Robustness of ℓ_1-Norm Estimation: From Folklore to Fact

Salvador A. Flores

***Abstract*—The advantages of using ℓ_1-norm rather than ℓ_2-norm in terms of robustness for signal processing and other data analysis procedures are largely recognized across the scientific literature. However, from the robust statistics viewpoint, at least that based in the concept of breakdown point, ℓ_1-norm regression has no better resistance to outliers than least squares, and it is believed that it degrades in higher dimension. We explain this seeming contradiction between theory and practice by the different contamination models used to assess robustness to outliers. After a brief review of the existing notions of robustness, we adopt a model where carriers are not subject to contamination, and only the response variable can be contaminated with outliers. We prove two new positive results concerning breakdown point robustness of ℓ_1-norm regression under this model. First, we show that ℓ_1-norm regression can have a positive breakdown point in any dimension, and this is rather common. We elaborate further in a second result, showing that random designs with unit normal rows yield to a high breakdown point, around 30% for moderate dimension growing asymptotically to 50%, with very large probability. These results provide a theoretical support to the practical success of ℓ_1-norm based procedures and are, at the same time, consistent with statistical robust regression theory.**

***Index Terms*—Robust regression, Outliers, ℓ_1 norm minimization, Breakdown point**

I. Introduction

DESPITE the formidable body of works providing empirical evidence on the robustness properties of ℓ_1-norm regression, such as Least Absolute Deviations (LAD) or Huber's regression, its resistance to outlying observations has taken a long time to be firmly established, and some aspects remain unclear or misunderstood. The practitioner looking for theoretical results supporting the robustness of ℓ_1-norm based methods will find very few positive results, and a look into the robust regression literature leaves the impression that the only way to perform outlier-robust data analysis is by using the computationally intensive robust estimators. The reasons for this are manifold. The lack of a closed-form solution to the estimation problem and the non-differentiability of the ℓ_1-norm complicated for a long time both the computation of the estimator and the analysis of its statistical properties, in particular, those related to breakdown in non-trivial scenarios. In addition, the introduction of high-breakdown point estimators and the development of its associated theory could have discouraged the study of robustness properties of classical M-estimators, often presented without distinction as examples of non-robust estimators.

The advent of interior point methods has made the computation of the ℓ_1 regression estimator not only feasible, but even faster than Least Squares (LS) for very large datasets [1]. Also, several results [2]–[8] have shed light into its fine robustness properties. A widespread measure of robustness to outliers is the *Finite-sample Breakdown Point* (FBP), which is the minimum fraction of data that, if replaced by arbitrary contamination, can drive the estimation out of any bound. It is well known that if contamination by high-leverage outliers is allowed in the explanatory variables, the FBP of the LS and LAD estimators is 0% [9], [10] . Nonetheless, the robustness of ℓ_1 norm estimators in most practical situations, where the presence of high-leverage outliers is not the main concern, such as signal and image basis or sensor arrays, is not equally simple to depict.

The characterization of the FBP for ℓ_1 regression is difficult, of combinatorial nature [2], [3], and depends on the design of the explicative variables, also called carriers. For this reason, attempts to obtain general conclusions on its FBP have not been conclusive. Moreover, a wrong interpretation of results in [11], stating that random designs from spherical distributions have a breakdown point decreasing as a function of the dimension of the carriers, led [3] to conclude that the FBP of LAD regression is at most 25%, which is the maximum FBP for univariate regression. On the other hand, [12] showed that for two-way contingency tables, LAD attains the maximum possible FBP among regression equivariant estimators, and [13] obtained theoretical bounds in the range $16\% - 20\%$ for the number of outliers identifiable by LAD estimation in a source localization problem, and showed numerically that the number of outliers that can be identified for particular configurations yielding well equilibrated designs is often higher.

The question whether a high FBP, close to 50%, is attainable at all by ℓ_1 norm regression is open. Subsidiary questions such as the dependence of the FBP on the dimension and which designs can attain such a high FBP are open as well.

We bridge this gap by formally showing that well balanced designs yield to very robust LAD regression, as expected, in any dimension. We show that a fraction of about 30% of vertical outliers can be tolerated in the most favorable situations for moderate dimension, and this fraction approaches 50% asymptotically. Designs obtained by drawing the explicative variables uniformly on the sphere achieve this FBP with very high probability.

In this way, we identify the precise sense in which the assertion "ℓ_1 norm is more robust than ordinary least squares" is true, and provide a formal proof of it.

A. notations

We shall use the notation $N = \{1, ..., n\}$ for the index set of all the observations. For a set of indexes M, $|M|$ denotes its

S. Flores is with the Mathematics Engineering Department and the Center for Mathematical Modelling, Universidad de Chile, Chile. e-mail: sflores@dim.uchile.cl

cardinality. For a subset M of N and $p \in [1, +\infty[$, we define $\|\cdot\|_{p,M} \colon \mathbf{x} \mapsto (\sum_{i \in M} |x_i|^p)^{1/p}$ and $\|\mathbf{x}\|_p = \|\mathbf{x}\|_{p,N}$. For a vector $\mathbf{x} \in \mathbb{R}^n$, we denote by $\operatorname{supp}(\mathbf{x})$ its support, *i.e.*, the index set of nonzero components, $\operatorname{supp}(\mathbf{x}) = \{i \in N \mid x_i \neq 0\}$. The so-called "$\ell_0$-norm" is defined as $\|\cdot\|_0 := |\operatorname{supp}(\cdot)|$.

II. The Robust Linear Regression Model

Let us consider the archetypal linear regression model, where each observation y_i follows

$$y_i = \mathbf{x}_i^\top \beta + \delta_i, \qquad i = 1, ..., n \tag{1}$$

for some parameter $\beta \in \mathbb{R}^p$, where the $\mathbf{x}_i \in \mathbb{R}^p$ are the corresponding vector of independent variables or *carriers* and δ_i are random deviations from the model, assumed to be independent identically distributed (i.i.d) with zero mean and finite variance. M-estimators are defined as solutions, for different loss functions $\rho(\cdot)$, to the minimization problem:

$$\underset{g \in \mathbb{R}^p}{\text{minimize}} \qquad \sum_{i=1}^{n} \rho(r_i(g)) \tag{2}$$

where $r_i(g) := y_i - \mathbf{x}_i^\top g$ are the components of the residuals vector $\mathbf{r}$. If the errors follow a normal distribution $\delta_i \sim N(0, \sigma^2)$, then the best linear unbiased estimator is the popular LS estimator $\hat{\beta}_2$, defined by (2) with $\rho(t) = t^2/2$. Huber's criterion, defined for a given estimate of the noise magnitude $\sigma > 0$ as

$$\rho_H(r) = \begin{cases} \frac{1}{2} r^2 & \text{if } |r| \leq \sigma \\ \sigma |r| - \frac{1}{2}\sigma^2 & \text{if } |r| > \sigma \end{cases} \tag{3}$$

combines ℓ_1-norm and ℓ_2-norm in order to behave like LS for Gaussian noise and like ℓ_1-norm face to outliers. Several generalizations combining ℓ_1-norm with other estimators are studied in detail in [14].

The robustness of M-estimators depends both on the function ρ and the type of contamination that we allow in the model. On the extremes, we have robust estimators [10], defined by non-convex ρ functions, which allow arbitrary contamination both in the carriers $\mathbf{x}_i$ and the responses y_i, from one side, and non-robust estimators, such as LS, which can be strongly affected by any type of contamination, either in $\mathbf{x}_i$ or in y_i, on the other side. In the middle, there is a class of ℓ_1-norm based M-estimators which are robust face to impulsive noise in y, but lose their robustness when there are outliers in the $\mathbf{x}$-variables. They have the great advantage of being defined by convex (although not strictly convex) ρ functions, making their computation affordable for large scale data and/or for real-time estimation.

In this paper, we consider the Robust Linear Regression Model (RLRM),

$$y_i = \mathbf{x}_i^\top \beta + \delta_i + e_i, \tag{4}$$

where δ is a dense vector with i.i.d components modeling background noise in y, and e is a sparse vector used to model outliers in the dependent variable y, called vertical outliers.

Our results concern this model, where impulsive noise affects the dependent variable y and the independent variables are under control or can be reliably measured, which is the case for instance in signal processing applications where the design matrix X represents a dictionary or basis, such as Fourier or Wavelets, or models a physical configuration of antennas or sensors. This is the scenario more often encountered in practice, and where the ℓ_1 estimator has forged its reputation as a robust alternative to ordinary LS. Extensions of our results to the error-in-variables model, which considers the possibility to have dense, bounded error in the independent variables $\mathbf{x}$ are possible through the robust counterparts of ℓ_1 and Huber estimators studied in [15] (see also [16]).

The RLRM has numerous applications in signal processing. [17] combines continuous wavelet transformation with linear regression for artifact removal in electroencephalogram signals. In localization problems, such as GPS systems, the measured position can be considerably corrupted when the signals are reflected by the environment, creating a multi-path phenomenon or jamming. Much in the same vein of this work, [13] obtained bounds on the number of outliers that a ℓ_1-norm based method for source localization can detect. In that case the design matrix X is related to the configuration of the sensor network. Similarly, [18] introduces a robust ℓ_1 beamformer for estimating the direction-of-arrival in a phased array with antenna switching. In [19], a mixed $\ell_1 - \ell_2$ norm minimization approach is used to cope with impulsive noise in an adaptive filter for system identification. This is extended in [20] to sparse signal estimation via continuous mixed norm. In image processing, [21] uses model (4) combined with an ℓ_0 regularization in a background/foreground separation algorithm for image segmentation.

III. Measuring robustness

Before showing our results, we are going to review and update the robustness properties of the LAD estimator $\hat{\beta}_1$, defined through

$$\|y - X\hat{\beta}_1\|_1 = \min_{g \in \mathbb{R}^p} \|y - Xg\|_1, \tag{5}$$

where X denotes the *design matrix*, whose rows are the carriers $\mathbf{x}_1, ..., \mathbf{x}_n$. All the results in this paper also apply to Huber's M-estimator, which is a denoised version of LAD estimation [8], and to most of the mixed $\ell_p - \ell_1$ norm estimation methods [14].

A. Leverage and Breakdown Point

Robustness measures can be roughly divided into finite sample ones, such as the finite sample breakdown point or signal-to-noise ratio (SNR), and those based on distributions, or asymptotic. The focus of this paper is on the former ones, since their practical use does not need to have an a-priori on the distributions of the data or the errors, enlarging their applicability. Also, their ease of interpretation is greater.

For a given estimator β, we define the finite sample breakdown order as :

$$\theta_f(\beta, n) := \max\{|M| : \tilde{\beta} \text{ is bounded for any } e \in S_M\}, \tag{6}$$

where $\tilde{\beta}$ denotes the estimate obtained from data contaminated as in (4) with background noise and outliers, and $S_M = \{u \mid$

$\operatorname{supp}(u) = M\}$ is the set of arbitrary data with support on M. The FBP $m_r(\beta, n)$ and the asymptotic FBP $\varepsilon_r(\beta, n)$ of an estimator β are defined respectively as

$$m_r(\beta, n) := \frac{\theta(\beta) + 1}{n} \quad \text{and } \varepsilon_r(\beta) = \lim_{n\to\infty} \theta(\beta, n). \tag{7}$$

The breakdown properties of the ℓ_1 estimator under the RLRM have been established by [2] and [3]. The finer analysis presented here and the extensions to Huber M-estimator can be found in [8]. There are two aspects to consider when dealing with outliers. Firstly, breakdown, when there are many large outliers, and bias, when the outliers are not many enough to trigger breakdown, they can nonetheless have a large impact on the bias of the estimation. We are going to see that for LAD regression these two aspects are closely related.

For a $n \times p$ matrix X, define for every $k \in \{1, \ldots, n\}$ the *leverage constants* c_k of X as

$$c_k(X) = \min_{\substack{M\subset N \\ |M|=k}} \min_{\substack{g\in\mathbb{R}^p \\ \|g\|_2=1}} \frac{\sum_{i\in N\setminus M} |\mathbf{x}_i^\top g|}{\sum_{i\in N} |\mathbf{x}_i^\top g|} \tag{8}$$

and

$$m(X) = \max\Big\{k \in N \;\Big|\; c_k(X) > \frac{1}{2}\Big\}. \tag{9}$$

The constants $c_k(X)$ and $m(X)$ determine the robustness of LAD regression as follows [3], [5], [8].

Theorem 1: Let $y = X\beta + \delta + e$ and $M = \operatorname{supp}(e)$ satisfying $|M| = k \leq m(X)$. Then the following hold for the ℓ_1 estimator $\hat{\beta}_1$,

$$\|X(\hat{\beta}_1 - \beta)\|_1 \leq \frac{2}{2c_k - 1} \|\delta\|_{1,N\setminus M}. \tag{10}$$

When the number of outliers k exceeds $m(X)$, the ℓ_1 estimator may breakdown, and the bound (10) is not longer valid.

Note that $X\hat{\beta}_1$ represents the signal reconstructed from noisy signal measurements contaminated with outliers, by ℓ_1-norm regression, while $X\beta$ is the true, noiseless, signal. From (10) we see that in the signal reconstruction error bound the noise is amplified by a factor $(c_k - 1/2)^{-1}$. Therefore, even if there is no breakdown, outliers can deteriorate the SNR of the reconstructed signal, specially if $c_k(X)$ is close to $1/2$.

Intuitively, for well equilibrated carriers the quantity $c_k(X)$ should be proportional to $(n-k)/n$. In this way, the condition $k \leq m(X)$ boils down to "there is no breakdown if outliers are minority". In order to leave aside the trivial scenario of extreme leverage points ($\|\mathbf{x}_i\| \to \infty$ for some i), let us suppose that carriers are normalized in such a way that $\|\mathbf{x}_i\| = 1$ for every i. In this situation, no individual observation can exert an unduly influence on the fit. The only possibility for $c_k(X)$ to be "oversized" is to have a group of highly correlated carriers, indexed by M, in such a way that it is possible to find a direction $\bar{g}$ making $\sum_{N\setminus M} |\mathbf{x}_i^\top \bar{g}|$ comparable to $\sum_N |\mathbf{x}_i^\top \bar{g}|$. In Figure 1 we illustrate this situation in dimension two. Plots a) and b) show the same uniformly distributed points on the disk and two dashed arc segments of the same length. It is evident that the fraction of points covered in the dashed arc is proportional to the arc length, and this is independent of the orientation of the dashed arc. Plot c) shows a situation that can

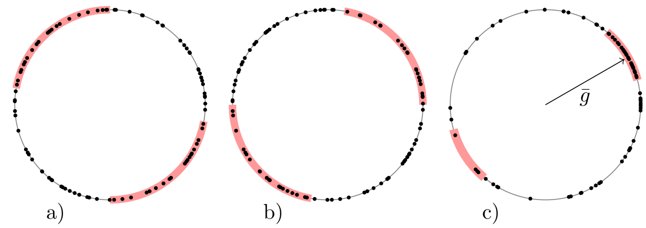

Fig. 1. Illustration of the rationale under which well balanced designs, a) and b), promote robustness in LAD linear regression, while groups of aligned carriers, in c), lead to low breakdown point.

lead to low FBP. A group of observations concentrate around direction $\bar{g}$, gaining an increased leverage.

Note that this result specializes in a sense the upper bound for the FBP of regression equivariant estimators given by $(n - \kappa(X) + 1)/2$, where $\kappa(X)$ denotes the maximum number of carriers lying in a $(p-1)$-dimensional subspace of $\mathbb{R}^p$ [4]. In the same way as the upper bound for the breakdown point increases as predictors do not concentrate around lower dimensional subspaces, the actual breakdown point of LAD improves as far as predictors do not concentrate around a narrow solid angle in $\mathbb{R}^p$.

B. Other Robustness Measures

As mentioned before, there exists another approach to measuring robustness, based on probability distributions. If we consider that the carriers $\mathbf{x}$ are a random p-variate and y a real random variable with joint distribution P on $\mathbb{R}^{p+1}$, then the data $\mathbf{x}_i, y_i$ can be thought of as finite samples from a contaminated distribution $(1-\varepsilon)P + \varepsilon Q$. In this way, it is possible to define the Gross Error Breakdown Point (GEBP) ε_G at a distribution P as the supremum over all $\varepsilon \in [0, 1]$ such that $\beta((1-\varepsilon)P + \varepsilon Q)$ remains bounded when the distribution Q ranges over a predefined class $\mathcal{Q}$ of contaminating distributions. [11] studied the GEBP of regression estimators, and showed that for most M-estimators, including least squares, the GEBP is $\varepsilon_G = 0$ if the contaminating distributions $\mathcal{Q}$ are point masses at arbitrary points $(\mathbf{x}_0, y_0)$, independently of P. They also showed that, if P is a spherically symmetric distribution and $\mathcal{Q}$ are point masses at arbitrary points $(\mathbf{x}_0, y_0)$, then the best GEBP that an ℓ_1-type estimator can attain behaves like $(2p)^{-1/2}$ for large p. Nothing can be deduced about the FBP of a given estimator from these results, since in general both quantities do not coincide, for any combination of P and $\mathcal{Q}$. Other possible exist in this category; see [22] for a discussion on the relation of the FBP to other robustness measures in this family.

IV. Breakdown Point of ℓ_1 Regression Without Leverage Points

In this section we go deeper into the study the robustness properties of the ℓ_1 regression estimator in the RLRM of Section II. We address the open issues on the subject, namely, the behavior of its FBP as the dimension p varies, and how robust is it in the most favourable scenario, when the design has no high-leverage points.

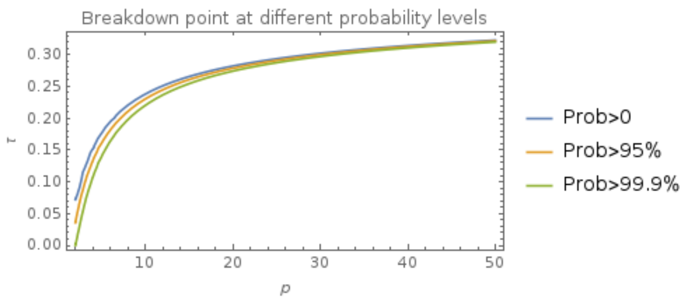

Fig. 2. Evolution of the FBP of LAD regression as p grows . The upper curve (in blue) shows the τ such that (11) holds. The second curve (in orange) and the bottom curve (in green) show the FBP with probability 95% and 99.9% respectively, with $n = 5p$.

Both questions are settled favorably to the ℓ_1 estimator. The first result, rather theoretical, shows that in any dimension and for any sample size, as far a $n > p$, there are "many" designs where the ℓ_1 estimator has a positive FBP.

Theorem 2: Let p and n be any natural numbers with $p < n$. There exists a set Ξ of $n \times p$ matrices with Grassmanian measure bounded below by a positive constant independent of n and p such that for any design matrix in Ξ, the ℓ_1 regression estimator $\hat{\beta}_1$ has breakdown order $\theta(\hat{\beta}_1, n) = k$ whenever

$$\frac{k}{n} < \alpha\phi\left(\frac{n-p}{n}\right)$$

for an absolute constant $\alpha > 0$ independent of n and p, where $\phi(t) = t/(1 - \log(t))$.

Remark 1: The function ϕ is monotone and satisfies $0 < \phi(t) \leq 1$ for $0 < t \leq 1$, $\lim_{t\to 0}\phi(t) = 0$ and $\phi(1) = 1$. In particular, $\liminf_{n\to\infty,p\to\infty} \phi((n-p)/n) > 0$ whenever $\liminf_{n\to\infty,p\to\infty} p/n < 1$.

See [23] for a proof. This result is essentially a re-writing in our setting of the results reported by [24]–[26] in the compressed sensing context. See Appendix B (in supplementary materials) for more on this connection between robust regression and compressed sensing. We do not exploit further this connection since the results based on restricted isometry conditions, which are necessary conditions, cannot approach the highest FBP of 50% that we are looking for.

Theorem 2 disproves a wrong belief in robust regression that the ℓ_1 estimator has a FBP going to 0 as p increases, even without leverage points [3], [27] . This belief originated in a result by [11] showing that if the rows of X are sampled from a spherically symmetric distribution, then the best GEBP that ℓ_1-norm M-estimators can attain behaves like $(2p)^{-1/2}$ for large p. As discussed in the previous section, the GEBP does not coincide in general with the FBP, even for finite samples. Furthermore, even if both definitions happened to coincide for some particular model, that result considers contamination with outliers both in $\mathbf{x}$ and y.

The next result gives a formal statement of the intuitive arguments explained in Figure 1 to support the case for the robustness of LAD regression for fixed designs with no leverage points.

Theorem 3: Let the rows $\mathbf{x}_1, ..., \mathbf{x}_n$ of the $n \times p$ matrix X be iid from the uniform distribution on the unit sphere $\mathbb{S}^{p-1}$. Let $0 < \tau < 1/2$ be such that

$$\frac{\tau\ln(e/\tau)}{\ln(p)} + \tau\alpha_p < 1/2, \tag{11}$$

where $\alpha_p = \left(\frac{1+p\ln(p)}{p\ln(p)}\right)$. The LAD regression estimator has a finite sample breakdown point greater or equal than τ with probability at least

$$1 - p^{n(\tau\ln(e/\tau)/\ln(p)+\tau\alpha_p-1/2)}. \tag{12}$$

The proof is provided in the supplementary materials, Appendix A. In Figure 2 we show, for each p, the greatest τ satisfying (11). For any τ under the curve, there is a positive probability, depending both in p and n, of having a breakdown point at least τ for LAD regression. The curve reaches very quickly a value around 30% and then keeps growing at a very small rate to approach its asymptotic value of 50%.

Note that the sample size n does not play any role in determining the breakdown point (11), but only in the probability of achieving it. The following result on the asymptotic FBP is a direct consequence of this fact.

Corollary 1: Under the conditions of Theorem 3, asymptotically as $n \to \infty$, the ℓ_1 regression estimator has a FBP greater or equal than any τ satisfying (11), with probability one.

V. Conclusion

We have presented two results substantiating the ubiquitous evidence of the robustness of LAD regression in practical contexts. These results are also of importance for robust regression, since most robust estimators need an initial estimate of scale, which is often provided by LAD regression. Our detailed treatment of the issue aims to clarify the existing results, which seem sometimes contradictory. The main conclusion is that LAD regression can be highly robust when there are not high-leverage points, but is quite sensitive to this kind of contamination. For these reasons, and taking into account the computational advantages over robust estimation, LAD regression is a competitive robust inference method under controlled experimental conditions and for large data, while high-breakdown point estimators might be preferred for small to mid-sized datasets, if there is not certainty about the experimental conditions. Besides its theoretical interest, our results have an impact in each of the applications mentioned in Section II, since they provide a lot of intuition on the factors of the design that influence the most the resistance of LAD regression to outliers.

Acknowledgment

The author gratefully acknowledges the support of Conicyt through grants Fondecyt Iniciacion 11150773 and AFB 170001.

REFERENCES

[1] S. Portnoy and R. Koenker, "The gaussian hare and the laplacian tortoise: computability of squared-error versus absolute-error estimators," *Statist. Sci.*, vol. 12, no. 4, pp. 279–300, 11 1997.

[2] X. He, J. Jurečková, R. Koenker, and S. Portnoy, "Tail behavior of regression estimators and their breakdown points," *Econometrica*, vol. 58, no. 5, pp. 1195–1214, 1990.

[3] S. P. Ellis and S. Morgenthaler, "Leverage and breakdown in L_1 regression," *J. Amer. Statist. Assoc.*, vol. 87, no. 417, pp. 143–148, 1992.

[4] C. H. Müller, "Breakdown points for designed experiments," *Journal of Statistical Planning and Inference*, vol. 45, no. 3, pp. 413 – 427, 1995.

[5] I. Mizera and C. Müller, "The influence of the design on the breakdown points of ℓ_1-type M-estimators," in *MODA6 – Advances in Model-Oriented Design and Analysis*, ser. Contributions to Statistics, A. Atkinson, P. Hackl, and W. Müller, Eds. Physica-Verlag, 2001, pp. 193–200.

[6] A. Giloni and M. Padberg, "The finite sample breakdown point of ℓ_1-regression," *SIAM J. Optim.*, no. 14, pp. 1028–1042, 2004.

[7] Y. Sharon, J. Wright, and Y. Ma, "Minimum sum of distances estimator: Robustness and stability," in *2009 American Control Conference*, 2009, pp. 524–530.

[8] S. Flores, "Sharp non-asymptotic performance bounds for ℓ_1 and Huber robust regression estimators," *Test*, vol. 24, no. 4, pp. 796 – 812, 2015.

[9] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection*. John Wiley & Sons, 1987.

[10] R. A. Maronna, R. D. Martin, and V. J. Yohai, *Robust statistics*, ser. Wiley Series in Probability and Statistics. John Wiley & Sons, 2006.

[11] R. Maronna, O. Bustos, and V. Yohai, "Bias- and efficiency-robustness of general M-estimators for regression with random carriers," in *Smoothing Techniques for Curve Estimation*, ser. Lecture Notes in Mathematics, T. Gasser and M. Rosenblatt, Eds. Springer Berlin / Heidelberg, 1979, vol. 757, pp. 91–116.

[12] M. Hubert, "The breakdown value of the L_1 estimator in contingency tables," *Statistics & Probability Letters*, vol. 33, no. 4, pp. 419 – 425, 1997.

[13] J. Picard and A. Weiss, "Bounds on the number of identifiable outliers in source localization by linear programming," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2884–2895, 2010.

[14] Y. Dodge, J. Jureckova, and J. Jurečková, *Adaptive regression*. Springer Science & Business Media, 2000.

[15] M. Pinar, "Linear Huber M-estimator under ellipsoidal data uncertainty," *BIT Numerical Mathematics*, vol. 42, no. 4, pp. 856 – 866, 2002.

[16] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms and Engineering Applications*, ser. MPS-SIAM Series on Optimization. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 2001.

[17] C. Dora and P. K. Biswal, "Robust ECG artifact removal from EEG using continuous wavelet transformation and linear regression," in *2016 International Conference on Signal Processing and Communications (SPCOM)*, June 2016, pp. 1–5.

[18] M. Lee and Y. Kim, "Robust l_1-norm beamforming for phased array with antenna switching," *IEEE Communications Letters*, vol. 12, no. 8, pp. 566–568, Aug 2008.

[19] J. Chambers and A. Avlonitis, "A robust mixed-norm adaptive filter algorithm," *IEEE Signal Processing Letters*, vol. 4, no. 2, pp. 46–48, Feb 1997.

[20] A. Javaheri, H. Zayyani, M. A. T. Figueiredo, and F. Marvasti, "Robust sparse recovery in impulsive noise via continuous mixed norm," *IEEE Signal Processing Letters*, vol. 25, no. 8, pp. 1146–1150, Aug 2018.

[21] G. Xue, L. Song, J. Sun, and M. Wu, "Foreground estimation based on robust linear regression model," in *2011 18th IEEE International Conference on Image Processing*, Sept 2011, pp. 3269–3272.

[22] D. Pastor and F. Socheleau, "Robust estimation of noise standard deviation in presence of signals with unknown distributions and occurrences," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1545–1555, April 2012.

[23] N. Linial and I. Novik, "How neighborly can a centrally symmetric polytope be?" *Discrete & Computational Geometry*, vol. 36, no. 2, pp. 273–281, 2006.

[24] W. Yin and Y. Zhang, "Extracting salient features from less data via ℓ_1-minimization," *SIAG/OPT Newsletter: Views & News*, vol. 19, no. 1, pp. 11–19, 2008.

[25] Y. Zhang, "A simple proof for recoverability of ℓ_1 minimization: Go over or under?" *Rice University CAAM technical report*, vol. TR05-09, 2005.

[26] ——, "On theory of compressive sensing via ℓ_1-minimization: Simple derivations and extensions," *Tech. Rep. TR08-11, Dep. Comp. Appl. Math., Rice Univ*, 2008.

[27] R. A. Maronna and V. J. Yohai, "Robust regression with both continuous and categorical predictors," *Journal of Statistical Planning and Inference*, vol. 89, no. 1, pp. 197–214, 2000.