Robust Data Processing in the Presence of Uncertainty and Outliers: Case of Localization Problems

Anthony Welte[1], Luc Jaulin[1],
Martine Ceberio[2], and Vladik Kreinovich[2]

[1]Lab STICC, École Nationale Supérieure de Techniques
Avancées Bretagne (ENSTA Bretagne), 2 rue François Verny
29806 Brest, France, tony.welte@gmail.com, lucjaulin@gmail.com

[2]Department of Computer Science, University of Texas at El Paso
El Paso, Texas 79968, USA, mceberio@utep.edu, vladik@utep.edu

1. Outline

- To properly process data, we need to take into account:
  - the measurement errors and
  - the fact that some of the observations may be outliers.
- This is especially important in radar-based localization, where some signals may reflect:
  - not from the analyzed object,
  - but from some nearby object.
- There are known methods for situations when we have full information about the probabilities.
- There are methods for dealing with measurement errors when we only have partial info about prob.
- In this talk, we extend these methods to situations with outliers.

2. Need for Data Processing

- We are often interested in the quantities $p_1, \ldots, p_m$ which are difficult to measure directly.
- We find a measurable quantity y that depends on p_i and settings x_j: $y = f(p_1, \ldots, p_m, x_1, \ldots, x_n)$.
- For example, locating an object (robot, satellite, etc.), means finding its coordinates $p_1, \ldots$
- We cannot directly measure coordinates, but we can measure, e.g., a distance $y = \sqrt{\sum_{i=1}^{3}(p_i - x_i)^2}$.
- In general, we measure y_k under different settings $(x_{k1}, \ldots)$, and reconstruct p_i from the condition

$$y_k = f(p_1, \ldots, p_m, x_{k1}, \ldots, x_{kn}).$$

- This is an important case of *data processing*.

3. Need to Take into Account Measurement Uncertainty and Outliers

- Measurement are never absolutely accurate.
- There is always a non-zero difference between the measurement result y_k and the actual (unknown) value:

$$\Delta y_k \stackrel{\text{def}}{=} y_k - f(p_1, \ldots, p_m, x_{k1}, \ldots, x_{kn}) \neq 0.$$

- Sometimes, the measuring instrument malfunctions.
- Then, we get *outliers* – values which are very different from the actual quantity.
- This is especially important in radar-based localization, where some signals may reflect:
  - not from the analyzed object,
  - but from some nearby object.

4. Case When We Know the Probability Distribution $\rho(\Delta y)$ of the Measurement Error

- In this case, for each p, the probability to observe y_k is proportional to $\rho(\Delta y_k) = \rho(y_k - f(p, x_k))$.
- Measurement errors corresponding to different measurements are usually independent.
- So, the prob. of observing all the observed values $y_1, \ldots, y_K$ is equal to the product $\prod\limits_{k=1}^{K} \rho(y_k - f(p, x_k))$.
- It is reasonable to select the most probable value p, for which this product is the largest.
- This idea is known as the *Maximum Likelihood Method*.
- For Gaussian distributions, this leads to the usual Least Squares Method $\sum\limits_{k=1}^{K} (\Delta y_k)^2 \to \min$.

5. What If We Only Have Partial Information About the Probabilities: First Case

- Sometimes, we know that the probability distribution belongs has the form $\rho(\Delta y, \theta)$ for some $\theta = (\theta_1, \ldots, \theta_\ell)$.
- In this case, the corresponding "likelihood function" L takes the form $L = \prod\limits_{k=1}^{K} \rho(\Delta y_k, \theta)$.
- We then select a pair (p, θ) for which the probability is the largest:

$$L = \prod_{k=1}^{K} \rho(y_k - f(p, x_k), \theta) \to \max_{p,\theta}.$$

6. What If We Only Have Partial Information About the Probabilities: Non-Parametric Case

- In many practical situations, we do not know the finite-parametric family containing the actual distribution.
- Each possible distribution $\rho(\Delta y)$ can be characterized by its entropy $S = -\int \rho(\Delta y) \cdot \ln(\rho(\Delta y))\, d\Delta y$.
- Entropy describes how many binary questions we need to ask to uniquely determine Δy.
- We want to select a distribution that to the largest extent reflects this uncertainty.
- In other words, it is reasonable to select a distribution for which the entropy is the largest possible.
- For example, among the distributions $\rho(\Delta y)$ located on $[-\Delta, \Delta]$, uniform distribution has the largest entropy.

7. Need for Interval Computations

- For uniform distributions:
  - the value $\rho(\Delta y_k) = 0$ if Δy_k is outside the interval $[-\Delta, \Delta]$ and
  - it is equal to a constant when Δy_k is inside this interval.
- Thus, the product L of these probabilities is constant when $|\Delta y_k| \le \Delta$ for all k.
- So, instead of a *single* tuple p, we now need to describe *all* the tuples p for which

$$|y_k - f(p, x_k)| \le \Delta \text{ for all } k = 1, \ldots, k.$$

- This is a particular case of *interval computations*.

8. What If We Have No Information About the Probabilities of Measurement Errors

- This situation is similar to the previous one, except that now, we do not know the bound Δ.
- A reasonable idea is to select Δ for which the corresponding likelihood $L = \dfrac{1}{(2\Delta)^K}$ is the largest possible.
- Selecting the largest possible L is equivalent to selecting the smallest possible Δ.
- The only constraint on Δ is that $\Delta \geq |\Delta y_k|$ for all k.
- The smallest Δ satisfying it is $\Delta = \max\limits_k |\Delta y_k|$.
- Thus, minimizing Δ means selecting p for which $\max\limits_k |\Delta y_k| = \max\limits_k |y_k - f(p, x_k)|$ is the smallest.
- This minimax approach is indeed frequently used in data processing.

9. How to Take Both Uncertainty and Outliers into Account

- We considered 4 cases:
  - we know the exact distribution;
  - we know the finite-parametric family of distributions;
  - we know the upper bound on the (absolute value) of the corresponding difference; and
  - we have no information whatsoever, not even the upper bound.
- In principle, we may have the same four possible types of information about the outlier probabilities $\rho_0(\Delta y)$.
- At first glance, it may therefore seem that we can have $4 \times 4 = 16$ possible combinations.
- In reality, however, not all such combinations are possible.

10. Which Combinations Are Possible?

- Indeed, once we gather enough data, we can determine the corresponding probability distributions. Thus:
  - that we do not know the probability distribution of the measurement error
  - means that we have not yet collected a sufficient number of measurement results.
- The number of outliers is usually much smaller than the number of actual measurement results. So:
  - if we cannot determine the probability distribution for the measurement errors,
  - then we cannot determine the probability distribution for the outliers either.
- In general, we have less info about outliers than about the measurement errors.

11. Case When We Know Distributions $\rho(\Delta y)$ of the Measurement Error and $\rho_0(\Delta y)$ of Outliers

- If we know the set $M \subseteq \{1, \ldots, K\}$ of indices k of non-outliers, then $L = \left(\prod\limits_{k \in M} \rho(\Delta y_k)\right) \cdot \left(\prod\limits_{k \notin M} \rho_0(\Delta y_k)\right)$.
- Now, we can use the Maximum Likelihood approach to determine both the parameter tuple p and the set M.
- Max L is when we assign k to M if $\rho_0(\Delta y_k) < \rho(\Delta y_k)$, thus $L = \prod\limits_{k=1}^{K} \max(\rho(\Delta y_k)), \rho_0(\Delta y_k)) \to \max\limits_{p}$.
- From the computational viewpoint, this is similar to the usual maximum likelihood, with

$$g(\Delta y) \stackrel{\text{def}}{=} \max(\rho(\Delta y), \rho_0(\Delta y)) \text{ instead of } \rho(\Delta y).$$

- The difference is that $\int g(\Delta y)\,dy > \int \rho(\Delta y)\,dy = 1$.

12. Full Information about $\rho(\Delta y)$, Finite-Parametric Family $\rho_0(\Delta y, \varphi)$ for $\rho_0(\Delta y)$

- We can determine all the parameters (p and φ) from the requirement that the likelihood is the largest:

$$L = \prod_{k=1}^{K} \max(\rho(\Delta y_k), \rho_0(\Delta y_k, \varphi)) =$$

$$\prod_{k=1}^{K} \max(\rho(y_k - f(p, x_k)), \rho_0(y_k - f(p, x_k), \varphi)) \to \max_{p,\varphi}.$$

13. Full information about $\rho(\Delta y)$, bound W on the outlier-related differences Δy_k

- Maximum entropy approach selects uniform distr. $\rho_0(\Delta y)$ on $[-W, W]$, with $\rho_0(\Delta y_k) = \dfrac{1}{2W}$.
- We determine p that maximizes the likelihood

$$L = \prod_{k=1}^{K} \max\left(\rho(\Delta y_k), \frac{1}{2W}\right) =$$

$$\prod_{k=1}^{K} \max\left(\rho(y_k - f(p, x_k)), \frac{1}{2W}\right)$$

under the constraint $|\Delta y_k| = |y_k - f(p, x_k)| \le W$ for all k.

14. Full Information about $\rho(\Delta y)$, No Information About the Outlier-Related Differences Δy_k

- As before, in this case, we take

$$W = \max_\ell |\Delta y_\ell| = \max_\ell |y_\ell - f(p, x_\ell)|.$$

- Thus, we select the parameters p that maximize the likelihood

$$L = \prod_{k=1}^{K} \max \left(\rho(y_k - f(p, x_k)), \frac{1}{2 \cdot \max_\ell |y_\ell - f(p, x_\ell)|} \right).$$

15. Finite-Parametric Information About $\rho(\Delta y)$ and About $\rho_0(\Delta)$

- We have families of distributions $\rho(\Delta y, \theta)$ and $\rho_0(\Delta y, \varphi)$ with unknown parameters θ and φ.
- In such a situation, we find the parameters p, θ, and φ that maximize the likelihood

$$L = \prod_{k=1}^{K} \max(\rho(\Delta y_k, \theta)), \rho_0(\Delta y_k, \varphi)) =$$

$$\prod_{k=1}^{K} \max(\rho(y_k - f(p, x_k), \theta), \rho_0(y_k - f(p, x_k), \varphi)).$$

16. Finite-Parametric $\rho(\Delta y)$, Bound W on the Outlier-Related Differences Δy_k

- We have a family of distributions $\rho(\Delta y, \theta)$ with unknown parameters θ.
- In such a situation, we find the parameters p and θ that maximize the likelihood

$$L = \prod_{k=1}^{K} \max\left(\rho(\Delta y_k, \theta), \frac{1}{2W}\right) =$$

$$\prod_{k=1}^{K} \max\left(\rho(y_k - f(p, x_k), \theta), \frac{1}{2W}\right)$$

under the constraint $|\Delta y_k = y_k - f(p, x_k)| \le W$ for all k.

17. Finite-Parametric $\rho(\Delta y)$, No Information About the Outlier-Related Differences Δy_k

- Like in similar cases, we should select the smallest possible W:

$$W = \max_{\ell} |\Delta y_\ell|.$$

- Thus, we need to select the parameters p and θ that maximize the likelihood

$$L = \prod_{k=1}^{K} \max\left(\rho(y_k - f(p, x_k), \theta), \frac{1}{2 \cdot \max\limits_{\ell} |y_\ell - f(p, x_\ell)|}\right).$$

18. Bound Δ on the Measurement Errors, Bound W on the Outlier-Related Differences Δy_k

- In this case, by using the maximum entropy approach, we select the following distributions:
  - the measurement errors are uniformly distributed on the interval $[-\Delta, \Delta]$, with $\rho(\Delta y) = \dfrac{1}{2\Delta}$;
  - the differences Δy_k are uniformly distributed on the interval $[-W, W]$: $\rho_0(\Delta y) = \dfrac{1}{2W}$.
- In this case, we need to select the parameters p that maximize the likelihood $L = \prod\limits_{k=1}^{K} g(\Delta y)$, where

$$g(\Delta y) = \max(\rho(\Delta y), \rho_0(\Delta y)).$$

- Here, $g(\Delta y) = \dfrac{1}{2\Delta}$ when $|\Delta y| \le \Delta$, $g(\Delta y) = \dfrac{1}{2W}$ when $\Delta < |\Delta y| \le W$, and $g(\Delta y) = 0$ else.

19. Bound Δ on the Measurement Errors, Bound W on the Differences Δy_k (cont-d)

- Thus, maximizing the product $L = \prod\limits_{k=1} g(\Delta y_k)$ means minimizing the number of outliers under the constraint

$$|\Delta y_k| = |y_k - f(p, x_k)| \le W \text{ for all } k.$$

- So, we select p for which:
  - under these constraints,
  - the number of observations with $|y_k - f(p, x_k)| > \Delta$ is the smallest.

20. Bound Δ on the Measurement Errors, No Information About the Outlier Differences Δy_k

- In this case, since we take $W = \max\limits_{\ell} |y_\ell - f(p, x_\ell)|$, there are no longer any limitations on p.
- Thus, in this case, the maximum likelihood method simply means:
  - electing the values of the parameters p
  - for which the number of outliers (i.e., values for which $|y_k - f(p, x_k)| > \Delta$) is the smallest possible.
- This idea has been effectively used, as a heuristic idea, to deal with data processing under outliers.
- Thus, we get a probability-based justification for this heuristics.

21. Final Case, When We Have No Information About the Probabilities

- Finally, let us consider the case when we have no information about the probabilities:
  - neither about the probabilities of different values of the measurement errors,
  - nor about the probabilities of different outlier-related differences $\Delta y = y - f(p, x)$.
- In this case, we need to select the corresponding bounds Δ and W for which the likelihood is the largest.
- For each parameter tuple p, the maximum of L is attained when $W(p) = \max\limits_{\ell} |\Delta y_\ell|$.
- So, it only remains to select p and Δ.
- For each p and Δ, let us denote by $n(p, \Delta)$ the number of values k for which $|y_k - f(p, x_k)| \le \Delta$.

22. Final Case (cont-d)

- In terms of this notation, the desired likelihood value $L(p, \Delta) = \prod_{k=1}^{K} g(y_k - f(p, x_k))$ has the form

$$L(p, \Delta) = \frac{1}{(2\Delta)^{n(p,\Delta)}} \cdot \frac{1}{(2W(p))^{K-n(p,\Delta)}}.$$

- Maximizing this expression is equivalent to minimizing its minus logarithm $\psi(p, \Delta) = -\ln(L(p, \Delta)) =$

$$K \cdot \ln(2W(p)) + n(p, \Delta) \cdot (\ln(\Delta) - \ln(W(p))).$$

- Thus, we then select p for which the following expression is the smallest possible:

$$\psi(p) = \min_{\Delta}(K \cdot \ln(2W(p)) + n(p, \Delta) \cdot (\ln(\Delta) - \ln(W(p)))),$$

$$\text{where } W(p) = \max_{\ell} |y_\ell - f(p, x_\ell)| \text{ and}$$

$$n(p, \Delta) = \#\{k : |y_k - f(p, x_k)| \le \Delta\}.$$

23. Final Case: Checking How Well This Method Works

- We applied this idea to situations when Δy_k are distributed according to several reasonable distributions:
  - normal,
  - heavy-tailed power law, etc.
- In all these cases, we get 5-20% values classified as outliers.
- This is in line with the usual case of normal distribution, where:
  - 5% of the values lie outside the 2σ interval and
  - are, thus, usually dismissed as outliers.

24. Acknowledgments

- This work was supported in part:
  - by the National Science Foundation grants:
    - HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and
    - DUE-0926721, and
  - by an award from Prudential Foundation.
- This research was performed during Anthony Welte's visit to the University of Texas at El Paso.
- The authors are also thankful:
  - to all the participants of the Summer Workshop on Interval Methods SWIM'2016 (Lyon, France)
  - for valuable discussions.