

A NEW ROBUST AND EFFICIENT ESTIMATOR FOR ILL-CONDITIONED LINEAR INVERSE PROBLEMS WITH OUTLIERS

Marta Martinez-Camara¹, Michael Muma², Abdelhak M. Zoubir², Martin Vetterli¹

¹School of Computer and Communication Sciences,
Ecole Polytechnique Fédérale de Lausanne (EPFL),
CH-1015 Lausanne, Switzerland,
marta.martinez-camara@epfl.ch
martin.vetterli@epfl.ch

²Signal Processing Group,
Technische Universität Darmstadt,
Merckstr. 25, DE-64283 Darmstadt, Germany,
muma@spg.tu-darmstadt.de
zoubir@spg.tu-darmstadt.de

ABSTRACT

Solving a linear inverse problem may include difficulties such as the presence of outliers and a mixing matrix with a large condition number. In such cases a regularized robust estimator is needed. We propose a new τ -type regularized robust estimator that is simultaneously highly robust against outliers, highly efficient in the presence of purely Gaussian noise, and also stable when the mixing matrix has a large condition number. We also propose an algorithm to compute the estimates, based on a regularized iterative reweighted least squares algorithm. A basic and a fast version of the algorithm are given. Finally, we test the performance of the proposed approach using numerical experiments and compare it with other estimators. Our estimator provides superior robustness, even up to 40% of outliers, while at the same time performing quite close to the optimal maximum likelihood estimator in the outlier-free case.

1. INTRODUCTION

Let a linear inverse problem be given by

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}. \quad (1)$$

The unknown source signal $\mathbf{x} \in \mathbb{R}^{n \times 1}$ is to be estimated. The measurement vector $\mathbf{y} \in \mathbb{R}^{m \times 1}$ and the mixing matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ are known. Finally, the noise $\mathbf{n} \in \mathbb{R}^{n \times 1}$ is in general unknown. This type of problem appears frequently in signal processing [1], as well as in physics, geosciences [2, 3], finance, and engineering in general.

Traditionally, \mathbf{n} is assumed to have independent and identically distributed (iid) Gaussian random entries. Under this assumption, the optimal estimator is the least squares (LS) estimator [4]:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2. \quad (2)$$

However, in many modern engineering problems this idealistic noise model does not hold. The distribution of \mathbf{n} is far from Gaussian when it contains outliers, which causes the distribution to be heavy

tailed [5]. In such scenarios, the LS-estimator's performance drastically degrades. As a solution, many statistically robust estimators have been proposed in the literature. The first systematic approach to statistically robust estimation is M-estimation [6], which generalizes maximum likelihood (ML) estimation. However M-estimators are non-robust, i.e., have a breakdown point (BP) equal to zero, when leverage points are present (errors in the entries of \mathbf{A}). The BP is the proportion of outliers in the data that an estimator can handle before giving an uninformative (e.g. arbitrarily large) result.

A second fundamental class of robust estimators are the ones that are based on the minimization of a robust residual scale. This class includes estimators such as the least-trimmed-squares (LTS) [7], the least-median-of-squares (LMS) [7], and the S-estimators [8]. The latter minimizes the M-estimate of scale of the residuals. S-estimators can be highly robust, i.e., have a high BP. However, in this case, they do not have high efficiency, i.e., the variance of the S-estimator, when it is tuned for high robustness, is much larger than for the ML-estimator. This led to the development of the third class of robust estimators, which are simultaneously robust and efficient. The τ -estimator [9] belongs to this class.

Another common difficulty is \mathbf{A} having a large condition number. In such cases, the LS solution $\hat{\mathbf{x}}$ in (2) does not depend continuously on the problem data \mathbf{y} and \mathbf{A} , i.e., it is unstable. Such inverse problems are said to be ill-posed. To stabilize $\hat{\mathbf{x}}$, it is possible to introduce additional a priori information in the problem. This is referred to as regularization. Tikhonov regularization is one of the oldest and most well-known techniques for dealing with ill-posed inverse problems [10]:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \|\mathbf{x}\|_2^2. \quad (3)$$

Moreover, some inverse problems suffer from both outliers and a matrix \mathbf{A} with a large condition number [11]. In these cases, a regularized robust estimator is needed.

Recently, there has been increased interest in regularized estimators in the robust statistics community. In particular, regularized M-estimators [12], S-estimators [13], and MM-estimators [14] were proposed. Of these, only the MM-estimator is simultaneously efficient and robust against leverage points [5].

We propose a regularized estimator of the τ -type, which is simultaneously highly robust, efficient, and stable when \mathbf{A} has a large condition number. An advantage of our proposed estimator over the

This work was supported by a SNF Grant: Non-linear Sampling Methods – FNS-200021_138081, by the project HANDiCAMS which acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 323944. and by an ERC Advanced Grant Support for Frontier Research SPARSAM Nr: 247006.

regularized MM-estimator is that it provides a high BP and a highly efficient estimate of the scale of the errors, in addition to the regularized estimate $\hat{\mathbf{x}}$. We also derive an algorithm to compute the estimates, which formulates the τ -minimization problem as a regularized iterative reweighted least squares (IRLS) problem of a non-convex function. A fast version of the algorithm is provided, which is highly practical. Simulation studies are also provided to compare our estimator to the regularized LS-estimator and M-estimator in terms of solving an outlier-contaminated inverse problem, where the matrix \mathbf{A} has a high condition number. A brief discussion about how the regularization affects the robustness of the estimators is provided by running the same experiment, but using a matrix \mathbf{A} with a low condition number.

This paper is organized as follows. Section 2 introduces the proposed regularized τ -estimator. Section 3 contains the derivation of an algorithm to solve the τ -minimization problem as a regularized IRLS problem of a non-convex function in a computationally feasible way. Section 4, provides a Monte-Carlo simulation study that compares the proposed approach to the regularized LS-estimator and M-estimator in terms of the norm of the error. Section 5 concludes the paper and provides a brief outlook on future research.

2. THE REGULARIZED τ -ESTIMATOR

The τ -scale of the residuals $\sigma_\tau(\mathbf{r}(\mathbf{x}))$ [9] is defined by

$$\sigma_\tau^2(\mathbf{r}(\mathbf{x})) = \sigma_M^2(\mathbf{r}(\mathbf{x})) \frac{1}{m} \sum_{i=1}^m \rho_2 \left(\frac{r_i(\mathbf{x})}{\sigma_M(\mathbf{r}(\mathbf{x}))} \right). \quad (4)$$

Here, $\sigma_M(\mathbf{r}(\mathbf{x}))$ is the robust M-scale estimate, given by

$$\frac{1}{m} \sum_{i=1}^m \rho_1 \left(\frac{r_i(\mathbf{x})}{\sigma_M(\mathbf{r}(\mathbf{x}))} \right) = b, \quad (5)$$

with $b = E_F[\rho_1(\mathbf{r}(\mathbf{x}))]$, $E_F[\cdot]$ denoting the expectation w.r.t. the standard Gaussian distribution F , and $r_i = y_i - \mathbf{A}\mathbf{x}$, for $i \in \{1 \dots m\}$.

The choice of ρ_2 , controls the efficiency of the τ -estimator, e.g., full efficiency for the Gaussian distribution can be obtained by setting $\rho_2(r) = r^2$, which corresponds to the ML-estimator under the Gaussian assumption. The BP of the τ -estimator, on the other hand, is controlled by ρ_1 and is the same as that of an S-estimator [8] that uses ρ_1 . In this way, one can obtain BP=0.5 simultaneously with high efficiency [9].

In this paper, we introduce the regularized τ -estimator of regression as the minimizer of the sum of the τ -estimate of scale of the residuals and the Tikhonov regularization:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \tau_R^2(\mathbf{r}(\mathbf{x})), \quad (6)$$

where $\tau_R^2(\mathbf{r}(\mathbf{x})) = \sigma_\tau^2(\mathbf{r}(\mathbf{x})) + \lambda \|\mathbf{x}\|_2^2$ and $\lambda \geq 0$ is the regularization parameter.

3. COMPUTING THE ESTIMATES

A major challenge of the regularized τ -estimator is its computation. We note that (6) is a non-convex function, so finding the exact global minimum may be computationally prohibitive.

Thus, to minimize (6), we propose a modification of the heuristic algorithm given in [15]. The main idea behind this heuristic algorithm is intuitively summarized as follows: find several local minima, and select the best one among them as the final solution. There is, of course, no guarantee that this solution is the global minimum of (6).

3.1. Finding local minima

The parameter values $\hat{\mathbf{x}}$ that locally minimize $\sigma_\tau^2(\mathbf{r}(\mathbf{x}))$ are the ones with a derivative w.r.t. \mathbf{x} being equal to zero.

Defining

$$\tilde{\mathbf{r}}(\mathbf{x}) := \frac{\mathbf{r}(\mathbf{x})}{\sigma_M(\mathbf{r}(\mathbf{x}))} \quad (7)$$

$$\psi_j(\mathbf{x}) := \frac{\partial \rho_j(\mathbf{x})}{\partial \mathbf{x}}, \quad (8)$$

yields

$$\begin{aligned} \frac{\partial \tau_R^2(\mathbf{r}(\mathbf{x}))}{\partial \mathbf{x}} &= 2\sigma_M(\mathbf{r}(\mathbf{x})) \frac{\partial \sigma_M(\mathbf{r}(\mathbf{x}))}{\partial \mathbf{x}} \frac{1}{m} \sum_{i=1}^m \rho_2(\tilde{r}_i(\mathbf{x})) + \\ &+ \sigma_M^2(\mathbf{r}(\mathbf{x})) \frac{1}{m} \sum_{i=1}^m \psi_2(\tilde{r}_i(\mathbf{x})) \left[\frac{-\mathbf{A}_i \sigma_M(\mathbf{r}(\mathbf{x})) - r_i(\mathbf{x}) \frac{\partial \sigma_M(\mathbf{r}(\mathbf{x}))}{\partial \mathbf{x}}}{\sigma_M^2(\mathbf{r}(\mathbf{x}))} \right] \\ &+ \lambda \sum_{i=1}^m 2x_i = 0. \end{aligned} \quad (9)$$

To find $\frac{\partial \sigma_M(\mathbf{r}(\mathbf{x}))}{\partial \mathbf{x}}$, we take the derivative of (5) w.r.t. \mathbf{x} :

$$\frac{\partial \sigma_M(\mathbf{r}(\mathbf{x}))}{\partial \mathbf{x}} = -\sigma_M(\mathbf{r}(\mathbf{x})) \frac{\sum_{i=1}^m \psi_1(\tilde{r}_i(\mathbf{x})) \mathbf{A}_i}{\sum_{i=1}^m \psi_1(\tilde{r}_i(\mathbf{x})) r_i(\mathbf{x})} \quad (10)$$

Replacing (10) in (9) we obtain

$$\begin{aligned} \left(\sum_{i=1}^m W_m(\mathbf{x}) \psi_1(\tilde{r}_i(\mathbf{x})) - \psi_2(\tilde{r}_i(\mathbf{x})) \right) \frac{\sigma_M(\mathbf{r}(\mathbf{x})) \mathbf{A}_i}{m} + \\ + \lambda \sum_{i=1}^m 2x_i = 0, \end{aligned} \quad (11)$$

where

$$W_m(\mathbf{x}) := \frac{\sum_{i=1}^m -2\rho_2(\tilde{r}_i(\mathbf{x})) + \sum_{i=1}^m \psi_2(\tilde{r}_i(\mathbf{x})) \tilde{r}_i(\mathbf{x})}{\sum_{i=1}^m \psi_1(\tilde{r}_i(\mathbf{x})) \tilde{r}_i(\mathbf{x})}. \quad (12)$$

We can see that (11) is also $\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}$ where $f(\mathbf{x})$ is the regularized weighted least squares estimator

$$f(\mathbf{x}) = \sum_{i=1}^m w_i(\mathbf{x}) (y_i - \mathbf{A}_i \mathbf{x})^2 + \sum_{i=1}^n \lambda x_i^2 \quad (13)$$

with weights

$$w_i(\mathbf{x}) = \frac{\psi_\tau(\mathbf{x})}{2\tilde{r}_i(\mathbf{x})}, \quad (14)$$

where

$$\psi_\tau(\mathbf{x}) = -\frac{1}{m} \left(\sum_{i=1}^m W_m(\mathbf{x}) \psi_1(\tilde{r}_i(\mathbf{x})) - \psi_2(\tilde{r}_i(\mathbf{x})) \right). \quad (15)$$

So $f(\mathbf{x})$ is just another way of writing (6). Hence, $f(\mathbf{x})$ can be minimized iteratively as a regularized IRLS [16] problem with weights given by (14). The regularized IRLS algorithm for our problem is detailed in Algorithm 1. As $f(\mathbf{x})$ is a non-convex function, the result of this minimisation depends on the initial condition $\mathbf{x}[0]$ given to the IRLS.

Algorithm 1 Regularized IRLS

INPUT: $\mathbf{y} \in \mathbb{R}^m$, $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\lambda, \epsilon, \mathbf{x}[0], K$
OUTPUT: $\hat{\mathbf{x}}$
require $\lambda \geq 0, \epsilon \geq 0, K \geq 0$
for $k = 0$ **to** $K - 1$ **do**
 $\mathbf{r}[k] \leftarrow \mathbf{y} - \mathbf{A}\mathbf{x}[k]$
 Compute $\sigma_M[k]$
 Compute $\mathbf{W}[k]$
 $\mathbf{x}[k+1] \leftarrow (\mathbf{A}^\top \mathbf{W}^\top[k] \mathbf{W}[k] \mathbf{A} + \lambda^2)^{-1} \mathbf{A}^\top \mathbf{W}^\top[k] \mathbf{W}[k] \mathbf{y}$
 if $\|\mathbf{x}[k+1] - \mathbf{x}[k]\| < \epsilon$ **then**
 break
 end if
end for
return $\hat{\mathbf{x}} \leftarrow \mathbf{x}[k+1]$

With this new point of view we can think of the regularized τ -estimator as an M-estimator with a data-adaptive ψ_τ function.

Then the basic algorithm is shown below as Algorithm 2.

Algorithm 2 Basic algorithm

INPUT: $\mathbf{y} \in \mathbb{R}^m$, $\mathbf{A} \in \mathbb{R}^{m \times n}$, λ, ϵ, K, Q
OUTPUT: $\hat{\mathbf{x}}$
require $\lambda \geq 0, \epsilon \geq 0, K \geq 0, Q \geq 0$
for $q = 0$ **to** $Q - 1$ **do**
 $\mathbf{x}_q[0] \leftarrow$ random initial condition
 $\hat{\mathbf{x}}_q \leftarrow$ Algorithm 1 ($\mathbf{y}, \mathbf{A}, \lambda, \epsilon, \mathbf{x}_q[0], K$)
end for
return $\hat{\mathbf{x}} \leftarrow \arg \min_{\hat{\mathbf{x}}_q} \tau_R^2(\mathbf{r}(\hat{\mathbf{x}}_q))$

Naturally, the value of K must be large enough to ensure sufficient convergence of the solutions, typically on the order of hundreds of iterations. Hence, Algorithm 2 may require large run times, especially when high numbers of candidate solutions, Q , are considered.

3.2. Fast algorithm

To reduce the computational cost of the basic version of the algorithm, we adopt the idea proposed in [15]. It is a two-step algorithm, as shown below.

Algorithm 3 Fast algorithm

INPUT: $\mathbf{y} \in \mathbb{R}^m$, $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\lambda, \epsilon, K, Q, M$
OUTPUT: $\hat{\mathbf{x}}$
require $\lambda \geq 0, \epsilon \geq 0, K \geq 0, Q \geq 0, M \geq 0$
for $q = 0$ **to** $Q - 1$ **do**
 $\mathbf{x}_q[0] \leftarrow$ random initial condition
 $\hat{\mathbf{x}}_q \leftarrow$ Algorithm 1 ($\mathbf{y}, \mathbf{A}, \lambda, \epsilon, \mathbf{x}_q[0], K$)
end for
 $\mathcal{Z} \leftarrow$ set of M best solutions $\hat{\mathbf{x}}_q$
for $z \in \mathcal{Z}$ **do**
 $\hat{\mathbf{x}}_z \leftarrow$ Algorithm 1 ($\mathbf{y}, \mathbf{A}, \lambda, \epsilon, \mathbf{x}_z, \infty$)
end for
return $\hat{\mathbf{x}} \leftarrow \arg \min_{\hat{\mathbf{x}}_z} \tau_R^2(\mathbf{r}(\hat{\mathbf{x}}_z))$

In the first step, a very small K is chosen, in the order of 3 to 5. Hence, this step, though identical to the basic algorithm, takes much less time. After the first step, instead of choosing just one best solution, as in Algorithm 2, M best solutions are kept as input to the second step. Note that in the second step, all the M solutions are forced to run to complete convergence by setting $K \leftarrow \infty$. From these M converged solutions, the best solution is selected as the final output.

4. NUMERICAL EXPERIMENTS

In this section, we compare the performance of the regularized τ -estimator with that of the regularized LS-estimator and regularized M-estimator. The simulation setup is as follows: we generate a matrix \mathbf{A} with random iid Gaussian entries and a condition number of 1000. We use the same piece-wise constant \mathbf{x} for all the experiments. Additive outliers and Gaussian noise are included in generating \mathbf{y} :

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}_G + \mathbf{n}_o, \quad (16)$$

where $\mathbf{n}_G \in \mathbb{R}^{m \times 1}$ has random iid Gaussian entries. On the other hand, $\mathbf{n}_o \in \mathbb{R}^{m \times 1}$ is a sparse vector. The few non-zero randomly selected entries have a random uniform value with zero mean and variance equal to 10 times that of $\mathbf{A}\mathbf{x}$. These entries represent the outliers.

With \mathbf{y} and \mathbf{A} as input, $\hat{\mathbf{x}}$ is estimated with the three different estimators. To show the efficiency and robustness of the estimators in terms of the norm of the error $\|\mathbf{e}\| = \|\hat{\mathbf{x}} - \mathbf{x}\|$, several experiments with different percentages of outliers (0%, 10%, 20%, 30%, 40%) are carried out.

For the loss function $\rho(t)$ of the M-estimator, we choose the Huber function [17], while for the τ -estimator we choose the so-called optimal weight function with clipping parameters $c_1 = 0.4046$ and $c_2 = 1.0900$. These loss functions produce a τ -estimator with BP = 0.5 and efficiency of 95% [15].

In each realization, the optimal λ was found experimentally and used in the experiment. Figure 2 shows the average of the $\|\mathbf{e}\|$ evaluated over 100 Monte-Carlo runs.

We can observe in Figure 2 that the $\|\mathbf{e}\|$ of the LS-estimator and M-estimator strongly increase already for 10% of outliers. The τ -estimator, on the other hand, provides superior robustness, even up to 40% of outliers. The LS-estimator, which coincides with the ML-estimator, provides the best performance in the outlier-free case. The τ -estimator has a high relative efficiency, producing estimates which

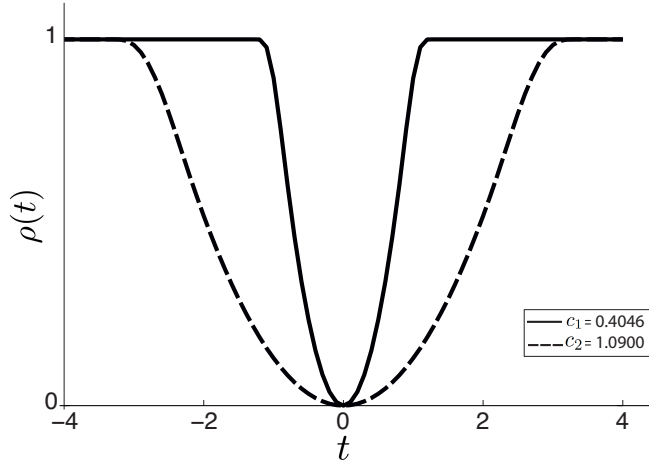


Fig. 1. 'Optimal weight' [15] loss functions $\rho_1(t)$ (solid line) and $\rho_2(t)$ (dashed line) used in the regularized- τ estimator.

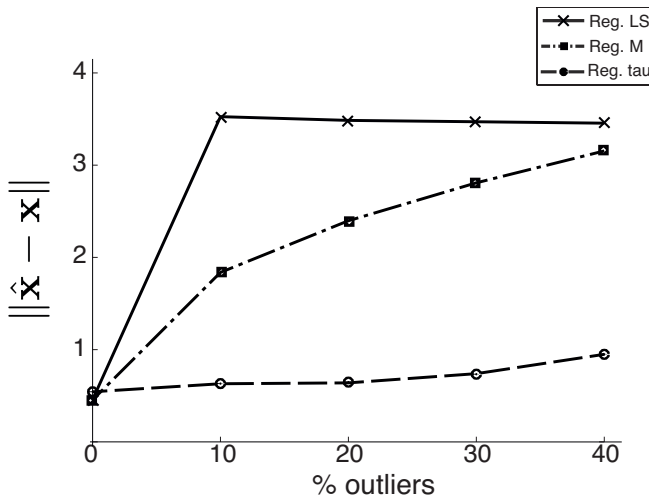


Fig. 2. Average of the $\|\hat{\mathbf{x}} - \mathbf{x}\|$ produced by the regularized LS-estimator, M-estimator, and τ -estimator evaluated over 100 Monte-Carlo runs.

are quite close to the LS-estimates.

In order to demonstrate how the regularization affects the robustness of the estimators, we run the same experiment, but using a matrix \mathbf{A} with a low condition number. Then, \mathbf{x} is estimated using the non-regularized LS-estimator, M-estimator, and τ -estimator, respectively.

First, we note that the errors in the non-regularized case are, in general, one order of magnitude larger than in the regularized case. This is due to the regularization limiting the maximum error: we chose the optimum regularization parameter. The maximum error is thus obtained when the parameter is large enough to make $\hat{\mathbf{x}} = \mathbf{0}$. Thus, the maximum $\|\mathbf{e}\| = \|\mathbf{x}\| = 3.6$ for the particular source \mathbf{x} we chose. This intuitively represents the BP of the regularized

estimator.

In the regularized case, the LS saturates to this maximum $\|\mathbf{e}\|$, while the M-estimator remains under this bound. Its performance is much better than in the non-regularized case. The τ -estimator performs well up to 40% of outliers in the regularized case, but in the non-regularized case it is not robust for 40% of outliers.

From these results, we can conclude that the regularization improves the robustness of the estimation. It includes in the problem more a priori information about the solution, which permits the estimator to compute a solution closer to the real one, and to discard unlikely solutions which cause a breakdown.

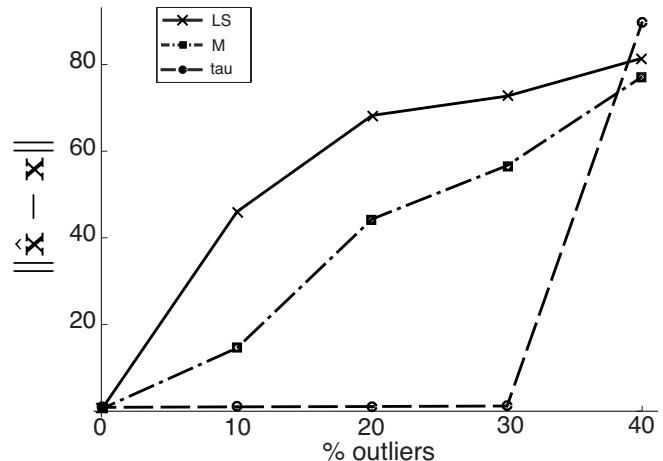


Fig. 3. Average of the $\|\hat{\mathbf{x}} - \mathbf{x}\|$ produced by the non-regularized LS-estimator, M-estimator, and τ -estimator evaluated over 200 Monte-Carlo runs.

5. CONCLUSIONS

In this paper, we proposed a new regularized robust estimator that is simultaneously highly robust against outliers, highly efficient in the presence of purely Gaussian noise, and also stable when the mixing matrix has a large condition number. A large part of this paper was dedicated to deriving a fast way of computing this estimator based on a regularized iterative reweighted least squares algorithm. We showed that our estimator provides superior robustness, compared to existing regularized estimators, and approximately maintains a performance quite close to the optimal maximum likelihood estimator in the outlier-free case, even up to 40% of outliers. The proposed estimator seems well suited to solve inverse transport modelling of atmospheric emissions, where the mixing matrix may also contain outliers. This real data application will be part of our future work. Furthermore, fundamental robustness concepts, such as the breakdown point, the maximum bias curve and the influence function are influenced by the regularization. We showed via numerical experiments that a breakdown occurs when the norm of the errors converges to a fixed value that depends on the regularization. A full statistical robustness analysis of the proposed estimator will provide new insights on the interrelation of robustness and regularization.

6. REFERENCES

- [1] A. Ribes and F. Schmitt, "Linear inverse problems in imaging," *IEEE Signal Process. Mag.*, pp. 84–99, July 2008.
- [2] R. H. Stolt and A. B. Weglein, *Seismic imaging and inversion*, Cambridge, 2012.
- [3] M. Martinez-Camara, I. Dokmanic, J. Ranieri, R. Scheibler, M. Vetterli, and A. Stohl, "The Fukushima inverse problem," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 4330–4334.
- [4] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection*, John Wiley & Sons, Ltd, 1987.
- [5] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: a tutorial-style treatment of fundamental concepts," *IEEE Signal Process. Mag.*, vol. 29, no. 4, pp. 61–80, July 2012.
- [6] P. J. Huber, "Robust estimation of a location parameter," *Ann. Math. Statist.*, vol. 35, no. 1, pp. 73–101, 1964.
- [7] P. J. Rousseeuw, "Least median of squares regression," *J. Am. Statist. Assoc.*, vol. 79, pp. 871–880, 1984.
- [8] P. J. Rousseeuw and V. J. Yohai, "Robust regression by means of S-estimators," *Lect. Notes Stat.*, vol. 26, pp. 256–272, 1984.
- [9] V. J. Yohai and R.H. Zamar, "High breakdown-point estimates of regression by means of the minimization of an efficient scale," *J. Amer. Statist. Assoc.*, vol. 83, no. 402, pp. 406–413, 1988.
- [10] A.N. Tikhonov and V.Y. Arsenin, *Solution of ill-posed problems*, Winston-Wiley, New York, 1977.
- [11] M. Martinez-Camara, B. Béjar Haro, A. Stohl, and M. Vetterli, "A robust method for inverse transport modelling of atmospheric emissions using blind outlier detection," *Geosci. Model Dev. Discuss.*, vol. 7, no. 3, pp. 3193–3217, 2014.
- [12] M. J. Silavapulle, "Robust ridge regression based on an M-estimator," *Aust. J. Stat.*, vol. 33, no. 3, pp. 319–333, 1991.
- [13] K. Tharmaratnam, G. Claeskens, C. Croux, and M. Salibian-Barrera, "S-estimation for penalised regression splines," *J. Comput. Graph. Stat.*, vol. 19, pp. 609–625, 2010.
- [14] R. A. Maronna, "Robust ridge regression for high-dimensional data," *Technometrics*, vol. 53, no. 1, pp. 44–53, 2011.
- [15] M. Salibian-Barrera, G. Willems, and R.H. Zamar, "The fast-tau estimator for regression," *J. Comput. Graph. Stat.*, vol. 17, pp. 659–682, 2008.
- [16] R. Wolke and H. Schwetlick, "Iteratively reweighted least squares: Algorithms, convergence analysis, and numerical comparisons," *SIAM J. Sci. and Stat. Comput.*, vol. 9, no. 5, pp. 907–921, 1988.
- [17] P. J. Huber and E. M. Rochetti, *Robust Statistics*, vol. 2, John Wiley & Sons, 2009.