Robust Target Localization Based on Squared Range Iterative Reweighted Least Squares

Alireza Zaeemzadeh, Mohsen Joneidi, Behzad Shahrasbi, and Nazanin Rahnavard
Presenter: Alireza Zaeemzadeh
Communications and Wireless Networks (CWN) Lab
Electrical and Computer Engineering Department
University of Central Florida
http://cwnlab.eecs.ucf.edu/

MASS 2017
14th IEEE International Conference on Mobile Ad-hoc and Sensor Systems

Outline

- Motivation and Related Work
- Our Contribution
- System Model
- Robust Localization From Squared-Range Measurements
  - SR-IRLS Algorithm
  - SR-GD Algorithm
  - Hybrid Algorithm
- Simulation Results
- Conclusion

Motivation

- In many real-world applications, some sensors might report *irrelevant data (outliers)*.
- The processing node should not blindly aggregate measurements from all sensors.

We aim to achieve *outlier distributional robustness*, which means the estimator performs well for different outlier probability distributions.

Related Work

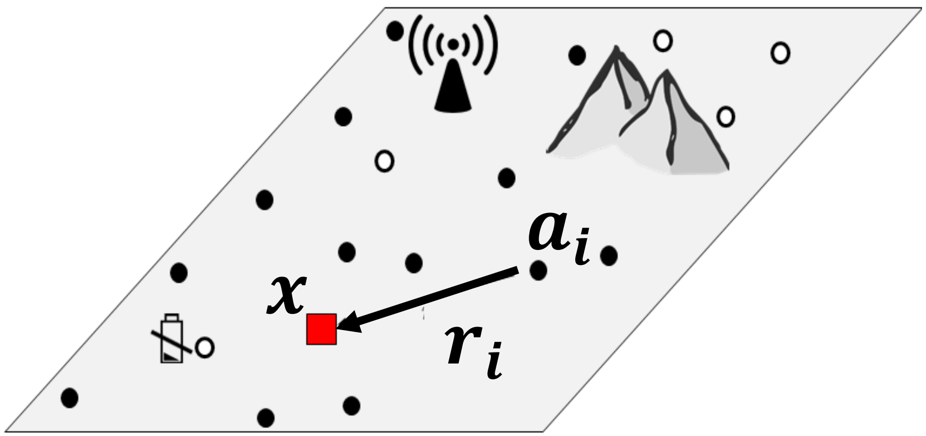
Y. Jiang et. al., ICASSP 2007.
This method is obtained by modeling the ToA estimation error as Cauchy-Lorentz distribution.
F. Yin et. al., IEEE Transactions on Signal Processing 2013
The authors have developed a robust geolocation method by estimating the probability density function (PDF) of the measurement error as a summation of Gaussian kernels. This method works best when the measurement error is drawn from a Gaussian mixture PDF.
S. Yousefi et. al., WPNC 2014.
Robust statistics, and specifically Huber norm, is exploited to localize sensors in a network in a distributed manner using the location of a subset of nodes.

Our Contribution

- A *robust* optimization problem is formulated, which disregards unreliable measurements, using squared range formulation.
- Two different algorithms are proposed to find the solution of the optimization problem.
- Convergence of the algorithms is analyzed theoretically.

Problem Statement and Model

- The system is comprised of R sensors, with known (possibly erroneous) locations, trying to localize a target.
- A portion of the measurements is irrelevant data with unknown distribution.
- The processing node has no information about the number of and the distribution of outlier measurements.



- Bluetooth
- CSI
- RSSI
- ToA
- …

Problem Statement and Model

- Range measurement:

$$r_i = \|\boldsymbol{x} - \boldsymbol{a_i}\|_2 + v_i, \qquad i = 1, \ldots, R,$$

Coordinates of the target

The location of the i$^{\text{th}}$ sensor

Measurement error

- PDF of measurement error:

$$p_V(v) = (1-\beta)\mathcal{N}(v; 0, \sigma^2) + \beta\mathcal{H}(v)$$

- Uniform
- Shifted Gaussian
- Exponential
- …

- β denotes the ratio of outlier measurements, also known as the *contamination ratio*.

Localization From Squared Range Measurements

- The conventional square-range-based least squares [1]:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \sum_{i=1}^{R} (\|\boldsymbol{x} - \boldsymbol{a_i}\|_2^2 - r_i^2)^2$$

- This formulation is not convex and is not optimal in the ML sense.
- But, it can be transformed into a special class of optimization problems with desirable characteristics:

$$\begin{aligned} \underset{\boldsymbol{x},\alpha}{\text{minimize}} \quad & \sum_{i=1}^{R} (\alpha - 2\boldsymbol{a_i}^T \boldsymbol{x} + \|\boldsymbol{a_i}\|^2 - r_i^2)^2, \\ \text{subject to} \quad & \|\boldsymbol{x}\|^2 = \alpha. \end{aligned}$$

1. A. Beck, P. Stoica, and J. Li, "Exact and Approximate Solutions of Source Localization Problems," Signal Processing, IEEE Transactions on, vol. 56, pp. 1770–1778, 5 2008.

Robustness

- *Robustness* signifies insensitivity to small deviation from the common assumption.
- Any robust proposed statistical procedure should have the following features:
  - It must be **efficient**, in the sense that it must have an optimal or near optimal performance at the assumed model
  - It must be **stable**, i.e., robust to small deviations from the assumed model.
  - Also, in the case of **breakdown**, or large deviation from the model, a catastrophe should not occur.

Robust Localization From Squared Range Measurements

- The general recipe to make any statistical procedure robust is to decompose the observations to *fitted values* and *residuals*.
- We use the residuals to assign weights to each observation.
- We define the new objective function as:

$$\mathcal{J}(\boldsymbol{y}, \boldsymbol{w}) = \sum_{i=1}^{R} w_i (\tilde{\boldsymbol{a}}_{\boldsymbol{i}}{}^T \boldsymbol{y} - b_i)^2 + \sum_{i=1}^{R} \epsilon^2 w_i - \ln w_i,$$

$$\tilde{\boldsymbol{a}}_{\boldsymbol{i}}{}^T = [\ -2\boldsymbol{a}_{\boldsymbol{i}}{}^T \quad 1\ ],\ \boldsymbol{y} = [\ \boldsymbol{x} \quad \alpha\ ]^T$$

$$b_i = r_i^2 - ||\boldsymbol{a}_{\boldsymbol{i}}||^2$$

- The new terms are added in such a way that result in the commonly used class of M-estimators known as Geman-McClure (GM) function.[1]

1. Inspired by: I. Daubechies, R. DeVore, M. Fornasier, and C. S. Gunturk, “Iteratively reweighted least squares minimization for sparse recovery,” Communications on Pure and Applied Mathematics, vol. 63, no. 1, pp. 1–38, 2010.

Robust Localization From Squared Range Measurements

- We are solving the following optimization problem:

$$\begin{aligned} \underset{\boldsymbol{y},\boldsymbol{w}}{\text{minimize}} \quad & \mathcal{J}(\boldsymbol{y},\boldsymbol{w}), \\ \text{subject to} \quad & \boldsymbol{y}^T\boldsymbol{D}\boldsymbol{y} + 2\boldsymbol{f}^T\boldsymbol{y} = 0, \\ & w_i > 0, \forall i, \end{aligned}$$

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{0}_{n\times 1} \\ \boldsymbol{0}_{1\times n} & 0 \end{bmatrix}, \ \boldsymbol{f} = \begin{bmatrix} \boldsymbol{0}_{n\times 1} \\ -0.5 \end{bmatrix}$$

Sub-problem #1	$\boldsymbol{y}^{(k+1)} = \arg\min \ \mathcal{J}(\boldsymbol{y},\boldsymbol{w}^{(k)})$, subject to $\boldsymbol{y}^T\boldsymbol{D}\boldsymbol{y} + 2\boldsymbol{f}^T\boldsymbol{y} = 0.$
Sub-problem #2	$\boldsymbol{w}^{(k+1)} = \arg\min \ \mathcal{J}(\boldsymbol{y}^{(k+1)},\boldsymbol{w})$, subject to $w_i > 0, \forall i.$

Robust Localization From Squared Range Measurements

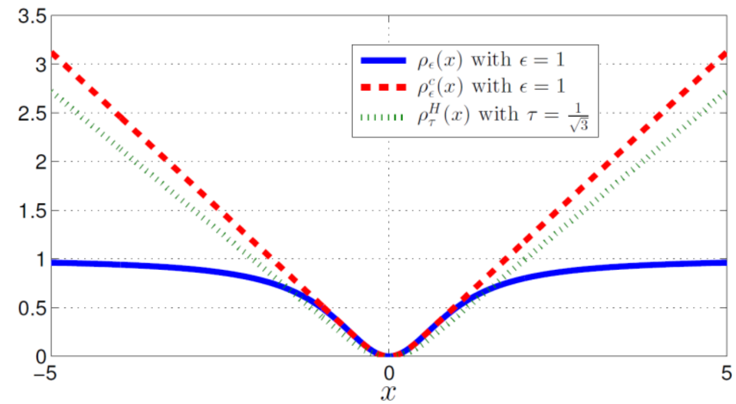
- Sub-problem #2 is convex and the global minimizer can be obtained easily.

$$w_i^{(k)} = \frac{1}{(e_i^{(k)})^2 + \epsilon^2},$$

$$\text{where} \quad e_i^{(k)} = \tilde{\boldsymbol{a}}_{\boldsymbol{i}}^T \boldsymbol{y}^{(k)} - b_i.$$

- Choosing such weights is common in iteratively reweighted least square (IRLS) methods.

- $e_i \ll \epsilon \Rightarrow w_i\left(\tilde{\boldsymbol{a}}_i^T \boldsymbol{y} - b_i\right)^2 \to 0$
- $e_i \gg \epsilon \Rightarrow w_i\left(\tilde{\boldsymbol{a}}_i^T \boldsymbol{y} - b_i\right)^2 \to 1$



Robust Localization From Squared Range Measurements

- Sub-problem #1:

$$\begin{aligned} \boldsymbol{y}^{(k+1)} = \quad & \arg\min \quad \mathcal{J}(\boldsymbol{y}, \boldsymbol{w}^{(k)}), \\ & \text{subject to} \quad \boldsymbol{y}^T \boldsymbol{D} \boldsymbol{y} + 2\boldsymbol{f}^T \boldsymbol{y} = 0. \end{aligned}$$

- Two different algorithms:
  - **SR-IRLS:** At each iteration, the global minimizer is derived.
  - **SR-GD:** We limit ourselves to be near the previous estimate.

The Squared Range Iterative Reweighted Least Squares (**SR-IRLS**)

- Matrix form of sub-problem #1:

$$\begin{aligned} \underset{\boldsymbol{y}}{\text{minimize}} \quad & (\boldsymbol{A}\boldsymbol{y}-\boldsymbol{b})^T \boldsymbol{W}^{(k-1)} (\boldsymbol{A}\boldsymbol{y}-\boldsymbol{b}), \\ \text{subject to} \quad & \boldsymbol{y}^T \boldsymbol{D}\boldsymbol{y} + 2\boldsymbol{f}^T \boldsymbol{y} = 0, \end{aligned}$$

with

$$\boldsymbol{A} = \begin{bmatrix} -2\boldsymbol{a_1}^T & 1 \\ \vdots & \vdots \\ -2\boldsymbol{a_R}^T & 1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} r_1^2 - \|\boldsymbol{a_1}\|^2 \\ \vdots \\ r_R^2 - \|\boldsymbol{a_R}\|^2 \end{bmatrix},$$

- A quadratic objective function is being minimized subject to a quadratic equality constraint.
- This special class of optimization problems is called **Generalized Trust Region Sub-problems (GTRS)**[1].

1. Moré, Jorge J. "Generalizations of the trust region problem." *Optimization methods and Software* 2.3-4 (1993): 189-209.

The Squared Range Iterative Reweighted Least Squares (**SR-IRLS**) (2)

- Hessian of the objective function is positive definite.
- Sub-problem #1 has a global minimizer for all the iterations.
- **KKT conditions**

$$\boldsymbol{y}(\lambda) = (\boldsymbol{A}^T \boldsymbol{W}^{(k-1)} \boldsymbol{A} + \lambda \boldsymbol{D})^{-1} (\boldsymbol{A}^T \boldsymbol{W}^{(k-1)} \boldsymbol{b} - \lambda \boldsymbol{f})$$

$$\boldsymbol{y}(\lambda)^T \boldsymbol{D} \boldsymbol{y}(\lambda) + 2 \boldsymbol{f}^T \boldsymbol{y}(\lambda) = 0$$

$$\boldsymbol{A}^T \boldsymbol{W}^{(k-1)} \boldsymbol{A} + \lambda \boldsymbol{D} \succeq 0.$$

- λ is the Lagrange multiplier.

The Squared Range Iterative Reweighted Least Squares (**SR-IRLS**) (3)

- It can be shown that:
  1. The interval of interest is not empty and can be derived using closed form expressions.
  2. The characteristic function is strictly decreasing over this interval.
  3. λ^* can be obtained using a bisection algorithm.
  4. The objective function converges to a stationary point of the optimization problem.
- SR-IRLS finds the **global minimizer** at each iteration.
  - Fast convergence for value of the objective.
  - *The convergence of the iterates* $\mathbf{y}^{(k)}$ *cannot be proved easily.*

The Squared Range Gradient Descent (**SRGD**)

- Instead of finding the global minimizer, we can find steepest descent at each iteration.
- convergence of the whole-sequence of the iterates can be proven theoretically.
- New objective function[1] :

$$\begin{aligned} \boldsymbol{y}^{(k)} = \arg\min_{\boldsymbol{y}} \quad & \langle \nabla_{\boldsymbol{y}} \mathcal{J}(\hat{\boldsymbol{y}}^{(k)}, \boldsymbol{w}^{(k-1)}), \boldsymbol{y} - \hat{\boldsymbol{y}}^{(k)} \rangle \\ & + l^{(k)} \|\boldsymbol{y} - \hat{\boldsymbol{y}}^{(k)}\|_2^2, \\ \text{subject to} \quad & \boldsymbol{y}^T \boldsymbol{D} \boldsymbol{y} + 2\boldsymbol{f}^T \boldsymbol{y} = 0, \end{aligned}$$

Find the steepest descent, while staying close to the prediction.

where

$$\hat{\boldsymbol{y}}^{(k)} = \boldsymbol{y}^{(k-1)} + \omega^{(k)}(\boldsymbol{y}^{(k-1)} - \boldsymbol{y}^{(k-2)}), \quad \omega^{(k)} = \frac{1}{12}\sqrt{\frac{l^{(k-1)}}{l^{(k)}}}$$

and $l^{(k)}$ is the Lipschitz constant of $\nabla_{\boldsymbol{y}} \mathcal{J}(\boldsymbol{y}, \boldsymbol{w}^{(k-1)})$ at the k^{th} iteration.

1. Inspired by: Xu, Yangyang, and Wotao Yin. "A globally convergent algorithm for nonconvex optimization based on block coordinate update." *Journal of Scientific Computing* (2017): 1-35.

The Squared Range Gradient Descent (**SRGD**) (3)

$$
\begin{aligned}
\boldsymbol{y}^{(k)} = \arg\min_{\boldsymbol{y}} \quad & \langle \nabla_{\boldsymbol{y}} \mathcal{J}(\hat{\boldsymbol{y}}^{(k)}, \boldsymbol{w}^{(k-1)}), \boldsymbol{y} - \hat{\boldsymbol{y}}^{(k)} \rangle \\
& + l^{(k)} \|\boldsymbol{y} - \hat{\boldsymbol{y}}^{(k)}\|_2^2, \\
\text{subject to} \quad & \boldsymbol{y}^T \boldsymbol{D} \boldsymbol{y} + 2\boldsymbol{f}^T \boldsymbol{y} = 0,
\end{aligned}
$$

- Again, a quadratic objective function is being minimized subject to a quadratic equality constraint. (**GTRS**)
- As before, this problem can be solved efficiently using a bisection algorithm.
- The value of the *new iterate is bounded to be around the previous iterate*, unlike the SR-IRLS method.

Hybrid Algorithm

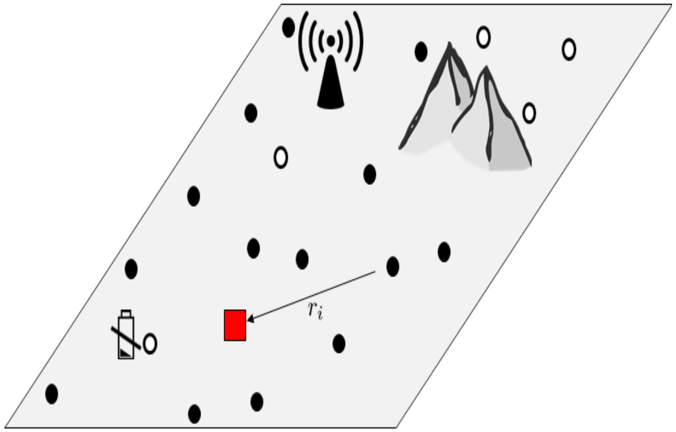
- SR-GD method needs more time to find the solution than SR-IRLS.
- To take advantage of the fast convergence of the SR-IRLS and the whole sequence convergence of the SR-GD:
  - Start with the SR-IRLS.
  - Switch to SR-GD, after convergence of the objective function, which is proven.

Results: Simulation Environment

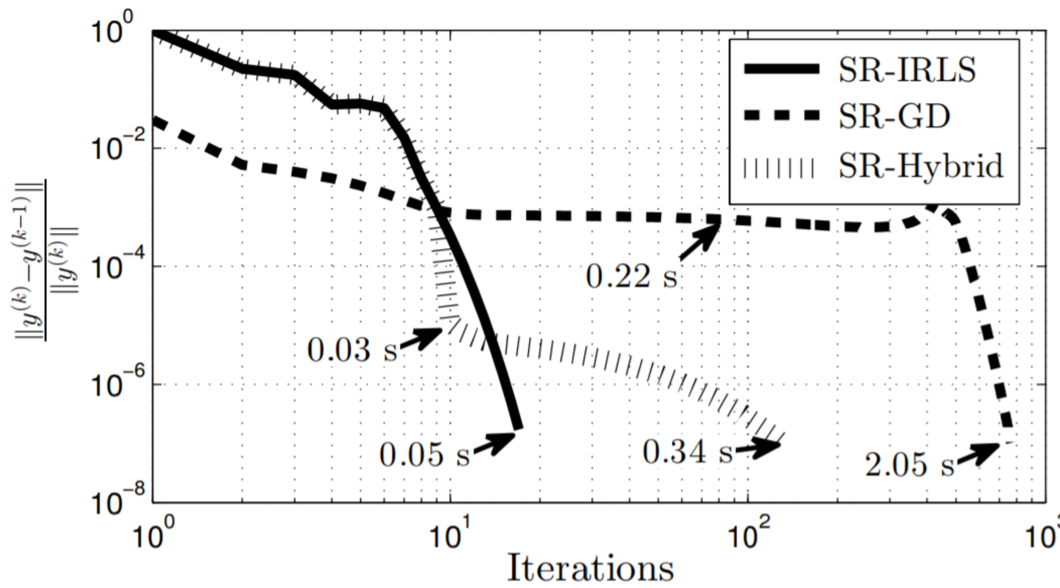
- Simulation environment $4000 \times 4000\ m^2$
- 10 sensors, distributed uniformly at random.
- 4 outlier sensors
- The noise of the outlier sensors are uniformly distributed in range $[-4000\sqrt{2}, 4000\sqrt{2}]$.
- Noise distribution:

$$p_V(v) = (1-\beta)\mathcal{N}(v; 0, \sigma^2) + \beta\mathcal{U}(v; -D_{max}, D_{max}),$$

- $\sigma = 55\ m$

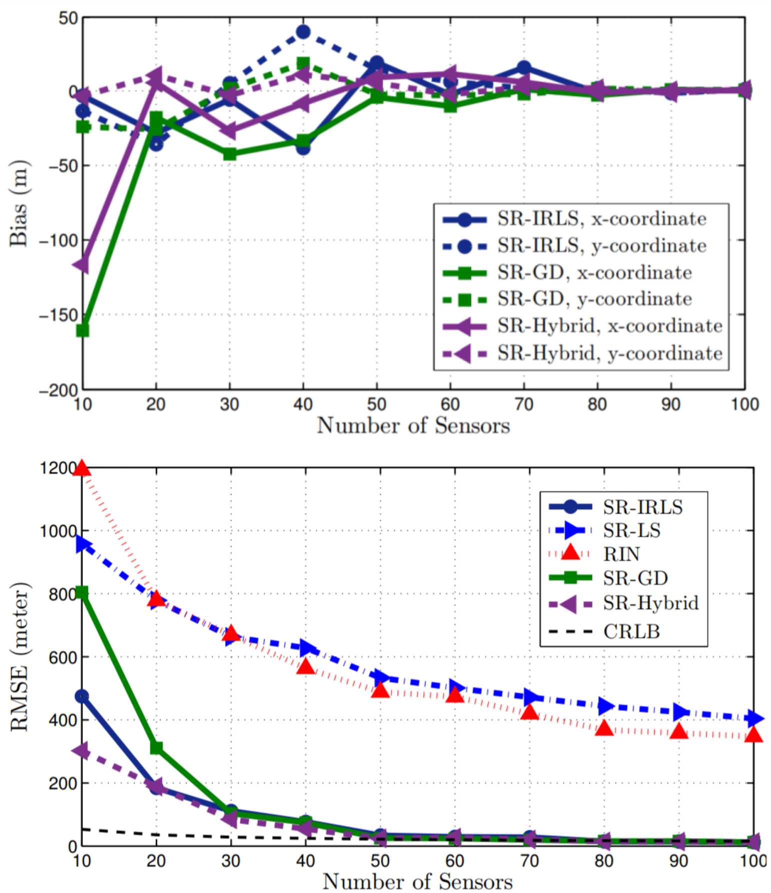


Results: Convergence



- SR-GD needs more iterations and more time to converge.
- SR-Hybrid needs less iterations than SR-GD, while its convergence is still theoretically provable.

Results: Cramer-Rao Bound

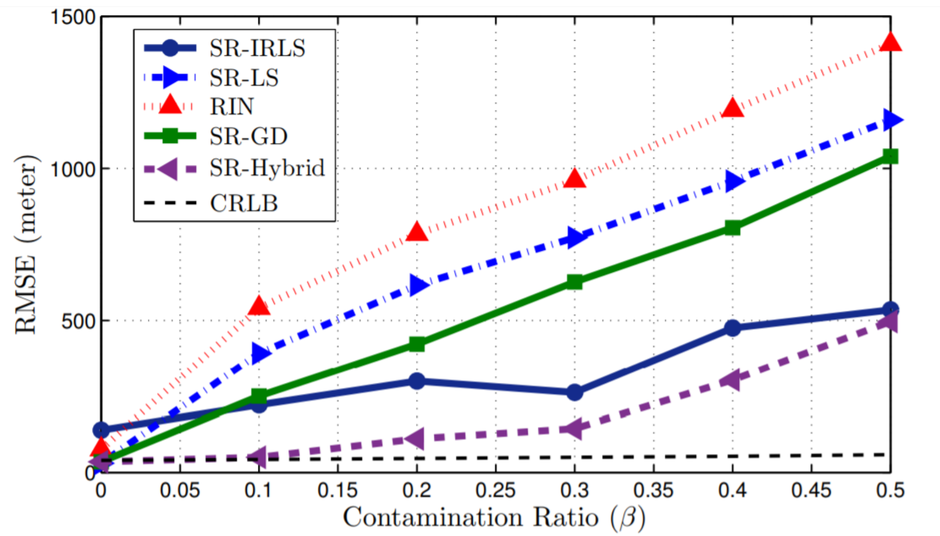


- **CRLB**: Cramer-Rao Lower Bound
- **RIN**: Robust Iterative Non-parametric[1]
- **SR-LS:** Conventional squared range solution
- 40% of the sensors are reporting unreliable data to the processing node, i. e., $\beta = 0.4$.
- For sufficiently large number of measurements, the proposed methods are efficient, *because they meet the CRLB and they are unbiased*.

F. Yin, C. Fritsche, F. Gustafsson, and A. M. Zoubir, "TOA-Based Robust Wireless Geolocation and Cramer-Rao Lower Bound Analysis in Harsh LOS/NLOS Environments," Signal Processing, IEEE Transactions on, 2013.

Results: Robustness

- The hybrid version performs the best for different values of contamination ratio.
- Robustness criteria:
  - Near optimal performance at the assumed method ($\beta = 0$),
  - Stability for small β,
  - and, for large β, a catastrophe is not occurred.



Conclusions

- Distributional robustness can be achieved by using M-estimators.
- The proposed method achieves CRLB, although it is not optimal in the ML sense.
- Future directions:
  - Distributed computation
  - Real-world dataset
  - Mobile sensors and target
  - Tracking multiple targets.

Thank you!