INCREMENTAL LOCALIZATION ALGORITHM BASED ON REGULARIZED ITERATIVELY REWEIGHTED LEAST SQUARE

Xiaoyong YAN^{1,2}, Zhong YANG^{3*}, Yu LIU^{4*}, Xiaoduo XU⁵, Huijun LI^{6,7}

Abstract. Incremental localization algorithm is a distributed localization method with excellent characteristics for wireless network. However, its estimated result is generally influenced by the heteroscedasticity arising from cumulative errors and the collineation among anchor nodes. We have proposed a novel incremental localization algorithm with consideration to cumulative errors and collinearity among anchors. Using iteratively reweighted and regularized method, the algorithm reduces the influences of errors accumulation and avoids collinearity problem between anchors. Simulation experiment results show that compared with the previous incremental localization algorithms, the proposed algorithm obtains a localization solution which not only has high accuracy but also high stability. Therefore, the proposed algorithm is suitable for different deployment environments and has high adaptability.

Keywords: wireless network, incremental localization, regularized iteratively reweighted least square, heteroscedasticity.

1. Introduction

Wireless and network have been research hotspot and focus in recent years [1]. In the research of wireless network, location information and wireless network are closely linked, since the

1. School of Intelligence Science and Control Engineering, Jinling Institute of Technology, Nanjing, China; 2. Remote Measurement and Control Key Lab of Jiangsu Province, Nanjing, China, Email: xiaoyong_yan@126.com

3. *Corresponding author, School of Intelligence Science and Control Engineering, Jinling Institute of Technology, Nanjing, China, Email: yz@jit.edu.cn

4. *Corresponding author, School of Computer Engineering, Jinling Institute of Technology, Nanjing, China, Email: ayu987@jit.edu.cn

5. School of Software and Microelectronics, Northwestern polytechnical university, Xi'an, China, Email: wsn2010@yeah.net

6. School of Instrument Science and Engineering, Southeast University, Nanjing, China;

7. Remote Measurement and Control Key Lab of Jiangsu Province, Nanjing, China, Email: lihuijun@seu.edu.cn

location information is not only calculated through the wireless network, but also reacts on the wireless network. That is to say, wireless network sensory data with geographic information can assist other wireless network functions and management. And for that reason, with the development of wireless communication technology, wireless localization technology has attracted more and more attentions, and has become another research hotspot after wireless and network [9].

The close contact between wireless localization technology and people's daily life benefits from the popularity of the satellite localization system such as GPS. With a base station as the center, satellite localization system is a long distance localization system, which by way of one-hop uses terminal equipments such as receiver to directly communicate with the satellite base station in the sky to construct a star network in the vast outdoor area [12]. And the indoor scenes, cities with intensive buildings, forests and underground environments belong to the medium and short-range system, where signal propagation environment is harsher and wireless localization system is required to have very powerful wireless capabilities and digital signal processing chip, and more sophisticated wireless localization algorithms [26]. With the development of medium and short range wireless ad-hoc network technology represented by the sensor network and the Internet of things, wireless location technology also correspondingly upgrades from the base station localization, that is, one-hop localization to network localization, which is multi-hop localization. Also is to say, a few nodes, commonly called anchors, manually configure location or equip with GPS receiver, and a large number of remaining nodes first obtain the relative position with neighbour anchors, then exchange data with each other through wireless network to obtain the global location [16].

In medium and short range systems, wireless localization algorithms can be further subdivided into one-hop and multi-hop localization. If the distance between nodes is less than 100 meters, it belongs to short range systems [27]; physical measurement is directly adopted for unknown nodes and anchors, and one-hop localization for location estimation is feasible; if the distance ranges from 100 meters to 2000 meters between nodes, it is medium range, and multi-hop localization is available, i.e., gradually spread to the edge of the network through incremental or iterative method from the vicinity of the anchors [24] (Figure 1 (a)) or indirectly convert hops into physical distances [25] (Figure 1 (b)).

Figure 1. Multi-hop localization

In the multi-hop localization method, the target nodes cannot directly communicate with the anchors, and the location estimation method and organization form become more complex, and computation load and communication traffic are even greater. In addition, some

problems difficult to be faced for one-hop localization will be encountered, such as nodes density, error propagation *etc.* According to the implementation and organization form of the algorithm, the multi-hop localization algorithms can be divided into: centralized algorithms and distributed algorithms [19]. In the centralized algorithms, each node in the network is responsible for returning the measurement results to the base station for the calculation and analysis. Centralized algorithms often require at least one node with strong computing ability in the network, and a large amount of communication is produced in data returning process, so the excessively complex calculation of centralized multi-hop localization hinders its wide application in practical problems. Common centralized localization algorithms such as: multidimensional scaling-based localization (MDS-based) [18], semi-definite programming-based localization (SDP-based) [6], kernel locality preserving projection-based localization (KLPP-based) [21], *etc.*

Distributed algorithms make up the defect of insufficient computing ability caused by the lack of centralized computing through parallel processing and communication among nodes, at the same time effectively reducing the high overhead of the communication between nodes and a central server. Different from the calculation method of using all measuring data for one time and localizing all network nodes in centralized algorithms, distributed algorithms usually only use a subset of measurement data at a time to gradually localize other network nodes, so as to achieve the approximate localization results corresponding to centralized algorithms. Common distributed location algorithms are incremental or iterative localizations. As shown in Figure 1 (a), initially an arbitrary unknown node tries to determine its own location through ordinary least square (OLS) or maximum likelihood (ML) [15] according to the location of its neighbour node; after determining its location, the node can be a new reference node, then assist in locating other unknown nodes. This process is repeated until all unknown nodes become known nodes, so that the entire network localization is realized. The advantage of these methods is that it only involves local anchors when locating nodes, thus avoiding too much communication traffic. However, as with a number of key technologies, in practical applications the performance of incremental localization techniques still faces many technical problems, of which the most deadly one is the error accumulation problem, that a prior localization error affects the next locating performance, and the influence is cumulative. The accumulation of the errors will inevitably lead to the difference between the variance of the previous locating error and the variance of the posterior locating error, which is known as heteroscedasticity [5,13]. If heteroskedasticity appears in the process of locating and estimating, the location estimation of the unknown nodes using the OLS may not be efficient estimators, nor even asymptotically efficient estimators. In incremental localization process, if the error accumulation cannot be effectively processed, the obtained degree of variation of estimated values increases, resulting in the increasing of the prediction errors of the coordinates of the unknown nodes and the reduction of prediction accuracy, so the estimated results sometimes become meaningless. In addition, in the process of locating the collinearity between original anchors, new anchors or original anchors and new anchors will have a significant impact on the location estimation of unknown nodes. In this paper, we focus on the problem of heteroskedasticity and collinearity in incremental localization and propose a novel incremental localization, namely Incremental localization algorithm-regularized iteratively reweighted least square (IL-RIRLS). We choose iteratively reweighted least square to reduce the cumulative impact of errors, and regularization method to weaken the collinearity problem.

The remainder of the paper is structured as follows. Section II presents the process of incremental localization algorithm. And section III shows our localization method performance, and section V gives the conclusions.

2. Incremental localization by regularized iteratively reweighted least square

2.1. Error analysis

Consider that $m+n$ nodes $\{S_i\}_{i=1}^{m+n}$ are deployed in a plane area $c \subseteq \mathbb{R}^2$. Assume that $x_i = (x_i, y_i) \in \mathbb{R}^2$ denotes the real coordinate of node S_i . Assume that the locations of the first m nodes are anchors whose locations are known, while the locations of the remaining n nodes are unknown. Suppose that each node needs to be able to transmit the locating data to the neighbour nodes within the communication radius. As shown in Figure 1(a), in the incremental localization methods there are three kinds of nodes, namely the original anchor, unknown node and the new anchor updated from the unknown node. Initially an unknown node, if possible, is located based on its neighbours by OLS or ML. After being aware of its location, it becomes a new anchor to localize other unknown nodes in the subsequent process. This step continues incrementally, gradually changing no-anchor to known ones. However, the use of localized unknown nodes as new reference inherently introduces substantial cumulative errors, especially for the nodes far away from anchors. In addition, the collinearity among anchors (original anchors or new anchors) affects the results of location estimation.

It is assumed that the location of the original anchor has no deviation, but due to the measurement error of distance between nodes, the estimated location error of the new anchor node exists, and that in the subsequent locating process, the location error of the new anchor node and the error of distance measurement are bound to enlarge the next localization error. And the collinearity problem can be analyzed by the following derivation.

Assume that in a 2D geographic area the unknown node u receives more than k ($k \geq 3$) anchors signals in the range of effective communication radius. For incremental localization, an anchor can be an original anchor and can also be a new anchor. There exists a coordinate-distance equation between anchors and unknown nodes, *i.e.*,

$$\begin{cases} (x-x_1)^2 + (y-y_1)^2 = d_1^2 \\ (x-x_2)^2 + (y-y_2)^2 = d_2^2 \\ \vdots \\ (x-x_k)^2 + (y-y_k)^2 = d_k^2 \end{cases} \quad k \geq 3 \quad (1)$$

Where (x, y) is the location of the unknown node, and $(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)$ are the coordinates of anchors. If subtract the k -th equation with the equations from the first to the $(k-1)$ -th equation, respectively, it is obtained that:

$$\begin{cases} 2(x_k - x_1)x + 2(y_k - y_1)y = d_1^2 - d_k^2 + y_k^2 + x_k^2 - y_1^2 - x_1^2 \\ 2(x_k - x_2)x + 2(y_k - y_2)y = d_2^2 - d_k^2 + y_k^2 + x_k^2 - y_2^2 - x_2^2 \\ \vdots \\ 2(x_k - x_{k-1})x + 2(y_k - y_{k-1})y = d_{k-1}^2 - d_k^2 + y_k^2 + x_k^2 - y_{k-1}^2 - x_{k-1}^2 \end{cases} \quad (2)$$

Set,

$$\mathbf{A} = 2 \times \begin{bmatrix} (x_1 - x_n) & (y_1 - y_n) \\ (x_2 - x_n) & (y_2 - y_n) \\ \vdots & \vdots \\ (x_{k-1} - x_k) & (y_{k-1} - y_k) \end{bmatrix} \quad (a)$$

$$\mathbf{b} = \begin{bmatrix} x_1^2 - x_k^2 + y_1^2 - y_k^2 + d_k^2 - d_1^2 \\ x_2^2 - x_k^2 + y_2^2 - y_k^2 + d_k^2 - d_2^2 \\ \vdots \\ x_{k-1}^2 - x_k^2 + y_{k-1}^2 - y_k^2 + d_k^2 - d_{k-1}^2 \end{bmatrix} \quad (b)$$

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} \quad (c)$$

The above equations can be converted into $\mathbf{Ax} = \mathbf{b}$. Due to measurement errors, the correct representation of the equation group is $\mathbf{Ax} = \mathbf{b} + \mathbf{\varepsilon}$. In order to obtain the optimal solution of the unknown node location, the sum of the squares of errors is used as the criterion, *i.e.*,

$$\begin{aligned} & \arg \min (\mathbf{b} - \mathbf{Ax})^T (\mathbf{b} - \mathbf{Ax}) \\ & \Rightarrow \arg \min (\mathbf{x}^T \mathbf{A}^T \mathbf{Ax} - 2\mathbf{b}^T \mathbf{Ax} + \mathbf{b}^T \mathbf{b}) \end{aligned} \quad (4)$$

Solve the gradient of equation (4), and the value is taken to be 0, so we can get,

$$\begin{aligned} & -2\mathbf{A}^T \mathbf{b} + 2\mathbf{A}^T \mathbf{Ax} = 0 \\ & \Rightarrow \mathbf{A}^T \mathbf{b} = \mathbf{A}^T \mathbf{Ax} \end{aligned} \quad (5)$$

If anchors are not in a line, when the square matrix $\mathbf{A}^T \mathbf{A}$ is reversible, it is easy to obtain the estimated location of unknown nodes.

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} \quad (6)$$

However, in the real environment, the situation is difficult to predict in advance. There may be one such instance:

There exist three anchors and an unknown node, in which Anchor 3 is a new anchor. Their locations are shown in table 1:

Table 1. Three anchors and an unknown node

coordinate	Anchor 1	Anchor 2	Anchor 3	unknown node
X	2	4	1	5
Y	6	12	5	7

Due to the influence of environment, there exists measurement error between the unknown nodes and the anchors, and then $\mathbf{b} = [12.6, 64.6]^T$. The estimated coordinate of unknown node is (5.9, 6.7) by equation (6). Anchor 3 is the new beacon node, since it is difficult to control error accumulation, the estimated coordinate is turned into (1, 3.1), and its real distance with the unknown node is unchanged, so $\mathbf{b} = [12.6, 64.6]^T$. At the moment, with OLS the estimated location of the unknown node is (-376, 134). This is due to the collinearity problem existing between the original anchors, new anchors or between the original anchors and the new anchors, resulting in large estimation errors. The collinearity problem between nodes is known as ill-posed problem in theory.

2.2. Improved localization algorithm

According to the error analysis of the previous section, in the process of incremental localization there are two kinds of influences, namely cumulative errors and collinearity among anchors. The cumulative error is caused by the accumulation of the preceding round of localization errors, so effectively eliminating the preceding round of localization errors can avoid the accumulation of errors. Inspired by this, the researchers [22] proposed that the weighted least squares method with the inverse of the variance of the errors as weight could eliminate the impact of the previous round of errors. There have been many discussions about the weight matrix $\mathbf{W}$, however, it is difficult to know the size of errors in the actual environment. We use the regularization iteratively reweighted least squares (RIRLS) method to construct the weight matrix guide by the residual error of the previous round of the estimated value. In each round of the locating, maximize the elimination of measurement error through iteratively reweighted least square (IRLS), and IRLS also has several capabilities [3, 23], such as weakening the outliers, network attack, data missing *etc.*, and then make the algorithm robust. In addition, in order to avoid the calculation problem of the location estimation which is caused by the collinearity among anchors, the regularization method is also used based on IRLS.

At the preliminary stage of locating, the unknown node utilizes the surrounding anchors to estimate location through the OLS method and these unknown nodes which can be located are upgraded to new anchors. The calculation method uses the formula (5).

In order to eliminate the effect of heteroscedasticity on the subsequent localization, the effect of abnormal data is reduced by weighting the sample of residuals. Therefore, the minimization is not for the sum of squared residuals, but for another residual function with low increasing rate.

$$\arg \min \rho(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) \quad (7)$$

Set the partial derivative of equation (7) to be 0.

$$-\sum_{i=1}^n \rho'(v_i) \hat{\mathbf{x}}' = 0 \quad (8)$$

Where ρ is the objective function. We define the weight function w as

$$w(\hat{v}) = \begin{cases} \rho'(\hat{v})/\hat{v} & \hat{v} \neq 0 \\ 1 & \hat{v} = 0 \end{cases} \quad (9)$$

Then formula (5) is turned into

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \mathbf{b} \quad (10)$$

Where $\mathbf{W} = \text{diag}(w_1, \dots, w_k)$.

The weighting matrix $\mathbf{W}$ is iteratively generated guide by the regression residuals, and the generation procedure is:

1. Set one iteration cursor as $\mathbf{I} = 0$, and in this case use OLS to estimate the estimated value $\hat{\mathbf{x}}^{(0)}$ of the initial coordinate.
2. From the initial OLS estimates, figure out the residual error $e_i^{(0)}$, which is used to calculate the initial weight.
3. Choose a weight function for the initial OLS residuals, producing the preparation weight $w(e_i^{(0)})$.

Thereinto, the common weight functions include Huber, Bisquare-weight, Andrew, Ramsay, and so on [2]. Their definitions are shown in table 2. In this paper weight function huber is selected.

Table 2. Several kinds of objective functions and influence functions

	objective function	influence function
Least Square	$\rho_{LS}(\mu) = \mu^2/2$	$\omega_{LS}(\mu) = 1$
Huber	$\rho_H(\mu) = \begin{cases} \mu^2/2 & \mu \leq c \\ c \mu - m^2/2 & \mu > c \end{cases}$	$\omega_H(\mu) = \begin{cases} 1 & \mu \leq c \\ c/ \mu & \mu > c \end{cases}$
Bisquare weight	$\rho_{BH}(\mu) = \begin{cases} \frac{c^2}{6} \left\{ 1 - \left[1 - (\mu/c)^2 \right]^3 \right\} & \mu \leq c \\ c^2/6 & \mu > c \end{cases}$	$\omega_{BH}(\mu) = \begin{cases} \left[1 - (\mu/c)^2 \right]^3 & \mu \leq c \\ 0 & \mu > c \end{cases}$
Andrew	$\rho_A(\mu) = \begin{cases} c \{ 1 - \cos(\mu/c) \} & \mu \leq c\pi \\ 2c & \mu > c\pi \end{cases}$	$\omega_A(\mu) = \begin{cases} \sin(\mu/c)/(\mu/c) & \mu \leq c\pi \\ 0 & \mu > c\pi \end{cases}$
Ramsay	$\rho_R(\mu) = \frac{1 - e^{-c \mu (1+c \mu)}}{c^2}$	$\omega_R(\mu) = e^{-c \mu }$

Where, c is a constant in the above table.

4. In the first iteration, set $\mathbf{I} = 1$. Minimize $\sum w_i e_i^2$ with the weighted least square method and get $\hat{\mathbf{x}}^{(1)}$, and the solution is $\hat{\mathbf{x}}^{(1)} = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \mathbf{b}$, where $\mathbf{W} = \text{diag}\{w_i^{I-1}\}$.
5. The program continues to use the residual error obtained by the initial WLS to calculate the new weight $w_i^{(2)}$.
6. The new weight $w_i^{(2)}$ will be used in the next WLS iteration. Set $\mathbf{I} = 2$, and estimate the new coordinate value $\hat{\mathbf{x}}^{(2)}$.
7. Repeat step 4 to step 6 until the coordinate value $\hat{\mathbf{x}}$ is stable.

In general, the iteration continues until $\hat{\mathbf{x}}^{(t)} - \hat{\mathbf{x}}^{(t-1)} \cong 0$. Generally speaking, when the change of estimated results is less than 0.01% [2] of the previous iteration, the solution is deemed to be convergent.

It's worth noting that the IRLS algorithm is iteratively forward, and every step of the derivation of optimal estimated value $\hat{\mathbf{x}}_i$ is based on the hypothesis that there exists no collinearity problem in $\mathbf{A}^T \mathbf{W} \mathbf{A}$, but unfortunately, the IRLS method can eliminate the heteroscedastic disturbances caused by accumulated errors, while the collinearity problem in $\mathbf{A}^T \mathbf{W} \mathbf{A}$ cannot be guaranteed to be eliminated. For the diagnosis of ill-posed problems, there are several frequently used methods, such as the characteristic analysis, the condition number and the condition index-variance decomposition proportion method and so on [10]. The most direct way to solve the ill-posed problem is to increase the amount of information, namely the way to increase the number of anchors, but this way is not suitable for incremental localization which is lack of anchors by itself. The use of the biased estimate method is another solution, which has strong anti-disturbance ability and can even overcome the ill-condition, so that the stability of parameter estimation can be enhanced. There are some common biased estimates; such as Stein estimate, ridge estimate, generalized ridge estimate, *etc.* [7]. Thereinto, the ridge estimate method is derived from the Tikhonov regularization principle, and it is the most influential algorithm in the biased estimates; what's more, its calculation amount is small, and it is simple and effective.

Therefore, in the iterative process it is necessary to do the corresponding strategy to avoid the problem that the algorithm cannot be solved, which is resulted from the collinearity. The ridge regression method is a regularization method which can effectively eliminate the collinearity problem. Proposed by A.E.Hoerl [1] in 1962, the method sacrifices the unbiasedness of estimation for greatly reduction of estimated variance by introducing an offset (also known as a ridge parameter), so as to ultimately improve the estimation accuracy and stability. For equation 4, after introducing the ridge parameter k , the new estimation model $\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{W} \mathbf{A} + k\mathbf{I})^{-1} \mathbf{A}^T \mathbf{W} \mathbf{b}$ is obtained. Compared with the OLS estimation, the ridge estimation transforms $\mathbf{A}^T \mathbf{W} \mathbf{A}$ into $\mathbf{A}^T \mathbf{W} \mathbf{A} + k\mathbf{I}$, and when the minimum eigenvalue of $\mathbf{A}^T \mathbf{W} \mathbf{A}$ is close to zero, the degree of the minimum eigenvalue of $\mathbf{A}^T \mathbf{W} \mathbf{A} + k\mathbf{I}$ closing to zero is improved. When the system is ill-posed, the ridge estimation can overcome the defect that the OLS estimation is unstable. The key of the ridge regression method is how to select the proper ridge parameter. There are some common selected methods, such as ridge trace method, L-curve method, generalized cross validation method, and so on [28]. Ridge trace method is intuitionistic, but has a certain degree of subjective randomness. Sometimes the general cross validation (GCV) function changes too slowly, and the minimum is difficult to determine. The L-curve method is a very rigorous method to determine the ridge parameters in theory with high accuracy and good applicability. However, the above methods all require a certain amount of computational cost. Considering that in general $\|\mathbf{A}^T \mathbf{W} \mathbf{A}\| < 0.01$ is an ill-conditioned matrix [4], so we also adopt the method of fixed ridge parameter.

3. Performance Evaluation

Incremental localization method has the characteristic of large-scale, and to verify such a localization algorithm may need to deploy hundreds of nodes, but there is no way to achieve such a large real network under the present experimental conditions. In addition, the evaluation of the merits of the incremental localization algorithm usually needs to verify its applicability in different scenarios; sometimes it is necessary to adjust the parameters of the algorithm in the same scene, all of which are more difficult to achieve under the present experimental conditions. Hence, in the research on large-scale incremental localization algorithm, simulations are usually used to evaluate the performance of the localization algorithm.

Subject to the measurement equipment and measurement methods, incremental localization process will be interfered by noise, so there is a certain deviation between the estimated location and the true location of node. The localization error is a function containing lots of network configuration parameters, including the number of anchors and nodes to-be-located, the geometric relationship among nodes, and the network connectivity *etc.*, combining to form a complex system. Understanding the error characteristic is a significant step how to find the control error. Currently there are some commonly used indexes to evaluate this localization error, such as Mean Square Error (*MSE*), Root Mean Square Error (*RMSE*), Cramer Rao lower bound (*CRLB*), *etc.* [8,17], in which *CRLB* can be used as a benchmark to measure the performance of localization algorithm. If the algorithm can be very close to this bound, then continuing to improve the calculation accuracy will only get a small profit. In addition, dependence of *CRLB* on network parameters can help to understand the characteristics of the network localization error, and *CRLB* can be computed efficiently, thereby avoiding the expensive Monte Carlo simulation. But *CRLB* is the lower bound for the covariance of the unbiased location estimate and it cannot be used as the performance boundary of a biased estimate. Regularization is a biased estimate for IL-RIRLS algorithm proposed in this paper. Therefore, it is not suitable for *CRLB* to evaluate the algorithm. *RMSE* is utilized to analyse and evaluate performance of IL-RIRLS and the ordinary incremental localization (OIL) algorithm.

Kleinrock [11] pointed out that in order to improve the utilization rate of network channel and maximize the network throughput, each node must has six neighbour nodes, while Takagi [20] thought that the performance of the static wireless network is the best when the number of neighbour nodes is 6 to 8. These results are all derived from the assumption that there is no interference between nodes. In order to ensure the best performance of wireless network in the actual deployment, a certain number of neighbouring nodes are required. To ensure the localization performance, each node needs to connect with a plurality of nodes. Therefore, in this experiment there is an assumption that nodes are distributed in a square area with 1000×1000, where there are 500 nodes and the anchor node rate is between 10% and 20%; nodes' RF transmission range is set to 80.

In the simulation experiment, the measurement of the distance between nodes is obtained by transforming the RSSI signal into the distance. For comparing the fairness of the experimental results, we use the signal model in the literature [14] to simulate the signal strength between nodes:

$$PL(d)[dB] = \overline{PL}(d) + X_\sigma = \overline{PL}(d_0) + 10n \log_{10}(d/d_0) + X_\sigma \quad (11)$$

where $d(m)$ is the distance between the receiver and the transmitter; $d_0(m)$ denotes the reference distance, whose general value is 1 m; $\overline{PL}(d)$ (dBm) is the received signal power at the receiver; $\overline{PL}(d_0)$ is the received signal power corresponding to the reference distance d_0 ; X_σ (dBm) denotes a Gaussian random variable with a mean value of 0 and reflects the change of the received signal power when the distance is certain; n represents the path loss index which is a value associated with the environment. The values of $\overline{PL}(d_0)$ and n can refer to the actual measured values in the literature [14].

Because of the higher coverage rate of the incremental algorithm, the experiment in this section is mainly to study the localization accuracy of nodes, and $RMSE$ is used as the evaluation basis for localization accuracy. Its definition is as follows:

$$RMSE = \sqrt{\frac{1}{N_t} \sum_{i=1}^{N_t} ((\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2)} \quad (12)$$

Where $(\hat{x}_i, \hat{y}_i)$ is the estimated coordinate location of the i -th node; (x_i, y_i) is the real coordinate of the i -th node; N_t denotes the number of nodes that can be located.

Under the real environment, there are a large number of non-line-of-sight(NLOS) problems, such as moving crowd, occlusion of large items, *etc.* Therefore, the experiment investigates two types of network topology, which is the random deployment of nodes without occlusion in the square area, and the random deployment of nodes with occlusion, the deployment with the shape of 'C'. Figure 2 shows the final estimation results of the two groups are examined. In the figure, the solid '★' denotes an anchor node; '*' represents the real location of the unknown node; '○' only uses the original anchor node to obtain the estimated location. '◇' expresses the obtained estimated location by using the new anchor and the original anchor. Each straight line connects a true node location and its estimation. The length of each line denotes the estimation error. In this group of simulation, 60 original anchors are placed randomly in the area.

(a) OIL with random deployment under square topology

(b) IL-RIRLS with random deployment under square topology

Figure 2. Localization results with two different algorithms under different topologies

Figure 2(a) shows the localization result of the OIL with nodes randomly deployed in the no-occluded area and its localization $RMSE=6.86$. While considering the cumulative error and the collinearity between anchors, from Figure 2 (b), it is easy to see that under the condition of the corresponding deployment, the localization result of the proposed IL-RIRLS method significantly outperforms that of the OIL, by this time $RMSE=5.28$. Figure 2 (c) shows the localization result of the OIL with nodes randomly deployed in the C-shape area. Due to without considering the cumulative error and the collinearity between anchor nodes, the estimated location is seriously deviated from the true location, and $RMSE=7.31$ in this region. Figure 2(d) shows the localization result under the conditions of the corresponding deployment. It is easy to see that it is obviously better than the OIL, and its $RMSE=6.26$.

In order to avoid the case that a single experiment cannot reflect performance of the algorithm under each area deployment, we have carried out a number of redeployment experiments (carried out 200 trials) on two kinds of localization algorithms. All of the reported simulation results are the average over 200 trials. Figure 3 shows the histogram of the average $RMSE$ of multiple experiments with the change of the number of anchor nodes (the number of anchor nodes increased from 10% to 20%). Among them, Figure 3(a) expresses the histogram of the change of localization error in the no-occluded area with nodes randomly deployed. Figure 3(b) represents the histogram of the change of localization error in the C-shape area with nodes randomly deployed.

Due to the deployment randomness, there exists a collinearity problem between the original anchors, and coupled with the cumulative error problem, the situation becomes more complicated. The OIL does not take the above problems into consideration, so $RMSE$ fluctuates widely. In the experiment, for random deployment in the no-occluded area, the value of $RMSE$ is between 6.63 to 9.26, while under the occluded situation, the value of $RMSE$ is between 4.28 to 9.20. In Figure 3 (a), when the number of anchor nodes is 80, the value of $RMSE$ is considerably greater than the value when the number of anchor nodes is 70. In Figure 3 (b), when the number of anchor nodes is 90, the value of $RMSE$ is significantly greater than the value when the number of anchor nodes is 80. This is caused by not considering that the probability of being collinear may increase as the number of the anchors is much more in the case of collineation. The method proposed in this paper has taken into account the two problems mentioned above, so the value of $RMSE$ decreases with the increase of the number of anchor nodes, and the corresponding $RMSE$ values range from 4.87 to 8.07,

and from 3.5 to 7.68, respectively. Because the number of nodes in the unit area in the C-shape area is greater than that in the no-occluded area, the *RMSE* value of the corresponding area is smaller than that of the no-occluded area.

Figure 3. Simulation results of two algorithms with different number anchors under two network topologies

4. Conclusion

Combining the regularized iteratively reweighted least square with the incremental localization, we use the iteratively reweighted method in localization process to guide weight selection so as to reduce the heteroscedasticity problem caused by cumulative errors; meanwhile the regularization method in calculation is used to avoid the collinearity problem between the original anchors, the new anchors or between the original anchor and the new anchor. And in the localization process, we use the experience ridge parameter that it can avoid the complex process of parameter selection and make the proposed method feasible, so the effect of localization is more stable and the accuracy of localization is higher. The simulations proved that the effect and applicability are better than the previous incremental localization algorithms.

Acknowledgement

The paper is sponsored by the NSF of China (61272379, 61403080), National Natural Science Foundation of Jiangsu Province (BK20140641), China Postdoctoral Science Foundation (2015M571633), Jiangsu Postdoctoral Science Foundation (1401016B), Natural Science Foundation of the Jiangsu Higher Education Institutions of China (15KJB520009), Project of Modern Educational Technology of Jiangsu Province (2015-R-42440), Jiangsu Province Undergraduate Training Programs for Innovation and Entrepreneurship (201513573005Z), Doctoral Scientific Research Start-up Foundation of Jinling Institute of Technology (JIT-B-

201411) and the Open Project Program of Jiangsu Key Laboratory of Remote Measurement and Control (YCCCK201603).

References

- [1] Ae H., Application of ridge analysis to regression problems, *Chemical Engineering Progress*, 58, 1962, 54-59.
- [2] Andersen R. *Modern Methods for Robust Regression*, SAGE Publications, New York, 2008.
- [3] Chen H, Lou W, Wang Z, et al., *Securing DV-Hop localization against wormhole attacks in wireless sensor networks*, *Pervasive and Mobile Computing*, **16**, 2015, 22-35.
- [4] Chen J, Chen X., *Special matrix*, Tsinghua University press, Beijing , 2001.
- [5] Cribari-Neto F, Silva W., *A New Heteroskedasticity consistent Covariance Matrix Estimator for the Linear Regression Model*, *Advances in Statistical Analysis*, **95**, 2, 2011, 129-146.
- [6] Doherty L, Pister K S J, El Ghaoui L., *Convex position estimation in wireless sensor networks*, INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, 2001.
- [7] Gruber M H J., *Regression Estimators: A Comparative Study*, Johns Hopkins University Press, Baltimore, 2010.
- [8] Huang B, Yu C, Anderson B D O., *Understanding Error Propagation in Multihop Sensor Network Localization*, *IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS*, 60, **12**, 2013, 5811-5819.
- [9] Jiaming S, Zhaowei T, Luoyi F, et al., Report on CCF A-level Conferences: Trend Analysis of Top Computer Network Conference, *Communications of The CCF*, **11**, 9, 2015, 62-66.
- [10] Kabanikhin S I., *Inverse and Ill-posed Problems: Theory and Applications*, DE GRUYTER, 2011.
- [11] Kleinrock L, Silvester J., *Optimum Transmission Radii for Packet Radio Networks or Why Six Is A Magic Number*, Proceedings of the IEEE National Telecommunications Conference, 1978, 4.3.1-4.3.5.
- [12] Leick A, Rapoport L, Tatarnikov D., *GPS Satellite Surveying*, JOHN WILEY & SONS, Canada, 2015.
- [13] Lim C, Sen P K, Peddada S D., *Accounting for uncertainty in heteroscedasticity in nonlinear regression*, *Journal of Statistical Planning and Inference*, **142**, 5, 2012, 1047-1062.
- [14] Neal Patwari, *Wireless Sensor Network Localization Measurement Repository*, http://span.ece.utah.edu/download/patwari03meas_sIV_v2.mat.
- [15] Nguyen C, Georgiou O, Doi Y., *Maximum likelihood based multihop localization in wireless sensor networks*, *Communications (ICC)*, 2015 IEEE International Conference on, 2015: 6663-6668.
- [16] Rawat P, Singh K D, Chaouchi H, et al., *Wireless sensor networks: a survey on recent developments and potential synergies*, *The Journal of Supercomputing*, **68**, 1, 2014, 1-48.

- [17] Savvides A, Garber W L, Moses R L, et al., *An Analysis of Error Inducing Parameters in Multihop Sensor Node Localization*, IEEE TRANSACTIONS ON MOBILE COMPUTING, **4**, 6, 2005, 567-577.
- [18] Shang Y, Rumi W, Zhang Y, et al., *Localization from connectivity in sensor networks*, IEEE Transactions on Parallel and Distributed Systems, **15**, 11, 2004, 961-974.
- [19] Simonetto A, Leus G., *Distributed Maximum Likelihood Sensor Network Localization*, IEEE Transactions on Signal Processing, **62**, 6, 2014, 1424-1437.
- [20] Takagi H, Kleinrock L., *Optimal transmission ranges for randomly distributed packet radio terminals*, Communications, IEEE Transactions on, **32**, 3, 1984, 246-257.
- [21] Wang C, Chen J, Sun Y, et al., *A graph embedding method for wireless sensor networks localization*, Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE, 2009.
- [22] Wu S, Xu D, Liu S., *Weighted Linear Least Square Localization Algorithms for Received Signal Strength*, Wireless Personal Communications, **72**, 1, 2013, 747-757.
- [23] Xiao Q, Bu K, Wang Z, et al., *Robust Localization Against Outliers in Wireless Sensor Networks*, Transactions on Sensor Networks (TOSN), **9**, 2, 2013, 1-26.
- [24] Yan X, Qian H, Chen J., *Incremental Localization Algorithm Based on Multivariate Analysis*, International Journal of Distributed Sensor Networks, **2013**, 2013, 1-13.
- [25] Yan X, Song A, Yang Z, et al., *An improved multihop-based localization algorithm for wireless sensor network using learning approach*, Computers & Electrical Engineering, **48**, 2015, 2015, 247-257.
- [26] Yang Z, Wu C, Liu Y., *Location-based Computing: Localization and Localizability of Wireless Networks*, Tsinghua University press, Beijing, 2014.
- [27] Yu K, Sharp I, Guo Y J., *Ground-Based Wireless Positioning*, John Wiley and Sons, Great Britain, 2009.
- [28] Zhang L, Hua C, Tang Y, et al., *Ill-posed Echo State Network based on L-curve Method for Prediction of Blast Furnace Gas Flow*, Neural Processing Letters, **43**, 1, 2016, 97-113.

This is an extended version of the paper presented at the International Conference on Big Data Intelligence and Computing (DataCom 2015), Chengdu, China, December 19-21, 2015