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

Marta Martinez-Camara\textsuperscript{1}, Michael Muma\textsuperscript{2},
Abdelhak M. Zoubir\textsuperscript{2}, Martin Vetterli\textsuperscript{1}

\textsuperscript{1}School of Computer and Communication Sciences
École Polytechnique Fédérale de Lausanne (EPFL)

\textsuperscript{2}Signal Processing Group
Technische Universität Darmstadt

23 April, 2015
1. Motivation

2. Problem formulation

3. Background on robust estimators

4. New regularized $\tau$-estimator

5. Algorithm

6. Results

7. Conclusions
Motivation
Motivation

Source location

Sensor location

Weather information

Measurements
Problem formulation

Consider the following linear inverse problem

$$y = Ax + e$$

- **y**: measurement vector
- **A**: known deterministic matrix
- **e**: error term
- **x**: unknown parameter vector
Problem formulation

Consider the following linear inverse problem

\[ y = Ax + e \]

- \( y \): measurement vector
- \( A \): known deterministic matrix
- \( e \): error term
- \( x \): unknown parameter vector

Typical assumptions

- \( A \) is well conditioned
- Distribution of \( e \) is Gaussian

Standard estimator: least-squares (LS)

\[ \hat{x} = \arg \min_x \| y - Ax \|_2^2 \]
Problems formulation

Consider the following linear inverse problem

\[ y = Ax + e \]

- **y**: measurement vector
- **A**: known deterministic matrix
- **e**: error term
- **x**: unknown parameter vector

Typical assumptions

- **A** is well conditioned
- Distribution of **e** is Gaussian

Standard estimator: least-squares (LS)

\[ \hat{x} = \arg \min_x ||y - Ax||_2^2 \]
Ill-conditioned problem:

\( \mathbf{A} \) has a large condition number \( \rightarrow \) LS estimate fails

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

e contains outliers $\rightarrow$ 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 $\mathbf{A}$ with a large condition number.

Proposed approach: regularized $\tau$-estimator
- A robust and efficient loss function.
- A penalty term for regularization.
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.
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 $\tau$-estimator
   ▶ A robust and efficient loss function.
   ▶ A penalty term for regularization.
Background: robust estimation

Least Squares estimator

\[ \hat{x}_{LS} = \arg \min_x \sum_{n=1}^{N} (r_n(x))^2 \]

- \( r_n(x) = y_n - A_n x \)
- \( A_n \) is the n-th row of \( A \)
Background: robust estimation

Least Squares estimator

\[ \hat{x}_{LS} = \arg \min_x \sum_{n=1}^{N} (r_n(x))^2 \]

- \( r_n(x) = y_n - A_n x \)
- \( A_n \) is the \( n \)-th row of \( A \)
- Optimal in the sense that the variance of the estimate \( (\sigma_{\hat{x}_{LS}}^2) \) is minimised with Gaussian noise.
Least Squares estimator

\[ \hat{x}_{LS} = \arg \min_x \sum_{n=1}^{N} \rho(r_n(x)) \]

- \( r_n(x) = y_n - A_n x \)
- \( A_n \) is the n-th row of \( A \)
- Optimal in the sense that the variance of the estimate \( \sigma^2_{\hat{x}_{LS}} \) is minimised with Gaussian noise.
Least Squares estimator

\[ \hat{x}_{LS} = \arg \min_x \sum_{n=1}^{N} \rho(r_n(x)) \]

- \( r_n(x) = y_n - A_n x \)
- \( A_n \) is the n-th row of \( A \)
- Optimal in the sense that the variance of the estimate \( \sigma_{\hat{x}_{LS}}^2 \) is minimised with Gaussian noise.
M estimation

$$\hat{x}_{LS} = \arg\min_x \sum_{n=1}^{N} \rho(r_n(x))$$

- $$r_n(x) = y_n - A_n x$$
- $$A_n$$ is the n-th row of $$A$$
M estimation

\[ \hat{x}_M = \arg\min_x \sum_{n=1}^{N} \rho(r_n(x)) \]

- \( r_n(x) = y_n - A_n x \)
- \( A_n \) is the n-th row of \( A \)
Background: robust estimation – M estimation

M estimation

\[ \hat{x}_M = \arg \min_x \sum_{n=1}^{N} \rho(r_n(x)) \]

- \( r_n(x) = y_n - A_nx \)
- \( A_n \) is the n-th row of \( A \)
M estimation

\[ \hat{x}_M = \arg \min_x \sum_{n=1}^{N} \rho \left( \frac{r_n(x)}{\hat{\sigma}_M(r(x))} \right) \]

- \( r_n(x) = y_n - A_n x \)
- \( A_n \) is the n-th row of \( A \)
- \( \hat{\sigma}_M(r(x)) \): residual scale M-estimate
- \( \rho(x) \): Symmetric, positive and non-decreasing on \([0, \infty]\)
- If \( \frac{\sigma_{\text{XLS}}}{\sigma_{\text{XM}}} \) close to 1 with Gaussian errors, efficient
Background: robust estimation – \( \tau \) estimation

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

\[
\hat{x}_\tau \text{ minimizes a robust and efficient } \tau \text{-scale estimate}
\]

\[
\hat{x}_\tau = \arg \min_x \hat{\sigma}_\tau(r(x))
\]

Martinez-Camara et al.
ICASSP’15
23.04.2015 14 / 25
Choosing $\rho(x)$ based on the data

Asymptotically equivalent to an M-estimator

$\rho(\cdot) = w(\cdot) \rho_1(\cdot) + \rho_2(\cdot)$

$w(\cdot)$: adapts to the distribution of the data
Background: robust estimation – $\tau$ estimation

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

- Asymptotically equivalent to an M-estimator

$$\rho(\cdot) = w(\cdot) \rho_1(\cdot) + \rho_2(\cdot)$$

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

- $\hat{x}_\tau$ minimizes a robust and efficient $\tau$-scale estimate

$$\hat{x}_\tau = \arg \min_{x} \hat{\sigma}_\tau(r(x))$$
### Recap

<table>
<thead>
<tr>
<th></th>
<th>Robust</th>
<th>Efficient</th>
<th>Ill posed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M estimator</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>robust</td>
<td>✔</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>efficient</td>
<td>✗</td>
<td>✔</td>
<td>✗</td>
</tr>
<tr>
<td><strong>T estimator</strong></td>
<td>✔</td>
<td>✔</td>
<td>✗</td>
</tr>
</tbody>
</table>
New regularized $\tau$-estimator

Proposed estimator

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

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

**Proposed** estimator

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

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

**Key difficulty:** How to compute the regularized $\tau$ estimate?

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

\[
\hat{x} = \arg \min_x \| W(x)(y - Ax) \|^2_2 + \lambda^2 \| x \|^2_2
\]

- \( W(x) \): data adaptive term that we derive from

\[
\frac{\partial (\hat{\sigma}_\tau^2(r(x)) + \lambda \| x \|^2_2)}{\partial x} = 0
\]
Step 2: finding the global minima

- We take many different initial solutions...

- ... and we hope to find the correct valley!
Step 3: speeding up the algorithm

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

Experimental setup

\[ y = Ax + e_G + e_o \]

- \( A \in \mathbb{R}^{300 \times 120} \): random iid Gaussian
- \( x \): piecewise constant
- \( e_G \): Gaussian noise
- \( e_o \): sparse vector, entries with large variance (outliers)
- \( \lambda \): determined experimentally
Results – with previous estimators

Non-regularized LS-estimator, M-estimator, and $\tau$-estimator

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

Regularized LS-estimator, M-estimator, and $\tau$-estimator

- $A$ with a condition number of 1000.
- $\|\hat{x} - x\|$: 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 $\tau$-estimator
  - Derivation of influence function
  - Application to real data

Reproducible Results
https://github.com/LCAV/RegularizedTauEstimator
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 $\tau$-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?