A new robust and efficient estimator for ill-conditioned linear inverse problems with outliers

Marta Martinez-Camara[1], Michael Muma[2],
Abdelhak M. Zoubir[2], Martin Vetterli[1]

[1]School of Computer and Communication Sciences
École Polytechnique Fédérale de Lausanne (EPFL)

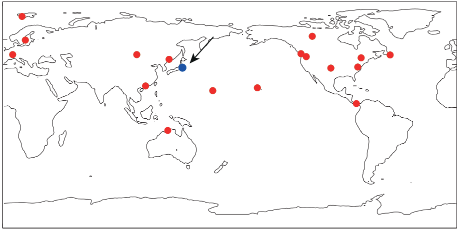
[2]Signal Processing Group
Technische Universität Darmstadt

23 April, 2015

Outline

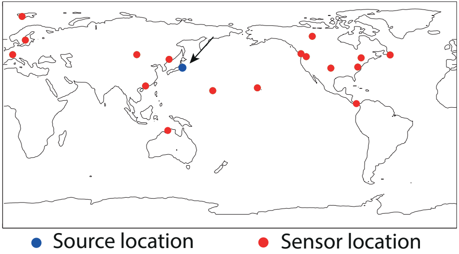
1. Motivation
2. Problem formulation
3. Background on robust estimators
4. New regularized τ-estimator
5. Algorithm
6. Results
7. Conclusions

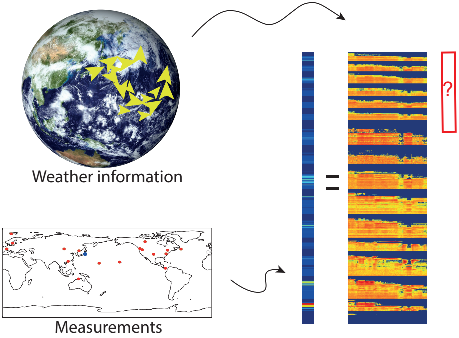
Motivation



● Source location ● Sensor location

Motivation


Source location
Sensor location


Weather information
Measurements
=
?

Problem formulation

Consider the following linear inverse problem

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$$

- $\mathbf{y}$: measurement vector
- $\mathbf{A}$: known deterministic matrix
- $\mathbf{e}$: error term
- $\mathbf{x}$: unknown parameter vector

Problem formulation

Consider the following linear inverse problem

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$$

- $\mathbf{y}$: measurement vector
- $\mathbf{A}$: known deterministic matrix
- $\mathbf{e}$: error term
- $\mathbf{x}$: unknown parameter vector

Typical assumptions

- $\mathbf{A}$ is well conditioned
- Distribution of $\mathbf{e}$ is Gaussian

Problem formulation

Consider the following linear inverse problem

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$$

- $\mathbf{y}$: measurement vector
- $\mathbf{A}$: known deterministic matrix
- $\mathbf{e}$: error term
- $\mathbf{x}$: unknown parameter vector

Typical assumptions

- $\mathbf{A}$ is well conditioned
- Distribution of $\mathbf{e}$ is Gaussian

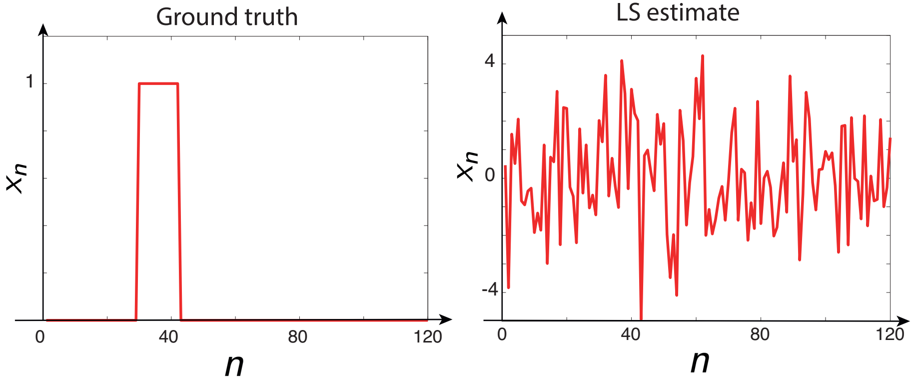
Standard estimator: least-squares (LS)

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$$

Difficulty 1

Ill-conditioned problem:

A has a large condition number → LS estimate fails

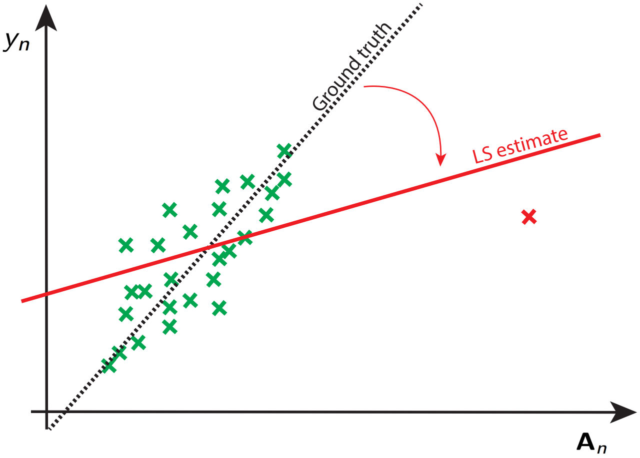


($\mathbf{A} \in \mathbb{R}^{300 \times 120}$, condition number = 1000, Gaussian errors SNR = 10 dB)

Difficulty 2

Impulsive noise and outliers

e contains outliers → LS estimate breaks down



Goals of this work

1. Design an estimator that is simultaneously
   - robust against outliers,
   - near optimal with Gaussian errors, and
   - can handle **A** with a large condition number.

Goals of this work

1. Design an estimator that is simultaneously
   - robust against outliers,
   - near optimal with Gaussian errors, and
   - can handle $\mathbf{A}$ with a large condition number.
2. Develop a fast and reliable algorithm to compute the estimates.

Goals of this work

1. Design an estimator that is simultaneously
    - robust against outliers,
    - near optimal with Gaussian errors, and
    - can handle **A** with a large condition number.
2. Develop a fast and reliable algorithm to compute the estimates.

Proposed approach: regularized τ-estimator

- A robust and efficient loss function.
- A penalty term for regularization.

Background: robust estimation

Least Squares estimator

$$\widehat{\mathbf{x}}_{LS} = \arg\min_{\mathbf{x}} \sum_{n=1}^{N} (r_n(\mathbf{x}))^2$$

- $r_n(\mathbf{x}) = y_n - \mathbf{A}_n\mathbf{x}$
- $\mathbf{A}_n$ is the n-th row of $\mathbf{A}$

Background: robust estimation

Least Squares estimator

$$\widehat{\mathbf{x}}_{LS} = \arg\min_{\mathbf{x}} \sum_{n=1}^{N} (r_n(\mathbf{x}))^2$$

- $r_n(\mathbf{x}) = y_n - \mathbf{A}_n \mathbf{x}$
- $\mathbf{A}_n$ is the n-th row of $\mathbf{A}$
- Optimal in the sense that the variance of the estimate ($\sigma^2_{x_{LS}}$) is minimised with Gaussian noise.

Background: robust estimation – M estimation

Least Squares estimator

$$\widehat{\mathbf{x}}_{LS} = \arg\min_{\mathbf{x}} \sum_{n=1}^{N} \rho(r_n(\mathbf{x}))$$

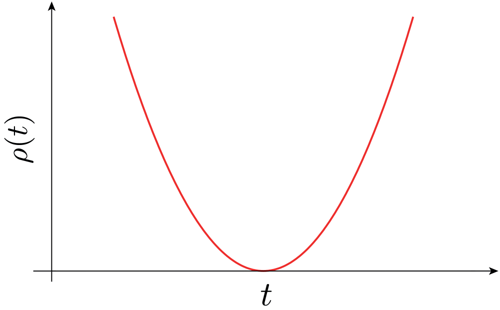
- $r_n(\mathbf{x}) = y_n - \mathbf{A}_n\mathbf{x}$
- $\mathbf{A}_n$ is the n-th row of $\mathbf{A}$
- Optimal in the sense that the variance of the estimate ($\sigma^2_{x_{LS}}$) is minimised with Gaussian noise.

Background: robust estimation – M estimation

Least Squares estimator

$$\widehat{\mathbf{x}}_{LS} = \arg\min_{\mathbf{x}} \sum_{n=1}^{N} \rho(r_n(\mathbf{x}))$$

- $r_n(\mathbf{x}) = y_n - \mathbf{A}_n \mathbf{x}$
- $\mathbf{A}_n$ is the n-th row of $\mathbf{A}$
- Optimal in the sense that the variance of the estimate ($\sigma^2_{x_{LS}}$) is minimised with Gaussian noise.

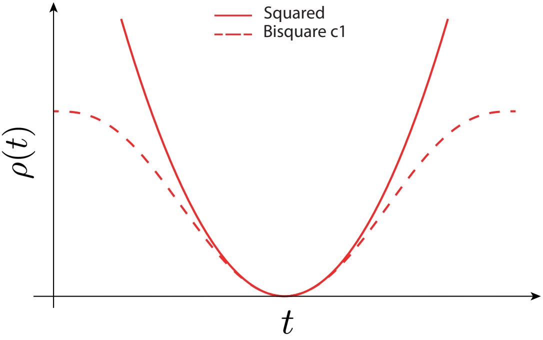


Background: robust estimation

M estimation

$$\widehat{\mathbf{x}}_{LS} = \arg\min_{\mathbf{x}} \sum_{n=1}^{N} \rho(r_n(\mathbf{x}))$$

- $r_n(\mathbf{x}) = y_n - \mathbf{A}_n\mathbf{x}$
- $\mathbf{A}_n$ is the n-th row of $\mathbf{A}$

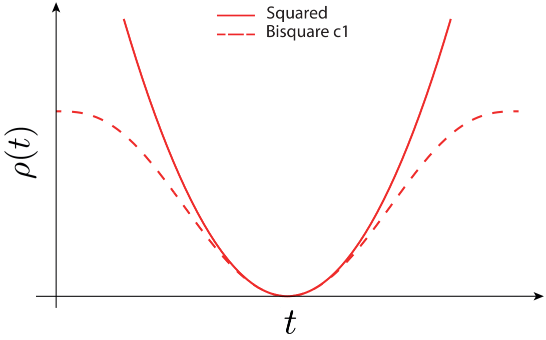


Background: robust estimation – M estimation

M estimation

$$\widehat{\mathbf{x}}_M = \arg\min_{\mathbf{x}} \sum_{n=1}^{N} \rho(r_n(\mathbf{x}))$$

- $r_n(\mathbf{x}) = y_n - \mathbf{A}_n\mathbf{x}$
- $\mathbf{A}_n$ is the n-th row of $\mathbf{A}$

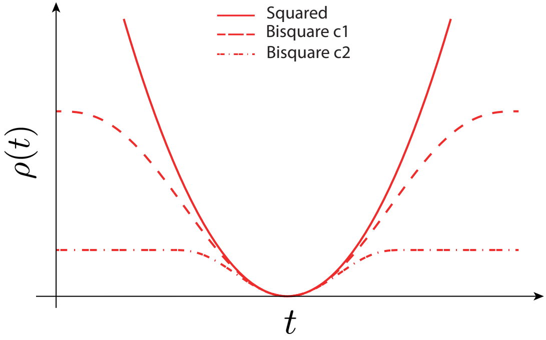


Background: robust estimation – M estimation

M estimation

$$\widehat{\mathbf{x}}_M = \arg\min_{\mathbf{x}} \sum_{n=1}^{N} \rho(r_n(\mathbf{x}))$$

- $r_n(\mathbf{x}) = y_n - \mathbf{A}_n\mathbf{x}$
- $\mathbf{A}_n$ is the n-th row of $\mathbf{A}$

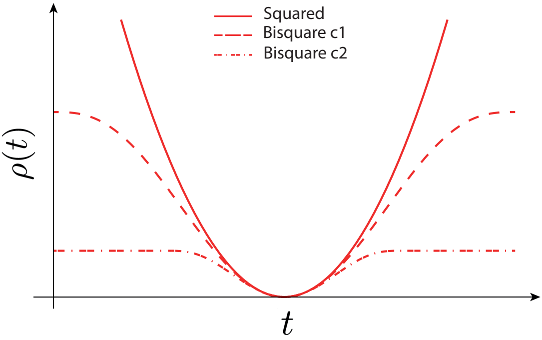


Background: robust estimation – M estimation

M estimation

$$\widehat{\mathbf{x}}_M = \arg\min_{\mathbf{x}} \sum_{n=1}^{N} \rho\left(\frac{r_n(\mathbf{x})}{\hat{\sigma}_M(\mathbf{r}(\mathbf{x}))}\right)$$

- $r_n(\mathbf{x}) = y_n - \mathbf{A}_n\mathbf{x}$
- $\mathbf{A}_n$ is the n-th row of $\mathbf{A}$
- $\hat{\sigma}_M(\mathbf{r}(\mathbf{x}))$: residual scale M-estimate
- $\rho(x)$: Symmetric, positive and non-decreasing on $[0, \infty]$
- If $\frac{\sigma x_{LS}}{\sigma x_M}$ close to 1 with Gaussian errors, efficient



Background: robust estimation – τ estimation

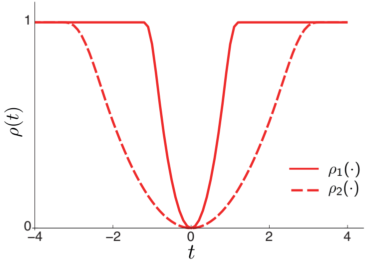
- Choosing $\rho(x)$ based on the data

Background: robust estimation – τ estimation

- Choosing $\rho(x)$ based on the data
- Asymptotically equivalent to an M-estimator

$$\rho(\cdot) = w(\cdot)\,\underbrace{\rho_1(\cdot)}_{\textit{robust}} + \underbrace{\rho_2(\cdot)}_{\textit{efficient}}$$

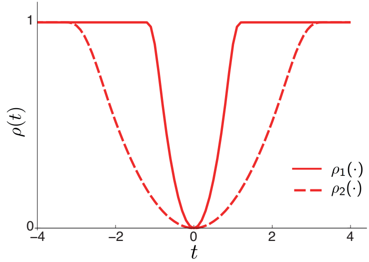
- $w(\cdot)$: adapts to the distribution of the data



Background: robust estimation – τ estimation

- Choosing $\rho(x)$ based on the data
- Asymptotically equivalent to an M-estimator

$$\rho(\cdot) = w(\cdot)\,\underbrace{\rho_1(\cdot)}_{robust} + \underbrace{\rho_2(\cdot)}_{efficient}$$



- $w(\cdot)$: adapts to the distribution of the data
- $\widehat{\mathbf{x}}_\tau$ minimizes a robust and efficient τ-scale estimate

$$\widehat{\mathbf{x}}_\tau = \arg\min_{\mathbf{x}} \hat{\sigma}_\tau(\mathbf{r}(\mathbf{x}))$$

Recap

	Robust	Efficient	Ill posed
M estimator robust	✓	✗	✗
M estimator efficient	✗	✓	✗
τ estimator	✓	✓	✗

New regularized τ-estimator

Proposed estimator

$$\widehat{\mathbf{x}}_\tau = \arg\min_{\mathbf{x}} \hat{\sigma}_\tau(\mathbf{r}(\mathbf{x})) + \lambda\|\mathbf{x}\|_2$$

- $\hat{\sigma}_\tau(\mathbf{r}(\mathbf{x}))$: τ-estimate of the scale
- $\lambda \geq 0$: regularization parameter

New regularized τ-estimator

Proposed estimator

$$\widehat{\mathbf{x}}_\tau = \arg\min_{\mathbf{x}} \hat{\sigma}_\tau(\mathbf{r}(\mathbf{x})) + \lambda\|\mathbf{x}\|_2$$

- $\hat{\sigma}_\tau(\mathbf{r}(\mathbf{x}))$: τ-estimate of the scale
- $\lambda \geq 0$: regularization parameter

Key difficulty: How to compute the regularized τ estimate?

- Non-convex function
- No guarantees of finding the global minimum

Algorithm

Steps:

1. How to find local minima
2. How to find the global one
3. Speeding up the algorithm

Algorithm

Step 1: finding local minima

- Equivalent to Iterative Reweighted Least Squares (IRLS)

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{W}(\mathbf{x})(\mathbf{y} - \mathbf{A}\mathbf{x})\|_2^2 + \lambda^2 \|\mathbf{x}\|_2^2$$

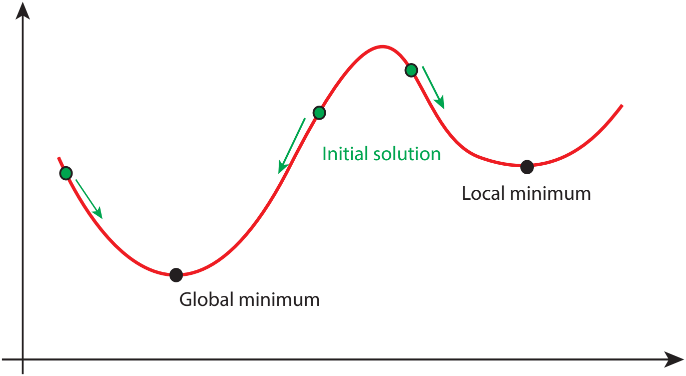
- $\mathbf{W}(\mathbf{x})$: data adaptive term that we derive from

$$\frac{\partial(\hat{\sigma}_\tau^2(\mathbf{r}(\mathbf{x})) + \lambda\|\mathbf{x}\|_2^2)}{\partial \mathbf{x}} = 0$$

Algorithm

Step 2: finding the global minima

- We take many different initial solutions...



- ... and we hope to find the correct valley!

Algorithm

Step 3: speeding up the algorithm

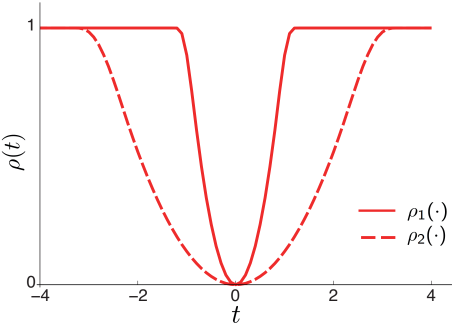
- For each initial solution, make only a few IRLS iterations.
  $\rightarrow$ fast convergence
- Pick the N best solutions.
- Use them as new initial solutions.
- Iterate IRLS until convergence.

Results

Experimental setup

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}_G + \mathbf{e}_o$$

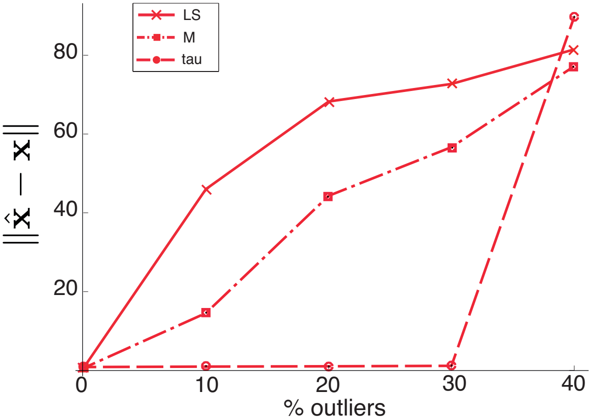
- $\mathbf{A} \in \mathbb{R}^{300\times 120}$: random iid Gaussian
- $\mathbf{x}$: piecewise constant
- $\mathbf{e}_G$: Gaussian noise
- $\mathbf{e}_o$: sparse vector, entries with large variance (outliers)
- λ: determined experimentally



Results – with previous estimators

Non-regularized LS-estimator, M-estimator, and τ-estimator

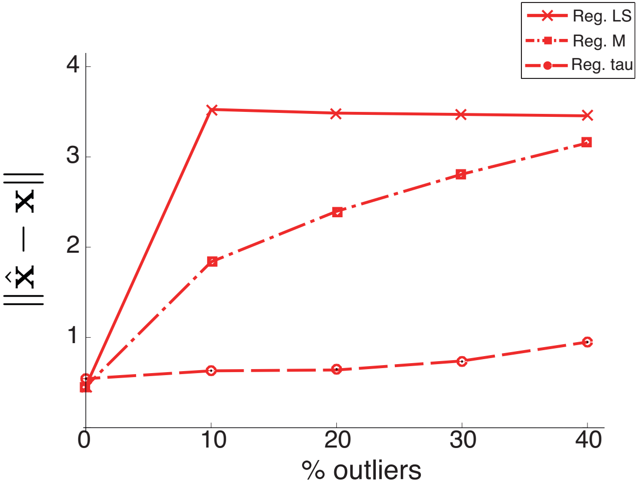
- $\mathbf{A}$ with a condition number of 50.
- $\|\hat{\mathbf{x}} - \mathbf{x}\|$: Monte Carlo average.



Results – with new estimator

Regularized LS-estimator, M-estimator, and τ-estimator

- $\mathbf{A}$ with a condition number of 1000.
- $\|\hat{\mathbf{x}} - \mathbf{x}\|$: Monte Carlo average.



Conclusions

- New regularized robust estimator
  - highly robust against outliers
  - highly efficient in the presence of Gaussian noise
  - stable when the mixing matrix has a large condition number

Conclusions

- New regularized robust estimator
  - highly robust against outliers
  - highly efficient in the presence of Gaussian noise
  - stable when the mixing matrix has a large condition number
- Current research
  - Study the interrelation of robustness and regularization
  - Develop Lasso-type regularized τ-estimator
  - Derivation of influence function
  - Application to real data

Conclusions

- New regularized robust estimator
  - highly robust against outliers
  - highly efficient in the presence of Gaussian noise
  - stable when the mixing matrix has a large condition number
- Current research
  - Study the interrelation of robustness and regularization
  - Develop Lasso-type regularized τ-estimator
  - Derivation of influence function
  - Application to real data
- Reproducible Results
  - `https://github.com/LCAV/RegularizedTauEstimator`

Thank you for your attention.

Thank you for your attention.

Questions?