

Robust Networked Federated Learning for Localization

Reza Mirzaeifard*, Naveen K. D. Venkategowda[§], Stefan Werner*

*Dept. of Electronic Systems, Norwegian University of Science and Technology-NTNU, Norway

[§]Department of Science and Technology, Linköping University, Sweden

E-mails: {reza.mirzaeifard, stefan.werner}@ntnu.no, naveen.venkategowda@liu.se

Abstract—This paper addresses the problem of localization, which is inherently non-convex and non-smooth in a federated setting where the data is distributed across a multitude of devices. Due to the decentralized nature of federated environments, distributed learning becomes essential for scalability and adaptability. Moreover, these environments are often plagued by outlier data, which presents substantial challenges to conventional methods, particularly in maintaining estimation accuracy and ensuring algorithm convergence. To mitigate these challenges, we propose a method that adopts an L_1 -norm robust formulation within a distributed sub-gradient framework, explicitly designed to handle these obstacles. Our approach addresses the problem in its original form, without resorting to iterative simplifications or approximations, resulting in enhanced computational efficiency and improved estimation accuracy. We demonstrate that our method converges to a stationary point, highlighting its effectiveness and reliability. Through numerical simulations, we confirm the superior performance of our approach, notably in outlier-rich environments, which surpasses existing state-of-the-art localization methods.

Index Terms—Federated learning, Robust learning, distributed learning, localization, non-convex and non-smooth optimization

I. INTRODUCTION

The increasing prevalence of edge devices in the era of the internet-of-things (IoT) and cyber-physical systems has resulted in a significant upsurge in available data. These devices, functioning within a federated network, present unique challenges. In particular, traditional machine learning methodologies often require a central server for data processing, raising privacy concerns [1]. In response to these concerns, federated learning (FL) has emerged as an innovative solution [2]–[4] for edge devices to collectively train a global model using their locally stored data, eliminating the need for direct data sharing [2], [4]. However, federated settings are frequently besieged by outlier data, often exhibiting heavy-tailed distributions [3], [5]. These outliers can considerably distort the accuracy of learning outcomes. Nevertheless, developing robust and computationally efficient solutions for managing outlier data in FL remains a critical research question, which this paper aims to address in the context of localization.

Localization is the process of estimating the location of an object or an event within a given environment. This problem has given rise to a broad range of applications in vehicle localization [6], aviation [7], healthcare [8], environmental [9], and industrial fields [10]. Based on the type of measurements

used to estimate position, the localization problem has different formulations which include time-of-arrival (ToA) [11], received-signal-strength (RSS) [12], time-difference-of-arrival (TDoA) [13], angle-of-arrival (AoA) [14], and frequency-difference-of-arrival (FDoA) [15]. In IoT applications, ToA and RSS are particularly effective as they provide relatively simple and cost-effective ways to estimate distances using the existing communication infrastructure. Therefore, in this paper, we focus on distance-based localization that inherently encompasses ToA and RSS techniques, as they can be interpreted as distance measurements [16]. This approach finds relevance across a spectrum of practical scenarios and continues to be an area of research interest.

Conventionally, various methods have been proposed for localization problems, including the parallel projection method [17], projection-onto-convex-sets method [18], nearest local minimum [19], boundary-of-convex-sets [20], recursive weighted least-squares algorithms [21], augmented Lagrangian-based method for localization [16], and iterative re-weighted least-squares [22]. However, these methods have limitations, such as being centralized or sequentially updating estimates, which hinders parallelization and flexibility. More recently, a parallel distributed alternating projection algorithm (DAPA) has been proposed for the distributed case, formulating the localization problem as a ring intersection problem [23]. Despite the advancements of DAPA, challenges remain, particularly in handling three-dimensional localization problems and targets lying outside the sensor’s convex hull [24]. To address these limitations, an alternative approach called EL-ADMM has emerged. EL-ADMM directly solves the non-convex and non-smooth event localization problem using the Alternating Direction Method of Multipliers [24]. By bypassing the need for high-computation convex relaxation techniques, EL-ADMM shows promise in handling three-dimensional environments [24]. However, EL-ADMM still faces challenges in outlier handling and lacks comprehensive convergence proof. Given these constraints, there is a compelling research imperative to develop a robust, distributed algorithm capable of efficiently handling three-dimensional localization problems with outliers.

In this paper, we address a challenging class of robust localization problems that require tackling non-convex and non-smooth optimization within a federated setting. The decentralization, energy constraints, and the presence of outliers further

exacerbate the difficulty of accurately estimating locations in such settings. To overcome these challenges, we propose a novel distributed sub-gradient-based algorithm specifically designed to address this problem. Our contributions can be summarized as follows:

- **Efficient Optimization Algorithm:** We introduce a novel optimization algorithm that operates within a single loop framework. This algorithm directly solves the localization problem in its original form, utilizing simple updating steps. The use of this approach leads to increased accuracy and eliminates the need for iterative approximation processes.
- **Theoretical Analysis:** We provide comprehensive theoretical insights into the convergence behavior of our algorithm. Through rigorous mathematical analysis, we elucidate the conditions under which our method is guaranteed to converge, thereby offering deeper insights into its reliability and robustness.
- **Empirical Validation:** We substantiate our theoretical results through extensive numerical simulations. Notably, our algorithm exhibits exceptional resilience in the presence of heavy-tailed noise and outliers. Furthermore, a comparative study with the state-of-the-art EL-ADMM algorithm highlights the superior accuracy and comparable fast convergence rate of our approach.

Mathematical Notations: Scalars are denoted by lowercase letters, column vectors by bold lowercase letters, and matrices by bold uppercase letters. The transpose of a matrix is signified by $(\cdot)^T$. The j th column of a matrix \mathbf{A} is represented as \mathbf{a}_j . The element in the i th row and j th column of \mathbf{A} is represented as a_{ij} . The sub-gradient of a function $f(\cdot)$ at a given point u is signified by $\partial f(u)$.

II. PRELIMINARIES

We consider the problem of localization with L sensors, where the location of the i -th sensor is denoted by \mathbf{a}_i , for $i = 1, 2, \dots, L$, and $\mathbf{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_L\}$ represents the complete set of sensors' locations. The unknown source is located at $\mathbf{x} \in \mathbb{R}^n$, which must be estimated using the range information between the sensors and the source. Let $d_i > 0$, $i = 1, 2, \dots, L$, represents a noisy observation representing the range between source location \mathbf{x} and i -th sensor location \mathbf{a}_i . The range d_i can be expressed as

$$d_i = \|\mathbf{x} - \mathbf{a}_i\| + \epsilon_i, \quad i = 1, 2, \dots, L, \quad (1)$$

where ϵ_i denotes the noise associated with each range observation.

Additionally, the network of L sensors is modeled as an unweighted undirected graph \mathcal{G} consisting of vertices $\mathcal{V} = 1, \dots, L$ and bidirectional communication links represented by the edge set \mathcal{E} . Each sensor $i \in \mathcal{V}$ can communicate with those in its neighborhood \mathcal{N}_i of cardinality $|\mathcal{N}_i|$. We assume that graph \mathcal{G} is strongly connected.

The maximum-likelihood estimate for the source location \mathbf{x} with additive independent and identically distributed Gaussian

noise affecting range measurements can be determined as the solution of the optimization problem [16]:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{1}{L} \sum_{i=1}^L (\|\mathbf{x} - \mathbf{a}_i\| - d_i)^2. \quad (2)$$

However, this approach encounters challenges when faced with outlier measurements that do not conform to a Gaussian distribution. These outliers can drastically bias location estimate due to the magnified impact of their squares in each term. One could use conventional filtering techniques to reduce this effect. However, such filtering methods often presuppose knowledge about the noise characteristics, which is not always available or accurate. Additionally, filtering can induce a 'masking effect', in which crucial signal components, particularly those smaller ones sharing frequency characteristics with the noise, are suppressed or obscured [22]. Such an outcome is undesirable in our context as it may lead to the loss of vital information [22]. To counteract this, robust estimation theory offers a way of softly rejecting outliers by using the L_1 loss [25] as given by

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{1}{L} \sum_{i=1}^L \|\|\mathbf{x} - \mathbf{a}_i\| - d_i\| = \arg \min_{\mathbf{x}} \sum_{i=1}^L f_i(\mathbf{x}), \quad (3)$$

where the local objective function $f_i(\mathbf{x})$ at sensor i is defined as $f_i(\mathbf{x}) = \frac{1}{L} \|\|\mathbf{x} - \mathbf{a}_i\| - d_i\|$. The L_1 objective function diminishes the impact of outliers by minimizing the average absolute discrepancy between the Euclidean distance from \mathbf{x} to each sensor \mathbf{a}_i and the corresponding observed range d_i . This approach is particularly beneficial as it ensures that smaller yet crucial variations in the data are preserved, and not overshadowed by outliers.

Even though the L_1 robust formulation can enhance accuracy and handle outliers more efficiently, it is inherently non-smooth and non-convex and not even a difference of convex functions, which makes (3) difficult to solve. Traditionally, convex relaxation approaches have been considered to tackle this problem [22], [25]. These methods, however, can compromise accuracy and are often time-consuming and computation-intensive iterative processes. In the distributed case, the problem exacerbates as one must ensure consensus among the nodes as well as convergence to a solution. To that end, this paper proposes a method for addressing the problem directly with the L_1 loss in a distributed manner.

III. DISTRIBUTED ROBUST LOCALIZATION

The proposed algorithm for localization in a distributed setting called distributed sub-gradient method for robust localization (DSRL), updates each node's estimate of the actual parameter \mathbf{x}^* through a diffusion step and a sub-gradient update step. In the diffusion step, each node uses a diffusion variable, \mathbf{v}_i , to incorporate information from its neighboring nodes. At each time instant k , the diffusion step takes place as follows:

$$\mathbf{v}_i^{(k)} = \mathbf{x}_i^{(k)} + \alpha^{(k)} \sum_{j \in \mathcal{N}_i} (\mathbf{x}_j^{(k)} - \mathbf{x}_i^{(k)}). \quad (4)$$

Algorithm 1: Distributed Sub-gradient Method for Robust Localization (DSRL)

Initialize $\mathbf{x}_i^{(1)}$ for each node i , the weight parameter $\{\alpha_k\}_{k=1}^K$, the step-size $\{\beta_k\}_{k=1}^K$ and the number of iterations K ;

for $k = 1, \dots, K$ **do**

for $i = 1, \dots, L$ **do**

Receive \mathbf{x}_j from neighbors in \mathcal{N}_i ;

Update $\mathbf{v}_i^{(k)}$ by (4);

Update $\mathbf{g}_i^{(k)}$ by (6);

Update $\mathbf{x}_i^{(k+1)}$ by (5);

end

end

In this formula, $\mathbf{x}_i^{(k)}$ is the current estimate at node i , while $\mathbf{x}_j^{(k)}$ are the estimates at each neighboring node j . The term $\alpha^{(k)}$ is a scalar weight parameter that adjusts how strongly the estimates of the neighbors influence the node's update. The diffusion step seeks to decrease the discrepancy between each node's estimate and the average estimate of its neighbors, working as a type of approximation to the average consensus algorithm. Instead of a simple average, we apply a weighted adjustment based on the difference between a node's estimate and its neighbors' estimates, thereby aiding the nodes to reach a consensus on the parameter estimate.

In the subsequent step, termed the sub-gradient step, each node updates its estimate of the true parameter \mathbf{x}^* in the manner depicted below:

$$\mathbf{x}_i^{(k+1)} = \mathbf{v}_i^{(k)} - \beta^{(k)} \mathbf{g}_i^{(k)}, \quad (5)$$

where $\beta^{(k)}$ is the step-size at iteration k and $\mathbf{g}_i^{(k)} \in \partial f_i(\mathbf{x}_i^{(k)})$ is any element of the sub-differential set of $f_i(\cdot)$. Each function $f_i(\cdot)$ is neither convex nor smooth, so the distributed sub-gradient method proposed in [26] is adopted for iterative computation of the estimate of \mathbf{x} at each node. The sub-gradient of the local robust localization function $f_i(\cdot)$ with respect to the coefficient $\mathbf{x}_i^{(k)}$ is computed as follows:

$$\partial f_i(\mathbf{x}) = \begin{cases} \frac{\mathbf{x} - \mathbf{a}_i}{L \|\mathbf{x} - \mathbf{a}_i\|} \text{sign}(\|\mathbf{x} - \mathbf{a}_i\| - d_i), & \|\mathbf{x} - \mathbf{a}_i\| > 0 \\ \mathbf{0}, & \|\mathbf{x} - \mathbf{a}_i\| = 0 \end{cases} \quad (6)$$

Algorithm 1 summarizes the proposed method for solving distributed robust localization.

IV. CONVERGENCE PROOF

In this section, a convergence analysis of the Algorithm 1 is presented. We construct our convergence proof based on the following assumptions.

Assumption 1. Consider the function $f(\cdot) = \sum_{i=1}^L f_i(\cdot)$ and let $CP_f \subset \mathbb{R}^n$ represent the set of its critical points. Then, the complement of CP_f in \mathbb{R}^n , expressed as $\mathbb{R}^n \setminus CP_f$, is a dense set in \mathbb{R}^n .

Assumption 2. The weight parameter $\alpha^{(k)}$ and the step-size $\beta^{(k)}$ are derived from $\alpha^{(k)} = \frac{a}{k^{\tau_\alpha}}$ and $\beta^{(k)} = \frac{b}{k^{\tau_\beta}}$ respectively. Here, $0 < \tau_\alpha < \tau_\beta$, $\frac{1}{2} < \tau_\beta \leq 1$, and both a and b are positive constants.

While Assumption 1 might appear to be technical, it is in fact a relatively mild condition that guides the positioning and quantity of sensors to ensure convergence to the critical points in our algorithm. As an illustration, this assumption is invariably satisfied when employing an odd number of sensors. Having a non-critical dense set ensures that, through communication and updates, the sensors have enough room to maneuver and collaborate towards converging to a consensus that aligns with the global objective. On the other hand, Assumption 2 guarantees that information can be effectively disseminated across the network.

In order to show the regularity of the objective function we provide the following Lemmas.

Lemma 1. each $f_i(\cdot)$ is locally Lipschitz continuous.

Proof. Given that the sub-gradient of each $f_i(\cdot)$ is bounded, it follows that these functions are locally Lipschitz continuous. \square

Lemma 2. There exists a radius $R > 0$ and constants $C_1, C_2 > 0$ such that:

$$\left\langle \frac{\mathbf{g}}{\|\mathbf{g}\|}, \frac{\mathbf{x}}{\|\mathbf{x}\|} \right\rangle \geq C_1 \quad \text{and} \quad \|\mathbf{g}\| \leq C_2 \|\mathbf{x}\|,$$

for all $\mathbf{g} \in \partial f_n(\mathbf{x})$ and $\|\mathbf{x}\| \geq R$.

Proof. We select a radius $R = (3 \max_i \max_j |a_{ij}| + \max_i d_i) + 1$. Consequently, $\text{sign}(\|\mathbf{x} - \mathbf{a}_i\| - d_i) = 1$. We proceed to examine the first condition of the lemma. It is evident that

$$\begin{aligned} \left\langle \frac{\mathbf{g}}{\|\mathbf{g}\|}, \frac{\mathbf{x}}{\|\mathbf{x}\|} \right\rangle &= \left\langle \frac{\mathbf{x} - \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\|}, \frac{\mathbf{x}}{\|\mathbf{x}\|} \right\rangle \\ &\geq \frac{\|\mathbf{x}\|}{\|\mathbf{x} - \mathbf{a}_i\|} - \frac{\mathbf{x}^T \mathbf{a}_i}{\|\mathbf{x} - \mathbf{a}_i\| \|\mathbf{x}\|} \\ &\geq 1 - \frac{\|\mathbf{a}_i\|}{\|\mathbf{x} - \mathbf{a}_i\|} \geq 1 - \frac{\|\mathbf{a}_i\|}{2\|\mathbf{a}_i\|} = \frac{1}{2}. \end{aligned}$$

These inequalities are derived from the conditions that $\|\mathbf{x}\|$ always surpasses $\|\mathbf{x} - \mathbf{a}_i\|$, and that $\|\mathbf{x}\|$ always exceeds $3\|\mathbf{a}_i\|$. These conditions are a direct result of our choice of R . This effectively verifies that $\left\langle \frac{\mathbf{g}}{\|\mathbf{g}\|}, \frac{\mathbf{x}}{\|\mathbf{x}\|} \right\rangle \geq C_1$ for $C_1 = \frac{1}{2}$. For the second condition of the lemma, we are aware that $\|\mathbf{g}\| = 1$ when $\|\mathbf{x} - \mathbf{a}_i\| > d_i$. This is always true given our selection of R . Also, taking into account that $1 \leq \|\mathbf{x}\|$, the inequality $\|\mathbf{g}\| \leq C_2 \|\mathbf{x}\|$ holds for $C_2 = 1$. \square

Theorem 1. Given that the static graph \mathcal{G} is strongly connected, the objective function fulfills Assumption 1, and both the step-size and weight parameter adhere to Assumption 2, the DSRL algorithm is ensured to converge to a critical point.

Proof. The convergence of our proposed DSRL algorithm is substantiated by verifying that all the prerequisites set forth in

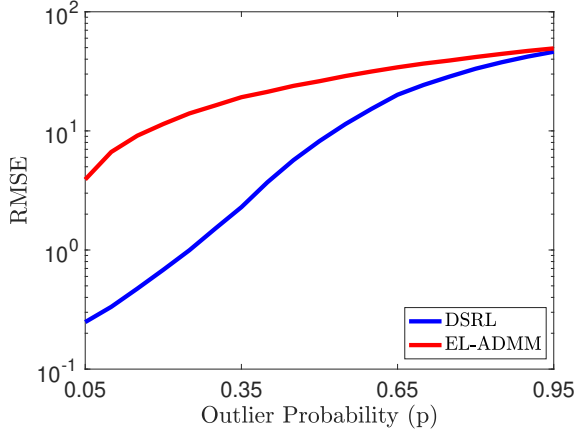


Fig. 1: Steady-state RMSE versus the outlier probability.

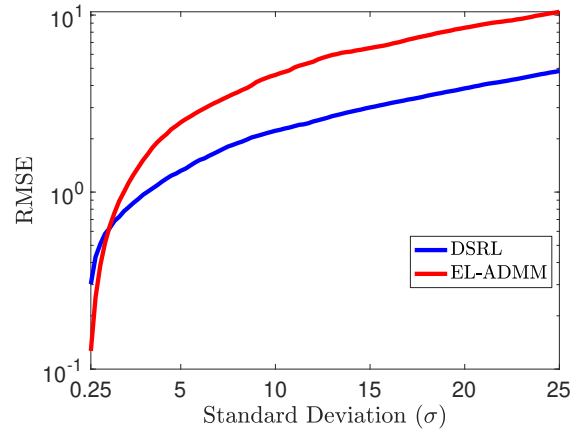


Fig. 2: Steady-state RMSE versus the standard deviation of mixture of Laplacian noise.

[26, Theorem 7], inclusive of the regularity conditions of our robust localization function, are met. The fulfillment of these conditions is demonstrated in Lemma 1 and Lemma 2. \square

V. SIMULATION RESULTS

In this section, we assess the performance of the proposed DSRL algorithm via simulations and draw comparisons with the distributed event localization via the Alternating Direction Method of Multipliers (EL-ADMM) algorithm, as outlined in [24]. The sensor network \mathcal{G} under consideration comprises $L = 31$ nodes, which are uniformly and randomly distributed within a $[-30, 30]^3$ cubic region, and is fully connected. Connections are established between nodes if their distance is less than 25, ensuring a minimum of 2 and a maximum of 10 neighbors per node. In addition, the target is assumed to be uniformly and randomly distributed within the same cubic region. For each iteration, the weight parameter $\alpha^{(k)}$ and the step-size $\beta^{(k)}$ are set to $\frac{0.3}{k^{0.55}}$ and $\frac{3.5}{k^{0.5}}$, respectively. We employ the root-mean-square error (RMSE), denoted by $\sqrt{\frac{\sum_{i=1}^L \|\hat{\mathbf{w}}_i - \mathbf{x}\|_2^2}{L}}$, as the performance metric. The results are obtained by averaging the outcomes of 1000 independent trials.

In the first scenario, we focus on comparing the accuracy of the algorithms when handling outliers in the measurements. We introduce outliers at varying probability, denoted as p , ranging from 0.05 to 0.95 with increments of 0.05. The outliers, represented as ϵ_i , are drawn from a uniform distribution over the interval $[0, 60\sqrt{3}]$. This is formalized as:

$$y_i = \begin{cases} \epsilon_i, & \text{if } u_i < p, \\ \|\mathbf{x} - \mathbf{a}_i\|, & \text{otherwise,} \end{cases} \quad (7)$$

where u_i is a realization of the uniform distribution on the interval $[0, 1]$. As illustrated in Figure 1, the DSRL algorithm outperforms EL-ADMM consistently across the various frequency of outliers.

In the second scenario, we evaluate the performance of the algorithms under noisy conditions, where the noise follows

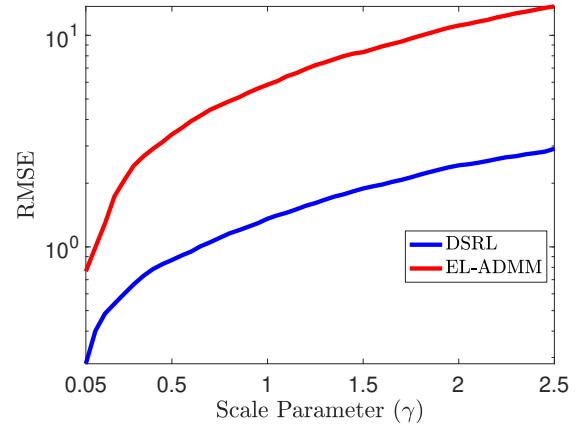


Fig. 3: Steady-state RMSE versus scale parameter of Cauchy noise.

a mixture of Laplacian distributions. The measurements are generated as follows:

$$y_i = \|\mathbf{x} - \mathbf{a}_i\| + \epsilon_i, \quad \forall i \in 1, \dots, N, \quad (8)$$

where $\{\epsilon_i\}_{i=1}^N$ denotes the components of the noise, which are independently and identically sampled from a mixture of Laplacian distributions given by $\sum_{i=1}^2 c_i \frac{\lambda_i}{2} e^{-\lambda_i |v|}$. Here, $c_1 = 0.9$, $c_2 = 0.1$, and $\lambda_2 = \frac{\lambda_1}{10}$. The parameter λ_1 is related to the standard deviation (σ) and can be calculated as $\lambda_1 = \sqrt{\frac{\sigma^2}{10.9}}$. The value of 10.9 in the denominator arises from the specific weights c_1 and c_2 , and it represents the effective contribution of the two components of the Laplacian mixture in relation to the variance. In this setup, 10% of the measurements can be considered outliers. We obtain simulation results for varying standard deviation values, ranging from 0.25 to 25, with increments of 0.25. As illustrated in Figure 2, the DSRL algorithm exhibits superior performance in terms of RMSE, particularly for standard deviation values greater than 1.

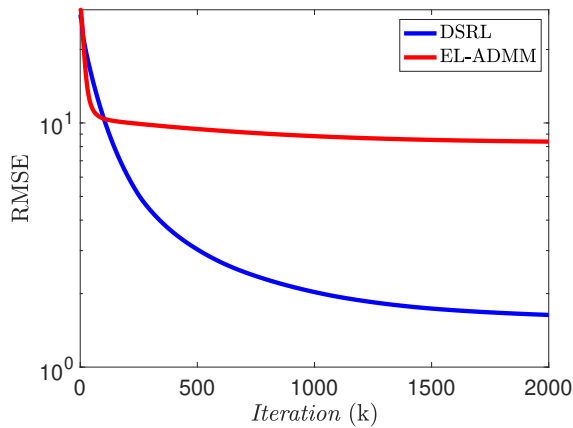


Fig. 4: RMSE versus iterations.

In the third scenario, we assess the accuracy of the algorithms in the presence of Cauchy noise. The measurements are formulated as in (8), where each component ϵ_i is independently and identically distributed according to the Cauchy distribution with the scale parameter γ , i.e., $\epsilon_i \sim \text{Cauchy}(0, \gamma)$. We conduct simulations for varying scale parameters, ranging from 0.05 to 2.5, in increments of 0.05. Figure 3 illustrates that the DSRL algorithm outperforms EL-ADMM in terms of RMSE across different scale parameters. Additionally, for the scale parameter set at 1, Figure 4 demonstrates that DSRL maintains a competitive convergence speed compared to EL-ADMM.

VI. CONCLUSION

This paper presents a sub-gradient algorithm devised for the distributed robust localization problem that directly tackles non-convex and non-smooth objective functions. This proposed algorithm has been rigorously proven to be convergent. Our simulation results highlight the superiority of our algorithm in terms of root-mean-squared error, compared to the existing state-of-the-art EL-ADMM algorithm. Particularly in outlier settings, the algorithm's effectiveness was consistently apparent.

REFERENCES

- [1] X. Yin, Y. Zhu, and J. Hu, "A comprehensive survey of privacy-preserving federated learning: A taxonomy, review, and future directions," *ACM Computing Surveys*, vol. 54, no. 6, pp. 1–36, July 2021.
- [2] X. Zhou, W. Liang, I. Kevin, K. Wang, Z. Yan, L. T. Yang, W. Wei, J. Ma, and Q. Jin, "Decentralized p2p federated learning for privacy-preserving and resilient mobile robotic systems," *IEEE Wireless Communications*, vol. 30, no. 2, pp. 82–89, Apr. 2023.
- [3] J. Zhao, X. Chang, Y. Feng, C. H. Liu, and N. Liu, "Participant selection for federated learning with heterogeneous data in intelligent transport system," *IEEE Transactions on Intelligent Transportation Systems*, Feb. 2022.
- [4] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE signal processing magazine*, vol. 37, no. 3, pp. 50–60, May 2020.
- [5] V. Tsouvalas, A. Saeed, T. Ozcelebi, and N. Meratnia, "Federated learning with noisy labels," *arXiv preprint arXiv:2208.09378*, Aug. 2022.
- [6] O. Kaiwartya, Y. Cao, J. Lloret, S. Kumar, N. Aslam, R. Kharel, A. H. Abdullah, and R. R. Shah, "Geometry-based localization for gps outage in vehicular cyber physical systems," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 3800–3812, Jan. 2018.
- [7] M. Strohmeier, I. Martinovic, and V. Lenders, "A k-nn-based localization approach for crowdsourced air traffic communication networks," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 3, pp. 1519–1529, Jan. 2018.
- [8] S.-B. Han, J.-H. Kim, and H. Myung, "Landmark-based particle localization algorithm for mobile robots with a fish-eye vision system," *IEEE/ASME Transactions on Mechatronics*, vol. 18, no. 6, pp. 1745–1756, Aug. 2012.
- [9] Z. Zhou, Z. Peng, J.-H. Cui, Z. Shi, and A. Bagtzoglou, "Scalable localization with mobility prediction for underwater sensor networks," *IEEE Transactions on Mobile Computing*, vol. 10, no. 3, pp. 335–348, Sep. 2010.
- [10] S. Petersen, P. Doyle, S. Vatland, C. S. Aasland, T. M. Andersen, and D. Sjong, "Requirements, drivers and analysis of wireless sensor network solutions for the oil & gas industry," in *2007 IEEE Conference on Emerging Technologies and Factory Automation*. IEEE, Sep. 2007, pp. 219–226.
- [11] Y.-M. Pun and A. M.-C. So, "Local strong convexity of source localization and error bound for target tracking under time-of-arrival measurements," *IEEE Transactions on Signal Processing*, vol. 70, pp. 190–201, Dec. 2021.
- [12] F. Yin, Y. Zhao, F. Gunnarsson, and F. Gustafsson, "Received-signal-strength threshold optimization using gaussian processes," *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 2164–2177, Jan. 2017.
- [13] N. Okello, F. Fletcher, D. Musicki, and B. Ristic, "Comparison of recursive algorithms for emitter localisation using tdoa measurements from a pair of uavs," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 3, pp. 1723–1732, July 2011.
- [14] S. Xu and K. Doğançay, "Optimal sensor placement for 3-d angle-of-arrival target localization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 3, pp. 1196–1211, Feb. 2017.
- [15] Y. Lin, Z. Zhang, and H. E. Najafabadi, "Underwater source localization using time difference of arrival and frequency difference of arrival measurements based on an improved invasive weed optimization algorithm," *IET Signal Processing*, vol. 16, no. 3, pp. 299–309, May 2022.
- [16] D. R. Luke, S. Sabach, M. Teboulle, and K. Zatlaway, "A simple globally convergent algorithm for the nonsmooth nonconvex single source localization problem," *Journal of Global Optimization*, vol. 69, pp. 889–909, Dec. 2017.
- [17] T. Jia and R. M. Buehrer, "A set-theoretic approach to collaborative position location for wireless networks," *IEEE Transactions on Mobile Computing*, vol. 10, no. 9, pp. 1264–1275, Dec. 2010.
- [18] D. Blatt and A. O. Hero, "Energy-based sensor network source localization via projection onto convex sets," *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3614–3619, Aug. 2006.
- [19] Q. Shi and C. He, "Distributed source localization via projection onto the nearest local minimum," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, Mar. 2008, pp. 2553–2556.
- [20] J. Wang and P. A. Regalia, "Sensor network localization via boundary projections," in *2009 43rd Annual Conference on Information Sciences and Systems*. IEEE, Mar. 2009, pp. 224–229.
- [21] C.-L. Wang, D.-S. Wu, S.-C. Chen, and K.-J. Yang, "A decentralized positioning scheme based on recursive weighted least squares optimization for wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 10, pp. 4887–4893, Nov. 2014.
- [22] A. Zaeemzadeh, M. Joneidi, B. Shahrabi, and N. Rahnavard, "Robust target localization based on squared range iterative reweighted least squares," in *2017 IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems*. IEEE, Oct. 2017, pp. 380–388.
- [23] Y. Zhang, Y. Lou, Y. Hong, and L. Xie, "Distributed projection-based algorithms for source localization in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3131–3142, Feb. 2015.
- [24] C. Zhang and Y. Wang, "Sensor network event localization via non-convex nonsmooth ADMM and augmented lagrangian methods," *IEEE Transactions on Control of Network Systems*, vol. 6, no. 4, pp. 1473–1485, Dec. 2019.
- [25] P. Oguz-Ekim, J. P. Gomes, J. Xavier, and P. Oliveira, "Robust localization of nodes and time-recursive tracking in sensor networks using

noisy range measurements,” *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3930–3942, May 2011.

- [26] B. Swenson, R. Murray, H. V. Poor, and S. Kar, “Distributed stochastic gradient descent: Nonconvexity, nonsmoothness, and convergence to local minima,” *The Journal of Machine Learning Research*, vol. 23, no. 1, pp. 14 751–14 812, Oct. 2022.