected.

In simulation (Section V) it will demonstrate that GBM estimator is biased. However, our goal is to minimize Mean Square Error (MSE) and MSE can rewritten as

$$\begin{aligned} MSE &= \mathrm{E}[(\widehat{\theta}-\theta)^2] \\ &= (\mathrm{E}[\widehat{\theta}]-\theta)^2 + \mathrm{E}[(\widehat{\theta}-\mathrm{E}[\widehat{\theta}])^2] \\ &= bias^2 + variance. \end{aligned} \tag{13}$$

It is also a classic problem occurring in Machine Learning. Estimator with small bias and big variance is called "over-fitting". In simulation and measured data verification part, we will show that conventional ML estimator is "over-fitting" compared to GBM.

III. Cramér-Rao Lower Bound (CRLB)

A. CRLB of Unbiased Estimator

For the path loss model, since the ith receiver's RSS $p_i \sim \mathcal{N}(-10\alpha\log_{10}d_i - C, {\sigma_i}^2)$, Fisher information can be obtained by

$$I(\boldsymbol{\theta}) = (\frac{10\alpha}{\sigma\ln 10})^2 \sum_{i=1}^{N}(\frac{\boldsymbol{\theta}-\mathbf{x}_i}{{\|\boldsymbol{\theta}-\mathbf{x}_i\|_2}^2})(\frac{\boldsymbol{\theta}-\mathbf{x}_i}{{\|\boldsymbol{\theta}-\mathbf{x}_i\|_2}^2})^T, \tag{14}$$

where σ is the standard deviation of noise. If the estimator is unbiased, CRLB covariance matrix can be obtained by

$$\begin{aligned} CRLB(\boldsymbol{\theta}) &= I(\boldsymbol{\theta})^{-1} \\ &\leq Cov(\widehat{\boldsymbol{\theta}}_{ML}). \end{aligned} \tag{15}$$

Since ML estimator is asymptotically efficient [9], according to (13), the approximate lower bound of ML RMSE is

$$RMSE \geq \sqrt{Tr(CRLB)}, \tag{16}$$

where $Tr(\cdot)$ is the trace of matrix.

B. CRLB of Biased Estimator

Suppose that the expectation of GBM estimator is $\boldsymbol{\phi}(\boldsymbol{\theta})$. If the estimator is biased, the CRLB covariance matrix can be obtained by

$$\begin{aligned} CRLB_{GBM}(\boldsymbol{\theta}) &= \frac{\partial\boldsymbol{\phi}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} I(\boldsymbol{\theta})^{-1} \frac{\partial\boldsymbol{\phi}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}^T \\ &\leq Cov(\widehat{\boldsymbol{\theta}}_{GBM}), \end{aligned} \tag{17}$$

where $\frac{\partial\boldsymbol{\phi}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$ is the Jacobian matrix. Note that if estimator is unbiased, then $\boldsymbol{\phi}(\boldsymbol{\theta}) = \boldsymbol{\theta}$ and $\frac{\partial\boldsymbol{\phi}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \boldsymbol{I}$, which coincides with (15).

According to (13), the lower bound of GBM RMSE is

$$RMSE \geq \sqrt{(\boldsymbol{\phi}(\boldsymbol{\theta})-\boldsymbol{\theta})^T(\boldsymbol{\phi}(\boldsymbol{\theta})-\boldsymbol{\theta}) + \mathrm{Tr}(CRLB_{GBM}(\boldsymbol{\theta}))}. \tag{18}$$

It is difficult to derive $\boldsymbol{\phi}(\boldsymbol{\theta})$ and $\frac{\partial\boldsymbol{\phi}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$ directly. However, we know the distribution of receivers' RSS, Monte Carlo method is suitable for this problem.

Firstly, it is easy to obtain $\boldsymbol{\phi}(\boldsymbol{\theta})$ by

$$\begin{aligned} \boldsymbol{\phi}(\boldsymbol{\theta}) &= \mathrm{E}[\widehat{\boldsymbol{\theta}}] \\ &\approx \frac{\sum_{k=1}^{K}\widehat{\boldsymbol{\theta}}}{K}, \end{aligned} \tag{19}$$

where K is the total number of simulation times.

$\frac{\partial\boldsymbol{\phi}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$ can be obtained by using finite difference method directly, but it can be time-consuming since we should simulate with different $\boldsymbol{\theta}$ in the 2-D plane.

In the following, we will derive a method to obtain $\frac{\partial\boldsymbol{\phi}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$ without simulating with different $\boldsymbol{\theta}$. we first write down the formulation of $\boldsymbol{\phi}(\boldsymbol{\theta})$

$$\begin{aligned} &Let \quad f(\boldsymbol{\theta}) = \prod_{i=1}^{N}(2\pi{\sigma_i}^2)^{-\frac{1}{2}} e^{-\frac{(p_i+10\alpha\log_{10}d_i+C)^2}{2{\sigma_i}^2}}, \\ &then \quad \boldsymbol{\phi}(\boldsymbol{\theta}) = \iiint_{\mathbb{R}^N}\widehat{\boldsymbol{\theta}}f(\boldsymbol{\theta})dp_1...dp_N, \end{aligned} \tag{20}$$

where d_i is defined in (2). Suppose that $\widehat{\boldsymbol{\theta}}f(\boldsymbol{\theta})$ and $\frac{\partial\widehat{\boldsymbol{\theta}}f(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$ is continuous, $\iiint_{\mathbb{R}^N}\widehat{\boldsymbol{\theta}}f(\boldsymbol{\theta})$ and $\iiint_{\mathbb{R}^N}\frac{\partial\widehat{\boldsymbol{\theta}}f(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$ is uniformly convergent, then

$$\begin{aligned} \frac{\partial\boldsymbol{\phi}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} &= \frac{\partial\mathrm{E}[\widehat{\boldsymbol{\theta}}]}{\partial\boldsymbol{\theta}} \\ &= \frac{\partial}{\partial\boldsymbol{\theta}}\iiint_{\mathbb{R}^N}\widehat{\boldsymbol{\theta}}f(\boldsymbol{\theta})dp_1...dp_N \\ &= \iiint_{\mathbb{R}^N}\frac{\partial\widehat{\boldsymbol{\theta}}f(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}dp_1...dp_N \\ &= \iiint_{\mathbb{R}^N}\sum_{i=1}^{N}\widehat{\boldsymbol{\theta}}(\frac{-10\alpha}{{\sigma_i}^2\ln 10}\frac{\boldsymbol{\theta}-\mathbf{x}_i}{{\|\boldsymbol{\theta}-\mathbf{x}_i\|_2}^2})^T f(\boldsymbol{\theta})dp_1...dp_N \\ &= \mathrm{E}[\sum_{i=1}^{N}\widehat{\boldsymbol{\theta}}(\frac{-10\alpha}{{\sigma_i}^2\ln 10}\frac{\boldsymbol{\theta}-\mathbf{x}_i}{{\|\boldsymbol{\theta}-\mathbf{x}_i\|_2}^2})^T]. \end{aligned} \tag{21}$$

Note that $\widehat{\boldsymbol{\theta}}$ is independent with $\boldsymbol{\theta}$ but a function of $(\mathbf{x}_1, ..., \mathbf{x}_N)^T$ and $(P_1, ..., P_N)^T$.

According to (21), we can utilize Monte Carlo method to approximate $\frac{\partial\boldsymbol{\phi}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$ by calculating the empirical average of $\sum_{i=1}^{N}\widehat{\boldsymbol{\theta}}(\frac{-10\alpha}{{\sigma_i}^2\ln 10}\frac{\boldsymbol{\theta}-\mathbf{x}_i}{{\|\boldsymbol{\theta}-\mathbf{x}_i\|_2}^2})^T$ instead of having to use finite difference method, which can save time in simulation.

IV. Confidence Interval of RMSE

The accuracy of estimators is affected by noise. So in order to distinguish whether different estimators are statistically significantly different, we need to import confidence interval (CI) of RMSE.

However, the distribution of RMSE is unknown. Fortunately, according to Central Limit Theorem, when the number of simulation times is large and $(\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})^T(\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})$ is independent identically distributed for different k, the distribution of MSE will tend to normal distribution, which means

$$\lim_{K\to\infty} MSE = \lim_{K\to\infty} \frac{\sum_{k=1}^{K}(\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})^T(\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})}{K} \sim \mathcal{N}(\mu, \frac{\sigma^2}{K}), \tag{22}$$

where μ and σ^2 are the expectation and variance of $(\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})^T(\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})$. Since we don't know the expectation and variance, we should use the empirical average instead, which means we should use T-test. However, when the number of simulation times is large (it is 100,000 in our situation), we can directly utilize Z-test. So we can derive the 95% confidence interval by

$$CI = (\sqrt{RMSE^2 - 1.96\frac{\sigma}{\sqrt{K}}}, \sqrt{RMSE^2 + 1.96\frac{\sigma}{\sqrt{K}}}). \tag{23}$$

Hence, we can use (23) to obtain the confidence interval of RMSE and check out if the RMSE of two estimators are statistically significantly different.

V. Numerical Simulation

A. Simulation Scene and Estimator Setting

A simulation experiment is performed to compare the performance of Geometry-Based Model (GBM), conventional Maximum Likelihood (ML) and Linear Least Square (LLS) (equation (29) in [7]). The scene of simulation is the same as that in [7]. The transmitter is located at (0, 0) and 10 receivers are located at (-126, 62), (-339, 166), (390, 76), (-178, 333), (482, 122), (-234, 141), (230, -45), (-352, 279), (-104, 465), (366, -139). The coordinate above is absolute and unit is meter (m). When generating data, the path loss exponent is set to be $\alpha = 4$, noise is normally distributed with 0 mean and standard deviation σ varies from 1 to 15 and C is set at -40.

Since both ML and GBM estimators are non-convex, we estimate the transmitter with different initial point 20 times and location with minimum loss is selected to be the final location. The coordinate of initial points are sampled from two independent identical uniform distribution $\mathcal{U}(-500, 500)$. Noted that we don't have anchor point, so we must choose a receiver as subtrahend. Receiver with maximum RSS is chosen here. For GBM, if there are more than one maximum RSS receivers, one of them is randomly chosen and the others are deleted. The σ_j in GBM are all set as 1. All of the three estimators (GBM, ML, LLS) will be simulated 100,000 times and then RMSE and confidence interval will be calculated. In addition, CRLB of unbiased estimator and the CRLB of GBM will be calculated according to (16) and (18).

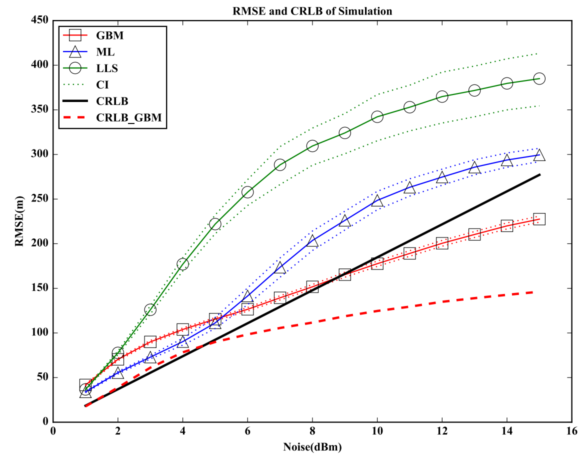

Fig. 1. RMSE, confidence interval and CRLB of 100,000 iterations of simulation.

B. Simulation Result

Fig. 1 demonstrates the RMSE, confidence interval and CRLB of simulation with 100,000 iterations. We can see that when the standard deviation of noise is less than 5dBm, the RMSE of ML is a bit lower than GBM. However, when the standard deviation of noise is more than 5dBm, the RMSE of GBM is much lower than ML and it is statistically significant. It is a common knowledge that the real-world scenarios are usually heavily noised because of multipath effect and shadowing effect. In next section, we will apply all of the above estimators to measured data and the results will give us some insight. What's more, the CRLB has the similar result compared to GBM and ML. Although the CRLB of GBM is much lower than the RMSE of GBM, we don't know if there exists an efficient estimator that can achieve the lower bound. It may be an interesting future research field.

Note that when the standard deviation of noise is more than 9, the RMSE of GBM will be lower than the CRLB of unbiased estimator, which implies that GBM is biased otherwise CRLB is the lower bound. According to (13), it means although GBM is a biased estimator, the increment of bias is little compared to the decrement of variance. In the view of Machine Learning, ML is an estimator with low bias and high variance, which is 'over-fitting' to the ideal path loss model, and GBM is more like a regularized version, which can better fit the data with larger variance of noise.

Finally, LLS estimator gets the highest RMSE. It has been shown that ML can get lower RMSE compared to LLS [2] and the performance of LLS estimator is affected heavily when noise is high [1], which coincide with our result.

VI. Measured Data Verification

In order to verify the results of simulation and test the practicability of estimators, we also test on experimental data

TABLE I
MEASURED DATA RESULTS

Estimator	GBM	ML	LLS
RMSE [m]	367.52	453.33	4898.10
MAE[a] [m]	282.43	380.08	936.19

[a]Mean Absolute Error.

set from commercial cellular networks.

A. Data Introduction

The data are collected by driving cars with base station signal receivers in cities. There are more than 120,000 base stations in data set. Data are collected every 5sec with Global Positioning System (GPS) coordinates, RSS, LAC (Location Area Code), CID (Cell ID) and other base station coefficients. In order to minimize the impact of shadowing, Global System for Mobile (GSM) signal is chosen for testing here. The frequency of GSM is 900 MHz and 1800 MHz.

Before localization, the data need to be preprocessed. Firstly, we need to convert GPS coordinates to Cartesian coordinates. Secondly, In order to reduce computing time and the effect of shadowing, points with RSS less than -60dBm will be deleted. Base station that satisfies all of the following conditions will be chosen:

- The type of signal is GSM;
- There are more than 10 receiver points in the range of base station;
- Distance of the nearest point to base station is less than 1,000 m.

After selection, there are 17321 base stations left and we will use them for testing. The average radius of base stations is 751m, which is taken as the average distance of the farthest point selected in cell to base station. We take path loss exponent as 4 for all three estimators.

B. Measured Data Results

As Table I demonstrates, GBM can both get the minimum in RMSE and MAE (Mean Absolute Error) compared to ML and LLS. Where MAE is computed by

$$MAE = \frac{\sum_{k=1}^{K} \left\| \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right\|_2}{K}. \tag{24}$$

According to simulation, we can derive that measured data is affected heavily by noise so that GBM can get much lower error, whose advantage is nearly 100m less in error which is 34.75% in MAE and 23.35% in RMSE relatively. It enlightens us that GBM is more practical to commercial cellular networks, which is frequently heavily noised, compared to ML. In addition, the score of LLS confirms the claim that the performance of LLS estimator is affected heavily when noise is high.

VII. CONCLUSION

In this paper, a novel geometry-based model is proposed to locate the transmitter based on RSS. In addition, a time-saving Monte Carlo approach to the CRLB of GBM RMSE and a method based on Central Limit Theorem to obtain confidence interval of RMSE are also proposed. Simulation results demonstrate that GBM can achieve much higher accuracy for lager variance of noise compared to ML, which means it is more robust. The CRLB of unbiased estimator and GBM have the similar results compared to above result. In order to verify the results of simulation and test the practicability of estimators, experimental data from commercial cellular networks are also collected for testing. The results of measured data demonstrate that GBM can achieve the lowest both in RMSE and MAE, which enlightens us that GBM may be more practical to real-world scenarios.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China (No. 61631013), and the Fundamental Research Funds for the Central Universities (No. 2016JBM010).

REFERENCES

[1] R. M. Vaghefi, M. R. Gholami, R. M. Buehrer, and E. G. Strom, "Cooperative received signal strength-based sensor localization with unknown transmit powers," *IEEE Transactions on Signal Processing*, vol. 61, no. 6, pp. 1389–1403, 2013.

[2] B. R. Jackson, S. Wang, and R. Inkol, "Received signal strength difference emitter geolocation least squares algorithm comparison," in *Electrical and Computer Engineering (CCECE), 2011 24th Canadian Conference on*. IEEE, 2011, pp. 001 113–001 118.

[3] B.-C. Liu, K.-H. Lin, and J.-C. Wu, "Analysis of hyperbolic and circular positioning algorithms using stationary signal-strength-difference measurements in wireless communications," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 2, pp. 499–509, 2006.

[4] I. Guvenc and C.-C. Chong, "A survey on toa based wireless localization and nlos mitigation techniques," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 3, 2009.

[5] A. Catovic and Z. Sahinoglu, "The cramer-rao bounds of hybrid toa/rss and tdoa/rss location estimation schemes," *IEEE Communications Letters*, vol. 8, no. 10, pp. 626–628, 2004.

[6] L. Cong and W. Zhuang, "Hybrid tdoa/aoa mobile user location for wideband cdma cellular systems," *IEEE Transactions on Wireless Communications*, vol. 1, no. 3, pp. 439–447, 2002.

[7] S. Wang and R. Inkol, "A near-optimal least squares solution to received signal strength difference based geolocation," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2600–2603.

[8] V. Erceg, L. J. Greenstein, S. Y. Tjandra, S. R. Parkoff, A. Gupta, B. Kulic, A. A. Julius, and R. Bianchi, "An empirically based path loss model for wireless channels in suburban environments," *IEEE Journal on selected areas in communications*, vol. 17, no. 7, pp. 1205–1211, 1999.

[9] S. M. Kay, *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.