Carrillo *et al. EURASIP Journal on Advances in Signal Processing* (2016) 2016:108
DOI 10.1186/s13634-016-0404-5

EURASIP Journal on Advances in Signal Processing

REVIEW Open Access

Robust compressive sensing of sparse signals: a review

Rafael E. Carrillo[1*], Ana B. Ramirez[2], Gonzalo R. Arce[3], Kenneth E. Barner[3] and Brian M. Sadler[4]

Abstract

Compressive sensing generally relies on the ℓ_2 norm for data fidelity, whereas in many applications, robust estimators are needed. Among the scenarios in which robust performance is required, applications where the sampling process is performed in the presence of impulsive noise, i.e., measurements are corrupted by outliers, are of particular importance. This article overviews robust nonlinear reconstruction strategies for sparse signals based on replacing the commonly used ℓ_2 norm by M-estimators as data fidelity functions. The derived methods outperform existing compressed sensing techniques in impulsive environments, while achieving good performance in light-tailed environments, thus offering a robust framework for CS.

Keywords: Compressed sensing, Sampling methods, Signal reconstruction, Impulsive noise, Nonlinear estimation

1 Introduction

The theory of compressive sensing (CS) introduces a signal acquisition and reconstruction framework that goes beyond the traditional Nyquist sampling paradigm [1–4]. The fundamental premise in CS is that certain classes of signals, such as natural images, have a succinct representation in terms of a sparsity inducing basis, or frame, such that only a few coefficients are significant and the remaining coefficients are negligibly small. In such cases, the signal is acquired taking a few linear measurements and subsequently accurately recovered using nonlinear iterative algorithms [4, 5]. CS has proven particularly effective in imaging applications due to the inherent sparsity, e.g., in medical imaging [6], astronomical imaging [7], radar imaging [8], and hyperspectral imaging [9].

Since noise is always present in practical acquisition systems, a range of different algorithms and methods have been proposed in the literature that enable accurate reconstruction of sparse signals from noisy compressive measurements using the ℓ_2 norm as the metric for the residual error (see [10] for a review of CS recovery algorithms). However, it is well known that least squares-based estimators are highly sensitive to outliers present in the measurement vector, leading to a poor performance when the noise does not follow the Gaussian assumption and is, instead, better characterized by heavier-than-Gaussian-tailed distributions [11–14]. A broad spectrum of applications exists in which such processes emerge, including wireless and power line communications, teletraffic, hydrology, geology, atmospheric noise compensation, economics, and image and video processing (see [14–16] and references therein).

As a motivating example, consider a CS system for wireless body area networks (WBAN). WBAN allows the transition from centralized health care services to ubiquitous and pervasive health monitoring in everyday life. Typical signals that are monitored by WBAN are electrocardiogram (ECG) signals, and CS is a promising framework to lower WBAN's energy consumption. However, ECG signals are typically corrupted by electromyographic noise which shows an impulsive behavior. Another application of interest is a nonintrusive load monitoring system that identifies house appliances and their energy consumption. A CS system can be used to acquire the power signal and then a sparse classification system used to classify the house appliances. However, the power signals exhibit impulsive behavior due to the switching nature of the appliances. If the compressive sampling process has infinite or even very large variance, the reconstructed signal obtained utilizing traditional approaches is far from the desired original signal. Thus, there are clear motivations

*Correspondence: rafael.carrillo@epfl.ch
[1]Signal Processing Laboratory (LTS5), Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland
Full list of author information is available at the end of the article

Springer Open

© 2016 The Author(s). **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

for developing robust CS techniques that address these challenging environments.

The need to describe impulsive data, coupled with computational advances that enable efficient processing methods based on models more complex than the traditional Gaussian distribution has thus led to the interest in heavy-tailed models. Robust statistics, more specifically, the stability theory of statistical procedures, systematically investigates the effects of deviation from modeling assumptions [17–19]. Maximum likelihood (ML) type estimators, also known as M-estimators, developed in the theory of robust statistics are of great importance in robust signal processing techniques [14, 16]. M-estimators are described by a cost function-defined optimization problem where properties of the cost function (or its first derivative, the so-called influence function) determine the estimator robustness [18].

The key idea in M-estimation is that the cost function, or the influence function, can be chosen in such a way to provide the estimator desirable properties (in terms of bias and efficiency) when the data are truly generated from the assumed model, and reliable albeit not optimal behavior when the data are generated from another model that is, in some sense, close to the assumed model.

Over the past decade, there have been several works addressing the reconstruction of sparse signals whose measurements are corrupted by outliers or by impulsive noise [20–55]. Parametric approaches that model the corrupting noise as a linear combination of a sparse vector of outliers (possibly gross errors) and a dense vector of small bounded noise have been proposed in the literature. Popilka et al., [21] were the first to analyze this model and proposed a reconstruction strategy that estimate first the sparse error pattern, and then estimate the true signal, in an iterative process. Related approaches are studied in [23–32]. These works assume a sparse error and estimate both signal and error at the same stage using a modified ℓ_1 minimization problem. This approach was originally proposed by Wright et al. for the face recognition problem with image occlusions [22]. A similar model was proposed by Candès et al. in [33] for the recovery of low rank matrices corrupted by outliers.

Sparse models coupled with sparse reconstruction algorithms have also been used to address the robust regression problem where the number of measurements (observations) is greater than the number of unknowns (explanatory variables) [34–38]. In the context of error correction coding, Candès et al. investigated ℓ_1 optimization approaches to solve the decoding problem when the received codeword (measurements) is assumed to be corrupted by gross outliers [39, 40].

Approaches based on M-estimators that replace the ℓ_2 data fidelity term by a more robust cost function have also been proposed. Carrillo et al. propose reconstruction approaches based on the Lorentzian norm as the data fidelity term [41, 42]. In addition, Ramirez et al. develop an iterative algorithm to solve a Lorentzian ℓ_0-regularized cost function using iterative weighted myriad filters [43]. A similar approach is used in [44] by solving an ℓ_0-regularized least absolute deviation (LAD) regression problem yielding an iterative weighted median algorithm. The authors of [45] propose an iterative approach based on a gradient descent median truncated Wirtinger flow algorithm to solve the phase retrieval problem when the magnitude measurements are corrupted by outliers.

Nonconvex optimization approaches based on ℓ_p norms as data fidelity functions have been proposed in [46, 47], while an ℓ_p-space greedy algorithm is proposed in [48]. Greedy algorithms [49, 50] and optimization-based approaches [51, 52] using the Huber function as the data fidelity term have also been proposed in the literature. Bayesian approaches, modeling the corrupting noise using a heavy-tailed probality distribution, are proposed in [53, 54]. Robust PCA approaches resilient to outliers are proposed in [55].

The purpose of this article is to provide an overview of robust reconstruction strategies for CS when the measurements are corrupted by outliers. We approach the problem first from a statistical point of view and then review nonlinear methods that have been proposed in the literature that are based on robust statistics, specifically methods that are based on M-estimators. The organization of this paper is as follows. A general overview of CS is introduced in Section 2, and a collection of robust estimators, known as M-estimators, are discussed in Section 3. We then present a review of nonlinear methods based on robust estimation in Section 4. Section 5 is devoted to illustrate the performance of the reviewed methods in the reconstruction of sparse signals from compressive contaminated samples. Concluding remarks are provided in Section 6.

2 Compressive sensing review

Let $\mathbf{x} \in \mathbb{R}^n$ be a signal that is either s-sparse or compressible in some representation basis Ψ such that $\mathbf{x} = \Psi\alpha$, where $\alpha \in \mathbb{R}^n$ is the vector of coefficients having at most s nonzeros values, i.e., $\|\alpha\|_0 \leq s$. Recall that the ℓ_p norm of a vector $\mathbf{u} \in \mathbb{R}^n$ is defined as $\|\mathbf{u}\|_p = \left(\sum_{i=1}^n |u_i|^p\right)^{1/p}$. The ℓ_0 "norm" is not a norm, since it does not meet the positive homogeneity and sub-additivity properties, but in practice simply counts the number of nonzero elements of a vector. Let Φ be an $m \times n$ sensing matrix that represents a dimensionality reduction operation since m is taken to be smaller than n, with rows that form a set of vectors incoherent with the sparsity representation basis.

The signal $\mathbf{x}$ is measured by $\mathbf{y} = \Phi\mathbf{x}$. Setting $\Theta = \Phi\Psi$, the measurement vector becomes $\mathbf{y} = \Theta\alpha$. In the

following, we assume, without loss of generality, that $\Psi = \mathrm{I}$, the canonical basis for $\mathbb{R}^n$, such that $\mathbf{x} = \alpha$. It has been shown that a convex program (basis pursuit) can recover the original sparse signal, $\mathbf{x}$, from a small set of measurements, $\mathbf{y}$ if the sensing matrix obeys the restricted isometry property (RIP) [56], defined as follows.

Definition 1. *A matrix Φ satisfies the restricted isometry property of order s if there exists a constant δ_s, defined as the smallest positive quantity such that*

$$(1-\delta_s)\|\mathbf{x}\|_2^2 \leq \|\Phi\mathbf{x}\|_2^2 \leq (1+\delta_s)\|\mathbf{x}\|_2^2$$

holds for all $\mathbf{x} \in \Omega_s$, where $\Omega_s = \{\mathbf{x} \in \mathbb{R}^n | \ \|\mathbf{x}\|_0 \leq s\}$. A matrix Φ is said to satisfy the RIP of order s if $\delta_s \in (0,1)$.

Basically, the RIP dictates that every set of columns of Φ with cardinality smaller than s approximately behaves like an orthonormal system, such that it approximately preserves the ℓ_2-distance between any pair of s-sparse vectors. It has also been shown that random matrices with Gaussian or sub-Gaussian entries meet the RIP with high probability provided that $m = O(s\log(n))$ [5, 57].

The RIP has some implications concerning the robustness to noise. In a realistic scenario, the measurements are corrupted by noise and can be modeled as $\mathbf{y} = \Phi\mathbf{x} + \mathbf{z}$, where $\mathbf{z}$ is zero-mean additive white noise. It has been shown that under some characteristics of the noise, notably finite second order statistics or bounded noise in the ℓ_2 sense, and if the measurement matrix Φ satisfies the RIP condition, then there exists a variety of algorithms that are able to stably recover the sparse signal from noisy measurements [10]. Among those, basis pursuit denoising (BPD) relaxes the requirement that the reconstructed signal exactly explain the measurements, yielding the convex problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{y}-\Phi\mathbf{x}\|_2 \leq \epsilon, \tag{1}$$

for some small $\epsilon > 0$. Candès shows in [56] that if $\|\mathbf{y}-\Phi\mathbf{x}\|_2 \leq \epsilon$ and $\delta_{2s} < \sqrt{2}-1$, then the solution of (1), $\hat{\mathbf{x}}$, is guaranteed to obey $\|\mathbf{x}-\hat{\mathbf{x}}\|_2 \leq C\epsilon$, where the constant C depends on δ_{2s}.

Variations of (1) are also found in the literature, such as the ℓ_1-regularized least squares (ℓ_1-LS) problem, also known as the least absolute shrinkage and selection operator (LASSO) [58],

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{y}-\Phi\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1, \tag{2}$$

where λ is a regularization parameter that balances the weight between the data fidelity term and the ℓ_1 regularization term. The ℓ_1-LS problem is sometimes preferred over BPD because of the availability of efficient methods to solve (2) [59]. Other sparse reconstruction approaches including greedy algorithms, which iteratively construct sparse approximations, can be found in the literature. Orthogonal matching pursuit (OMP) [60, 61], regularized OMP [62], and iterative hard thresholding (IHT) [63] are examples of this class.

The aforementioned methods use the ℓ_2 norm as the *data-fitting* term, and they perform adequately under the assumption that the contaminating noise has finite second order statistics. However, just as in classical least squares and mean-square error estimation methods, ℓ_2-based sparse reconstruction methods tend to be very sensitive to outliers or gross error present in the measurements. Thus, it is natural to draw on the rich theory of robust linear regression [17–19, 64, 65] as a plausible approach to address the CS reconstruction problem when the measurements are contaminated with heavy-tailed noise. Several key robust estimators are reviewed in the following section.

3 M-estimators overview

The presence of outliers in CS measurements leads to the study of robust estimators since the recovered sparse signal is highly affected by the presence of the large errors in the data. Robust M-estimators bring substantial benefits in this scenario because, rather than relying on classical Gaussian ML estimation, they are based on modeling the contamination noise of the measurements as heavy-tailed process.

M-estimators are a generalization of ML estimators and are described by a cost function-defined optimization problem where properties of the cost function determine the estimator robustness [17–19]. In robust estimation theory, two important concepts characterize the robustness of an estimator: the breakdown point and the influence function.

The break down point is used to characterize quantitative robustness of an estimator. It indicates the maximal fraction of outliers (highly deviating samples) in the observations, which an estimator can handle without breaking down. The influence function describes the bias impact of infinitesimal contamination at an arbitrary point on the estimator, standardized by the fraction of contamination. For M-estimators, the influence function is proportional to the first derivative of the cost function [18]. Desirable properties of the influence function are boundedness and continuity. Boundedness ensures that a small fraction of contamination or outliers can have only a limited effect on the estimate, whereas continuity means that small changes in the data lead to small changes in the estimate.

Several robust M-estimators have been studied in the literature, and the most commonly used methods in CS are reviewed in the following. For simplicity of the exposition, the reviewed M-estimators are presented in the location estimation setting, i.e., the one-dimensional case, though the cost functions, and their properties, can be extended to the multidimensional case [14].

3.1 Median estimator

Consider a set of observations $\{y_1, y_2, \ldots, y_n\}$, where each observation follows the linear model $y_i = \alpha + z_i$, and the elements z_i are independent samples obeying a zero-mean Laplacian distribution. This is the classical location parameter estimation problem, which seeks the best estimate of α from a set of observations $\{y_1, y_2, \ldots, y_n\}$, where each observation has a scale parameter σ_i. The resulting ML estimate of α is given by

$$\hat{\alpha} = \arg\min_{\alpha} \sum_{i=1}^{n} \frac{1}{\sigma_i} |y_i - \alpha|. \tag{3}$$

Note that the cost function in the problem defined above is the absolute deviation function which is the one-dimensonal case of the ℓ_1 norm and its influence function, $\mathrm{IF}(x) = \mathrm{sign}(x)$, is bounded but discontinuos at the origin. The solution to (3) is the well-known *weighted median* (WM). The WM operator is defined by [66]

$$\hat{\alpha} = \mathrm{MEDIAN}(w_1 \diamond y_1, w_2 \diamond y_2, \cdots, w_n \diamond y_n), \tag{4}$$

where $w_i = 1/\sigma_i$ denotes the weight associated with the i-th observation sample and the symbol $\diamond$ represents an operator that replicates w_i times the value y_i; i.e. $w_i \diamond y_i = \overbrace{y_i, y_i, \cdots, y_i}^{w_i \text{times}}$. Thus, the WM operator consists of replicating the ith sample w_i times and sorting all the samples to then find the median value of the entire set. If the weights are real numbers instead of integers, the threshold decomposition framework can be applied to compute the weighted median [67].

3.2 ℓ_p estimator

Consider a set of observations $\{y_1, y_2, \ldots, y_n\}$, where each observation follows the linear model $y_i = \alpha + z_i$ and the elements z_i are independent and follow the zero-centered GGD. The probability density function of the GGD is given by

$$f(z) = \frac{p}{2\sigma\Gamma(1/p)} \exp\left(-\frac{|x|^p}{\sigma^p}\right), \tag{5}$$

where $\Gamma(\cdot)$ is the gamma function, σ is a scale parameter and $p > 0$, the so-called shade parameter, controls the tail decay rate. If each observation has a different scale parameter σ_i, the ML estimate of α is given by

$$\hat{\alpha} = \arg\min_{\alpha} \sum_{i=1}^{n} \frac{1}{\sigma_i^p} |y_i - \alpha|^p. \tag{6}$$

There are two special cases of the GGD family that are well studied: the Gaussian ($p = 2$) and Laplacian ($p = 1$) distributions, which yield the well-known *weighted mean* and *weighted median* estimators, respectively. Conceptually, the lower the value of p, the more heavy tailed is the distribution leading to more impulsive samples. When $p < 2$, the GGD exhibits heavier than Gaussian tails (super-Gaussian) and when $0 < p < 1$, the model is very impulsive. The $p \neq \{1, 2\}$ cases yield the fractional lower order moment (FLOM) estimation framework [68].

Recall that the ℓ_p norm of a vector $\mathbf{u} \in \mathbb{R}^m$ is defined as $\|\mathbf{u}\|_p = \left(\sum_{i=1}^{m} |u_i|^p\right)^{1/p}$. Note that the ℓ_p norms are convex and everywhere continuous functions when $p > 1$. The especial case $p = 1$ is the ℓ_1 norm that is convex but piece-wise continuous. When $0 < p < 1$, the ℓ_p norms are nonconvex and piece-wise continuous. In the latter case, the ℓ_p norms are not really norms in the strict sense, but quasi-norms, since the sub-additivity property is not satisfied. The influence function for the ℓ_p norms, $\mathrm{IF}(x) = \mathrm{sign}(x)p|x|^{p-1}$, is bounded but discontinuous at the origin for $0 < p \leq 1$ and continuous everywhere but not bounded for $p > 1$. Also note that the influence function is asymptotically redescending, i.e., $\mathrm{IF}(x) \to 0$ as $x \to \pm\infty$ when $0 < p < 1$. Having a redescending influence function is a desirable property in a robust estimator since large outliers do not influence the output of the estimate. Thus, the ℓ_p norms are optimal under GGD noise and offer a powerful framework for impulsive noise applications when $0 < p < 2$ [69–71].

3.3 Huber estimator

Consider now a set of observations $\{y_1, y_2, \ldots, y_n\}$, where each observation follows the linear model $y_i = \alpha + z_i$ and the elements z_i are i.i.d. random variables from a continuous GGP symmetric distribution, with scale parameter $\sigma > 0$. A robust estimator that combines the ℓ_2 and ℓ_1 norms as cost function is defined as [17]

$$\hat{\alpha} = \arg\min_{\alpha} \sum_{i=1}^{n} \rho\left(\frac{y_i - \alpha}{\sigma}\right), \tag{7}$$

where ρ is a convex and piece-wise continuous function, named the Huber's cost function, and it is given by

$$\rho(e) = \begin{cases} \frac{1}{2}e^2 & ;\text{for } |e| \leq c \\ c|e| - \frac{1}{2}c^2 & ;\text{for } |e| > c. \end{cases} \tag{8}$$

In Eq. (8), the parameter c is a tuning constant that influences the degree of robustness of the estimator [50]. The Huber cost function is one of the most popular cost functions in M-estimators since it combines the sensitivity properties of the ℓ_2 norm and the robustness to outliers of the ℓ_1 norm [17]. Robustness properties of the Huber estimator are dictated bye the scale parameter σ and the tuning constant c. Since the Huber cost function is a combination of the ℓ_2 and ℓ_1 norms, its influence function is also a combination of the two related influence functions. Thus, its influence function is bounded and piece-wise continuous.

3.4 Myriad estimator

Now consider a set of observations $\{y_1, y_2, \ldots, y_n\}$, where each observation again follows the linear model $y_i =$

$\alpha + z_i$, as described in the previous sub-section. However, the elements z_i are now i.i.d. samples obeying the standard Cauchy distribution. The Cauchy probability density function is given by

$$f(z) = \frac{\sigma}{\pi[\sigma^2 + z^2]}, \tag{9}$$

where σ is the scale parameter. The ML best estimate of α from a set of observations $\{y_1, y_2, \cdots, y_n\}$, where each observation has a scale parameter σ, is given by

$$\hat{\alpha} = \arg\min_{\alpha} \sum_{i=1}^{n} \log\left[\sigma^2 + (y_i - \alpha)^2\right]. \tag{10}$$

The solution for (10) is the *myriad* estimate. In this case, instead of using the sample mean or the sample median, the optimal solution minimizes the sum of logarithmic square deviations, referred as the Least Lorentzian Squares (LLS) criterion [16]. The influence function of the myriad estimator is given by $\mathrm{IF}(x) = 2x/(\sigma^2 + x^2)$. Note that this influence function is everywhere continuous, bounded, and asymptotically redescending. The myriad estimate is denoted as

$$\hat{\alpha} = \mathrm{MYRIAD}(\sigma; y_1, y_2, \ldots, y_n). \tag{11}$$

Note that the myriad estimate is also the ML estimator when z_i follow the Student's T distribution with 1 degree of freedom. The sample myriad has different modes of operation that depend on the tuning of the scale parameter σ, the so-called linearity parameter [72]. When the noise is Gaussian, for example, values of σ larger than the sample range, i.e., $\sigma \geq y_{(1)} - y_{(0)}$, where $y_{(q)}$ denotes the sample q-th quantile, can provide the optimal performance associated with the sample mean. On the other hand, setting σ as half the interquartile range, i.e., $\sigma = (y_{(0.75)} - y_{(0.25)})/2$, considers implicitly half the samples unreliable, giving resilience to gross errors. For highly impulsive noise statistics, mode-type estimators can be achieved by using small values of σ [72]. Different approaches to automatically adapt σ under different noise scenarios [73] and to efficiently compute the myriad estimate [74, 75] have been proposed.

In the following, the cost function of the myriad estimator, defined in (10), is extended to define a robust metric, known as the Lorentzian norm, for vectors in $\mathbb{R}^m$. Formally, the Lorentzian norm of a vector $\mathbf{u} \in \mathbb{R}^m$ is defined as

$$\|\mathbf{u}\|_{LL_2,\gamma} = \sum_{i=1}^{m} \log\left(1 + \frac{u_i^2}{\gamma^2}\right), \quad \gamma > 0. \tag{12}$$

The Lorentzian norm (or LL_2 norm) is not a norm in the strict sense, since it does not meet the positive homogeneity and sub-additivity properties. However, it defines a robust metric that does not heavily penalize large deviations, with the robustness depending on the scale parameter γ, thus making it an appropriate metric for impulsive environments (optimal in ML sense under the Cauchy model) [16, 72, 76, 77]. Further justification for the use of the Lorentzian norm is the existence of logarithmic moments for algebraic-tailed distributions, as second moments are infinite or not defined for such distributions and therefore not an appropriate measure of the process strength [13, 16].

3.5 M-generalized Cauchy estimator

Consider a set of observations $\{y_1, y_2, \ldots, y_n\}$, where each observation follows the linear model $y_i = \alpha + z_i$, and the elements z_i are i.i.d. samples obeying a generalized Cauchy distribution (GCD). The probability density function of the GCD is given by

$$f(z) = a\sigma(\sigma^p + |z|^p)^{-2/p}, \tag{13}$$

where $a = p\Gamma(2/p)/2(\Gamma(1/p))^2$. In (13), the scale parameter is given by σ, and the tail decay of the distribution is given by p. For the particular case $p = 2$, we have the Cauchy distribution. The ML estimate of the location parameter for GCD distributed samples is given by [77]:

$$\hat{\alpha} = \arg\min_{\alpha} \sum_{i=0}^{n} \log[\sigma^p + (y_i - \alpha)^p]. \tag{14}$$

The particular cases of $p = 1$ and $p = 2$ yield the *meridian* [78] and *myriad* [76] estimators, respectively.

3.6 A few comments on the cost functions

Figure 1 compares the ℓ_1 norm, the Huber cost function, with $\sigma = 1$ and $c = 0.75$, and the Lorentzian norm with two different values of γ ($\gamma = 1$ and $\gamma = 0.1$) for the one-dimensional case. The squared ℓ_2 norm is plotted as reference. Compared to the squared ℓ_2 norm, the ℓ_1, Huber and Lorentzian functions do not over penalize large deviations, leading to more robust error metrics when outliers are present. Notably, the Lorentzian norm and the Huber cost function, for $c < 1$, are more robust to outliers since they do not increase their value as fast as the ℓ_1 norm when $u \to \infty$.

In the same manner as the myriad estimator, robustness properties of the Lorentzian norm are defined by the scale parameter γ. The Lorentzian norm is convex in the interval $-\gamma \leq u \leq \gamma$ behaving as an ℓ_2 cost function for small variations compared to γ and log-concave outside this interval. Thus, small values of γ make the Lorentzian more resilient to gross errors and large values of γ make the Lorentzian similar to the squared ℓ_2 norm. Robustness properties of the Huber cost function also depend on the scale parameter σ and on the parameter c.

Although the Lorentzian norm is a nonconvex function, it is everywhere continuous and differentiable, which are desirable properties when used as a cost function in optimization problems. On the other hand, the ℓ_1

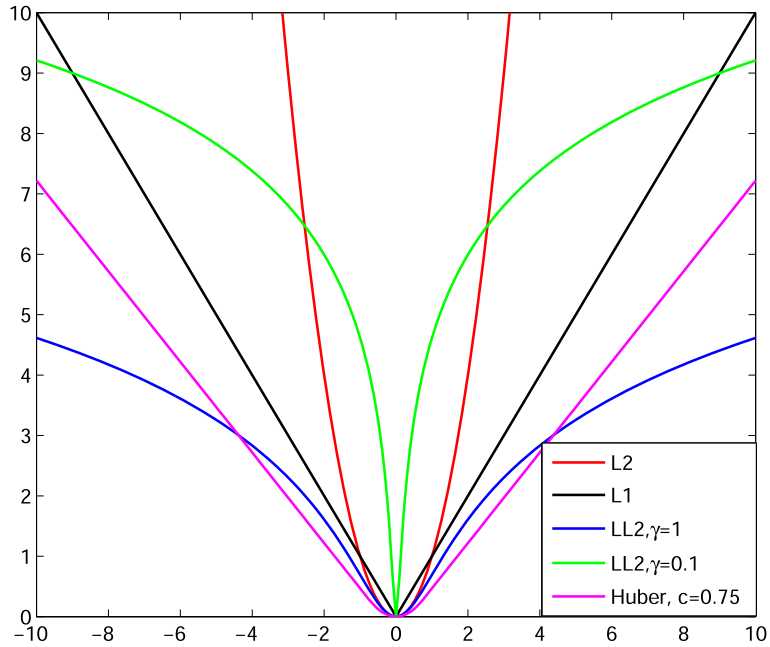

Fig. 1 Comparison of the ℓ_1 (*black*) norm, the Huber cost function with $c = 0.75$ (*magenta*), and the Lorentzian norm with $\gamma = 1$ (*blue*) and $\gamma = 0.1$ (*green*) for the-one dimensional case. The squared ℓ_2 norm (*red*) is plotted as reference

and Huber functions are convex and continuous functions, thus enjoying strong theoretical guarantees when used in optimization problems. However, the ℓ_1 norm is piece-wise continuous and not differentiable, which rules out traditional smooth optimization methods based on derivative information, whereas the Huber function is everywhere differentiable.

Figure 2 depicts the characteristics of the Lorentzian cost function (myriad estimator) for two different values of γ and for the ℓ_1 and Huber cost functions, in the location estimation problem. The observation samples are located in $x = \{-1, 0, 1, 10\}$. Note that for $\gamma = 0.1$, the Lorentzian cost function exhibits four local minima, whereas for $\gamma = 1$, the cost function is smoothed and only two local minima are present.

4 Review of robust sparse signal reconstruction methods

Recall that the CS signal estimation problem consists of reconstructing an s-sparse signal $\mathbf{x}_0 \in \mathbb{R}^n$ from a reduced set of noisy linear projections $\mathbf{y} \in \mathbb{R}^m$ given by

$$\mathbf{y} = \Phi\mathbf{x}_0 + \mathbf{z}, \tag{15}$$

where $\mathbf{z} \in \mathbb{R}^m$ is the noise vector with i.i.d. components following a common distribution $f_z(z)$. If the noise contains outliers, or is of impulsive nature, then it is better characterized by a distribution with heavier-than-Gaussian tails. A common model for the noise is to assume that $\mathbf{z} = \mathbf{r}_0 + \mathbf{w}$, where $\mathbf{r}_0$ is modeled as a sparse error whose locations of nonzero entries are unknown and whose magnitudes can be arbitrarily large and $\mathbf{w}$ is a small ℓ_2-bounded noise (possibly Gaussian). Another common model is to assume that $\mathbf{z}$ follows a heavy-tailed distribution such as the Laplace distribution or the alpha-stable distribution. In order to mitigate the effect of the impulsive noise in the compressive measurements, a robust *data-fitting* term should be used.

In this section, we present a set of formulations and methods for robust sparse signal reconstruction when the signals are acquired in the presence of impulsive noise. The approaches described herein are based on replacing the ℓ_2 norm by the previously described robust metrics for the data fidelity term.

4.1 ℓ_1-based methods

If the ℓ_2 norm is replaced by the ℓ_1 norm in the data-fitting term, the CS reconstruction problem reduces to solving a constrained LAD regression problem given by

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y} - \Phi\mathbf{x}\|_1 \text{ subjectto } \|\mathbf{x}\|_0 \leq s. \tag{16}$$

The problem in (16) is optimum under the ML assumption that the noise obeys a Laplacian distribution. The constraint term imposes sparsity in the estimated signal. The formulation in (16) can be rewritten as an unconstrained regularized problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y} - \Phi\mathbf{x}\|_1 + \tau\|\mathbf{x}\|_0. \tag{17}$$

Different strategies have been proposed to solve (17). Among these, a framework based on the coordinate

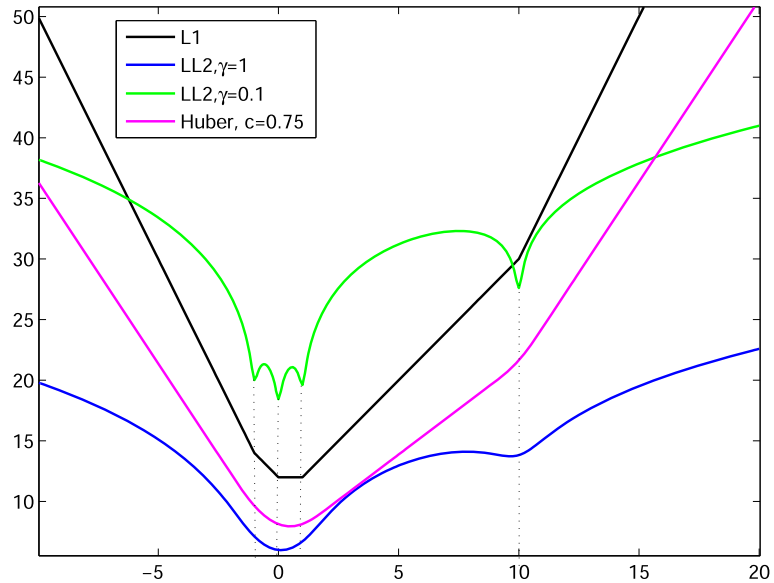

Fig. 2 Comparison of the ℓ_1 (*black*) norm, the Huber cost function with $c = 0.75$ (*magenta*) and the Lorentzian norm with $\gamma = 1$ (*blue*) and $\gamma = 0.1$ (*green*) for the location estimation problem, with observation samples located in $x = \{-1, 0, 1, 10\}$

descent approach is proposed in [44], though the problem in (17) is combinatorial and computationally expensive. Therefore, convex relaxations to the ℓ_0 constraint have been proposed. For instance, Wang et al. proposed the following convex problem [20]

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y} - \Phi\mathbf{x}\|_1 + \tau\|\mathbf{x}\|_1. \tag{18}$$

The reconstruction problem is thus formulated as a LAD ℓ_1 regularized problem (ℓ_1-LAD) whose theoretical properties for statistical regression are studied in [20]. The works in [23–28] study theoretical recovery conditions for the following equivalent convex problem

$$\min_{\mathbf{x}\in\mathbb{R}^n,\mathbf{r}\in\mathbb{R}^m} \tau\|\mathbf{x}\|_1 + \|\mathbf{r}\|_1 \text{ subject to } \mathbf{y} = \Phi\mathbf{x} + \mathbf{r}, \tag{19}$$

where $\mathbf{r}$ is a slack variable that represents the corrupting vector, i.e., $\mathbf{z} = \mathbf{r}_0$, where $\mathbf{r}_0$ is a sparse vector with unknown nonzero locations and possibly large magnitudes. The parameter τ controls the balance between the two ℓ_1 terms in (19). If a large value of τ is used, then the problem can recover a dense error for a sufficiently sparse signal. On the other hand, if a small value of τ is chosen, then only a small fraction of corrupted measurements can be corrected.

Approaches that model the corrupting noise as $\mathbf{z} = \mathbf{r}_0 + \mathbf{w}$, where $\mathbf{r}_0$ is assumed sparse and $\mathbf{w}$ is a small ℓ_2-bounded noise, are also studied [26–30]. These works study theoretical recovery conditions of the following convex program

$$\min_{\mathbf{x}\in\mathbb{R}^n,\mathbf{r}\in\mathbb{R}^m} \tau\|\mathbf{x}\|_1 + \|\mathbf{r}\|_1 \text{ subject to } \|\mathbf{y} - \Phi\mathbf{x} - \mathbf{r}\|_2 \le \epsilon, \tag{20}$$

where ϵ is an bound on the ℓ_2 norm of $\mathbf{w}$.

Recovery guarantees based on the RIP of the extended matrix $[\,\Phi\ \mathrm{I}]$ were reported in [23]. These recovery guarantees are particularly useful when Φ is, for example, i.i.d. Gaussian. Ngunyen and Tran reported results based on a structured model of the matrix Φ [28]. They assume that Φ is formed by selecting rows from an orthogonal matrix with a low incoherence parameter μ, which is the minimum value such that $|\Phi_{ij}|^2 \le \mu/n$ for any i, j. Under these assumptions, they showed that (20) can recover both $\mathbf{x}_0$ and $\mathbf{r}_0$ with high probability if $m \ge C\mu^2\|\mathbf{x}_0\|_0(\log n)^2$ and $\|\mathbf{r}_0\|_0 \le \gamma m$, $\gamma \in (0, 1)$, which are nearly optimal conditions for the number of measurements and the sparsity of the error vector, i.e., the number of gross errors that can be corrected. The following theorem, shown by Li [27], presents stable recovery guarantees under a probabilistic model.

Theorem 1. ([27]) *Let $\Phi \in \mathbb{R}^{m\times n}$ be a sensing matrix whose entries are i.i.d. Gaussian random variables with zero mean and variance $1/m$ and set $\tau = \sqrt{\log(n/m) + 1}$. Then, if $\|\mathbf{w}\|_2 \le \epsilon$, $\|\mathbf{x}_0\|_0 \le \gamma m/(\log(n/m) + 1)$ and $\|\mathbf{r}_0\|_0 \le \gamma m$, $\gamma \in (0, 1)$, the solution to the convex problem in* (20)*, $(\hat{\mathbf{x}}, \hat{\mathbf{r}})$, satisfies*

$$\|\mathbf{x}_0 - \hat{\mathbf{x}}\|_2 + \|\mathbf{r}_0 - \hat{\mathbf{r}}\|_2 \le K\epsilon \tag{21}$$

with probability at least $1 - C\exp(-cm)$, where K, C, and c are numerical constants.

The results in Theorem 1 show that the signal can be stably recovered if the number of gross errors is up to a fixed fraction of the number of measurements. The bounds on the number of measurements are nearly optimal compared to the standard CS problem. Deterministic recovery conditions based on the coherence of the matrix Φ and the number of nonzero entries of $\mathbf{x}$ and $\mathbf{r}$ were resported in [26, 30]. These coherence-based results do not assume any particular model for the matrix Φ.

Ngunyen and Tran proposed the extended Lasso (or robust Lasso, R-Lasso) estimator that solves the following convex problem [29]

$$\min_{\mathbf{x}\in\mathbb{R}^n,\mathbf{r}\in\mathbb{R}^m} \frac{1}{2}\|\mathbf{y}-\Phi\mathbf{x}-\mathbf{r}\|_2^2+\tau_x\|\mathbf{x}\|_1+\tau_r\|\mathbf{r}\|_1, \quad (22)$$

where τ_x and τ_r are regularization parameters. Recovery guarantees based on a extended restricted eigenvalue condition of the matrix Φ and bounds for the regularization parameters τ_x and τ_r are studied in [29]. Note that the problems in (20) and (22) are convex and can be efficiently solved using standard optimization algorithms to solve the ℓ_1-LS and BPD problems (see [59, 79]) by using the extended model $\tilde{\Phi}=[\,\Phi\ \mathrm{I}]$ and $\tilde{\mathbf{x}}=[\,\mathbf{x}^T,\mathbf{r}^T]^T$. In the following, we describe several approaches for solving (17) and (18).

ℓ_1-based coordinate descent algorithm The problem in (17) is combinatorial and nonsmooth, thus a greedy strategy based on the coordinate descent algorithm and the weighted median estimator is proposed in [44]. In this scheme, each element of the sparse vector $\mathbf{x}$ is estimated at each step, while keeping the other elements of the vector fixed. The solution for the one-dimensional problem is then given by

$$\tilde{x}_j=\mathrm{MEDIAN}\left\{|\phi_{i,j}|\diamond\frac{y_i-\sum_{k=1,k\neq j}^n\phi_{i,k}x_k}{\phi_{i,j}}\Big|_{i=1}^m\right\}. \quad (23)$$

The sparsity constraint given by the ℓ_0-regularization norm is included in the solution by computing the hard thresholding operator after computing the weighted median estimate. Thus, the solution is

$$\hat{x}_j=\begin{cases}\tilde{x}_j & ;\text{if } \|\mathbf{r}_j\|_1>\|\mathbf{r}_j-\phi_j\tilde{x}_j\|_1+\tau\\ 0 & ;\text{otherwise,}\end{cases}$$

where $\mathbf{r}_j=\mathbf{y}-\sum_{k=1,k\neq j}^n\phi_k x_k$ is the j-th residual term that remains after removing the contribution of all the components of the estimated vector except the j-th component, and ϕ_k denotes the k-th column vector of the measurement matrix. The coordinate-descent approach is computationally expensive because the estimation of the sparse vector requires cycling through all the components at each iteration of the algorithm.

ℓ_1-based alternating direction method The problems posed in (18) or (19) are convex but nonsmooth. However, they can be solved using the alternating direction method of multipliers (ADMM) [79, 80]. ADMM solves the extended ℓ_1 problem (19) by finding a saddle point of the augmented Lagrangian function

$$\tau\|\mathbf{x}\|_1+\|\mathbf{r}\|_1+\mathbf{z}^T(\Phi\mathbf{x}+\mathbf{r}-\mathbf{y})+\frac{\beta}{2}\|\Phi\mathbf{x}+\mathbf{r}-\mathbf{y}\|_2^2, \quad (24)$$

where $\mathbf{z}$ is the Langrange multiplier's vector and $\beta>0$ is a penalty constant.

The following iterative algorithm is derived in [80] to find a solution for (19):

$$\begin{aligned}\mathbf{r}^{(k+1)}&=\mathrm{Shrink}(\mathbf{z}^{(k)}/\beta-\Phi\mathbf{x}^{(k)}+\mathbf{y},1/\beta)\\ \mathbf{x}^{(k+1)}&=\mathrm{Shrink}(\mathbf{x}^{(k)}-\mu\mathbf{g}^{(k)},\mu\tau/\beta)\\ \mathbf{z}^{(k+1)}&=\mathbf{z}^{(k)}-\nu\beta(\Phi\mathbf{x}^{(k+1)}+\mathbf{r}^{(k+1)}-\mathbf{y}),\end{aligned} \quad (25)$$

where $\mathbf{g}^{(k)}=\Phi^T(\mathbf{z}^{(k)}/\beta+\Phi\mathbf{x}^{(k)}+\mathbf{r}^{(k+1)}-\mathbf{y})$ is the gradient of the differentiable part of the augmented Lagrangian function with respect to $\mathbf{x}$ and $\mathrm{Shrink}(\cdot,\rho)$ denotes the shrinkage operator defined as $\mathrm{Shrink}(\mathbf{a},\rho)_i=sgn(a_i)\max(|a_i|-\rho,0)$. The parameters μ and ν are step sizes. Convergence conditions for μ and ν and strategies to select β are detailed in [80].

4.2 ℓ_p-based methods

If the corrupting noise has heavier tails than the Laplacian distribution, the ℓ_p norm, with $0<p<1$, can be used as the data-fitting term yielding the following recovery optimization problem:

$$\min_{\mathbf{x}\in\mathbb{R}^n}\|\mathbf{y}-\Phi\mathbf{x}\|_p^p+\tau\|\mathbf{x}\|_1. \quad (26)$$

The problem in (26) is optimal under the ML criteria for GGD noise and robust to very impulsive noise. Numerical methods have been proposed to efficiently solve (26) for the $0<p<2$ case [47]. The algorithm is based on incorporating the proximity operator of the ℓ_p norm into the framework of ADMM. For the nonconvex case ($0<p<1$), a smoothing strategy has been employed to derive a convergent algorithm. Stability results similar to those derived in [41] are derived in [47] based on the RIP of Φ.

Filipovic studied the following related problem [46]

$$\min_{\mathbf{x}\in\mathbb{R}^n,\mathbf{r}\in\mathbb{R}^m}\tau\|\mathbf{x}\|_p^p+\|\mathbf{r}\|_p^p \text{ subject to } \mathbf{y}=\Phi\mathbf{x}+\mathbf{r}, \quad (27)$$

where the ℓ_1 norm is replaced by the ℓ_p norm as sparsity promoting function and $\mathbf{r}$ is a slack variable that represents the corrupting sparse vector. The following theorem presents theoretical recovery conditions based on the RIP of the extended matrix $[\,\Phi\ \mathrm{I}]$.

Theorem 2. ([46]) *Consider the extended sensing matrix* $\tilde{\Phi}=[\,\Phi\ \mathrm{I}]$. *Denote by* $K_1=\|\mathbf{x}_0\|_0$ *and* $K_2=\|\mathbf{r}_0\|_0$. *Let* $a_1\leq 1$ *and* $a_2\leq 1$ *be constants such that* a_1K_1 *and* a_2K_2

are integers and define $a = \min(a_1, a_2)$. *Let* $c \leq 1$ *and* b *be constants such that*

$$b = \frac{a^{\frac{1}{p}-\frac{1}{2}}}{2^{\frac{1}{p}} c^2} > 1.$$

If

$$\frac{1}{\tau} \in \left[\frac{1}{c^p} \left(\frac{a_1 K_1}{a_2 K_2} \right)^{1-\frac{p}{2}}, c^p \left(\frac{a_1 K_1}{a_2 K_2} \right)^{1-\frac{p}{2}} \right]$$

and $\tilde{\Phi}$ *satisfies*

$$\delta_{a_1 K_1 + a_2 K_2} + b^2 \delta_{(a_1+1)K_1 + (a_2+1)K_2} < b^2 - 1,$$

then the unique minimizer of (27) *yields exactly* $\mathbf{x}_0$ *and* $\mathbf{r}_0$.

Greedy methods that use the ℓ_p norm as data fidelity term are also studied. Zeng et al. proposed robust versions of MP and OMP, coined ℓ_p-MP and ℓ_p-OMP, respectively, based on the notion of ℓ_p-space correlation, with $0 < p < 2$, which is robust to outliers [48]. The ℓ_p-correlation is defined as follows. Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$ with finite ℓ_p norm. Then the ℓ_p-correlation, with $0 < p < 2$, is defined as

$$c_p(\mathbf{a}, \mathbf{b}) = 1 - \frac{\min_{\alpha \in \mathbb{R}} \|\mathbf{b} - \alpha \mathbf{a}\|_p^p}{\|\mathbf{b}\|_p^p}. \tag{28}$$

The function $\|\mathbf{b} - \alpha \mathbf{a}\|_p^p$ is the ℓ_p norm of the fitting error of the univariate linear regression model $\mathbf{b} = \alpha \mathbf{a} + \mathbf{z}$ where $\mathbf{z}$ denotes the error vector. If there exist an α such that $c_p(\mathbf{a}, \mathbf{b}) = 1$, then $\mathbf{a}$ and $\mathbf{b}$ are collinear. On the other hand, if $c_p(\mathbf{a}, \mathbf{b}) = 0$ then $\mathbf{a}$ and $\mathbf{b}$ are said to be orthogonal [48].

4.3 Huber loss-based methods

Consider now that data-fitting term in the CS reconstruction problem uses the Huber cost function that combines the ℓ_2 and ℓ_1 norms. Then, a sparse signal can be estimated by solving the following constrained problem [50]

$$\min_{\mathbf{x} \in \mathbb{R}^n, \sigma \in \mathbb{R}_+} \sigma \sum_{i=1}^{m} \rho\left(\frac{y_i - \phi_i^T \mathbf{x}}{\sigma} \right) + (m-s)\alpha\sigma \text{ s.t. } \|\mathbf{x}\|_0 \leq s, \tag{29}$$

where ρ is the piece-wise continuous and convex function defined in Eq. (8), ϕ_i denotes the column vector obtained by transposing the i-th row of Φ, and $\alpha > 0$ is a scaling factor. Note that the problem in (29) is combinatorial and that both $\mathbf{x}$ and the scale parameter σ are simultaneously estimated. Ollila et al. [50] derived an iterative hard thresholding algorithm coined Huber iterative hard thresholding (HIHT) to solve the problem (29). A detailed analysis of the selection of the parameters α and c is presented in [50]. Also note that this framework can be extended to any robust cost function that meets some regularity conditions, e.g., the Tukey's bi-weight function [49].

Convex optimization approaches have also been proposed by Pham et al. [51, 52]. In these works, the sparse signal is estimated by solving the following convex unconstrained problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \sigma \sum_{i=1}^{m} \rho\left(\frac{y_i - \phi_i^T \mathbf{x}}{\sigma} \right) + \lambda \|\mathbf{x}\|_1, \tag{30}$$

where σ is estimated beforehand and λ is a regularization parameter that controls the sparsity level of the solution. Efficient algorithms to solve (30) based on the fast iterative shrinkage algorithm (FISTA) and ADMM and the adequate selection of the parameter λ are presented in [51, 52].

4.4 Lorentzian-based methods

For a more general type of heavy-tailed noise, the reconstruction of sparse signals can be formulated using the Lorentzian norm as a fitting term. The formulations and algorithms described next are based on the Lorentzian norm as a robust error metric, which is appropriate for many impulsive environments.

Lorentzian-based basis pursuit Using the strong theoretical guarantees of ℓ_1 minimization for sparse recovery in CS, Carrillo et al. studied the following nonconvex constrained optimization problem to estimate a sparse signal from the noisy measurements [41]

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{y} - \Phi \mathbf{x}\|_{LL_2,\gamma} \leq \rho. \tag{31}$$

The following theorem presents an upper bound for the reconstruction error of the proposed estimator in (31).

Theorem 3. ([41]) *Let* $\Phi \in \mathbb{R}^{m \times n}$ *be a sensing matrix such that* $\delta_{2s} < \sqrt{2} - 1$. *Then for any signal* $\mathbf{x}_0$ *such that* $|supp(\mathbf{x}_0)| \leq s$, *and observation noise* $\mathbf{z}$ *with* $\|\mathbf{z}\|_{LL_2,\gamma} \leq \rho$, *the solution to* (31), $\mathbf{x}^*$, *obeys the following bound*

$$\|\mathbf{x}_0 - \mathbf{x}^*\|_2 \leq C_s \gamma \sqrt{m(e^\rho - 1)}, \tag{32}$$

where the constant C_s *depends only on* δ_{2s}.

Theorem 3 shows that the solution to (31) is a sparse signal with an ℓ_2 error that is dependent on logarithmic moments. Note that the dependence on the noise logarithmic moment, rather than its second order moment, makes the formulation in (31) robust and stable to algebraic-tailed and impulsively corrupted samples. The optimization problem in (31) is referred to as Lorentzian BP (LBP). The scale parameter γ controls the robustness of the norm and ρ the radius of the LL_2 ball thus defining the feasible set. The scale parameter is estimated as $\gamma = (y_{(0.875)} - y_{(0.125)})/2$, where $y_{(q)}$ denotes the q-th quantile of the corrupted measurement vector $\mathbf{y}$ [41]. The reader is referred to [41] for further details on strategies to estimate γ and ρ based on the Cauchy model.

The LBP problem is hard to solve since it has a nonsmooth convex objective function and a nonconvex, nonlinear constraint. A sequential quadratic programming (SQP) method with a smooth approximation of the ℓ_1 norm is used in [41] to numerically solve the problem in (31). However, a less expensive approach is to solve a sequence of unconstrained problems of the form

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y}-\Phi\mathbf{x}\|_{LL_2,\gamma} + \lambda\|\mathbf{x}\|_1, \tag{33}$$

where λ is a regularization parameter that is changed in a decreasing manner at every iteration following an homotopy approach. The solution of the previous problem is used as starting point for the next problem. Since the Lorentzian norm is differentiable (though not Lipschitz differentiable), a nonconvex proximal-gradient algorithm [81] can be used to efficiently solve (33).

Lorentzian-based iterative hard thresholding algorithm Even though Lorentzian BP provides a robust CS framework in heavy-tailed environments, as explained above, numerical algorithms to solve the proposed optimization problem are not efficient [41]. Therefore, Carrillo and Barner proposed a Lorentzian-based iterative hard thresholding (IHT) algorithm [42]. In order to estimate $\mathbf{x}_0$ from $\mathbf{y}$, the following optimization problem is proposed:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y}-\Phi\mathbf{x}\|_{LL_2,\gamma} \ \text{ subject to } \ \|\mathbf{x}\|_0 \le s. \tag{34}$$

The problem in (34) is nonconvex and combinatorial. Therefore, the authors derive a greedy algorithm to estimate $\mathbf{x}_0$ based on the gradient projection algorithm [42]. The proposed strategy is formulated as follows. Let $\mathbf{x}^{(t)}$ denote the solution at iteration time t and set $\mathbf{x}^{(0)}$ to the zero vector. At each iteration t, the algorithm makes the update

$$\mathbf{x}^{(t+1)} = H_s\left(\mathbf{x}^{(t)} - \mu_t\Phi^T\mathsf{W}_t(\Phi\mathbf{x}^{(t)}-\mathbf{y})\right), \tag{35}$$

where $H_s(\mathbf{a})$ is the nonlinear operator that sets all but the largest (in magnitude) s elements of $\mathbf{a}$ to zero, μ_t is a step size, and W_t is an $m\times m$ diagonal matrix with each element defined as

$$\mathsf{W}_t(i,i) = \frac{\gamma^2}{\gamma^2 + (y_i - \phi_i^T\mathbf{x}^{(t)})^2},\ i=1,\ldots,m,$$

where ϕ_i denotes the column vector obtained by transposing the i-th row of Φ.

The algorithm defined by the update in (35) is coined Lorentzian iterative hard thresholding (LIHT). Note that $\mathsf{W}_t(i,i)\le 1$, thus, the weights diminish the effect of gross errors by assigning a small weight (close to zero) for large deviations compared to γ, and a weight near one for deviations close to zero. In fact, if W_t is the identity matrix, the algorithm reduces to the ℓ_2-based IHT [63]. The algorithm is a fast and simple method that only requires the application of Φ and Φ^T at each iteration.

Although the algorithm is not guaranteed to converge to a global minimum of (34), it can be shown that LIHT converges to a local minimum [42]. In the following, we show that LIHT has theoretical stability guarantees similar to those of the ℓ_2-based IHT. For simplicity of the analysis, we set $\mu_t = 1$ and assume that $\|\Phi\| \le 1$, where $\|\cdot\|$ denotes the spectral norm of a matrix.

Theorem 4. ([42]) *Let $\mathbf{x}_0\in\mathbb{R}^n$ and define $S = supp(\mathbf{x}_0)$, $|S|\le s$. Suppose $\Phi\in\mathbb{R}^{m\times n}$ meets the RIP of order 3s with $\delta_{3s} < 1/\sqrt{32}$. Assume $\mathbf{x}^{(0)} = 0$. Then, if $\|\mathbf{z}\|_{LL_2,\gamma}\le\tau$, the reconstruction error of the LIHT algorithm at iteration t is bounded by*

$$\|\mathbf{x}_0 - \mathbf{x}^{(t)}\|_2 \le \alpha^t\|\mathbf{x}_0\|_2 + \beta\gamma\sqrt{m(e^\tau-1)}, \tag{36}$$

where $\alpha = \sqrt{8}\delta_{3s}$ and $\beta = \sqrt{1+\delta_{2s}}(1-\alpha^t)(1-\alpha)^{-1}$.

The results in Theorem 4 can be easily extended to compressible signals using Lemma 6.1 in [82]. The scale parameter γ is estimated from $\mathbf{y}$ in the same manner described previosly for LBP. The step size μ_t is adapted at every iteration using a line search scheme with backtracking. See [42] for details.

Lorentzian-based coordinate descent algorithm In the context of CS random projections contaminated with Cauchy distributed noise, a suitable formulation for the reconstruction of sparse signals is

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y}-\Phi\mathbf{x}\|_{LL_2,\gamma} + \tau\|\mathbf{x}\|_0 \tag{37}$$

where τ is a regularization parameter that balances the influence of the Lorentzian norm as fitting-term and the sparsity-inducing term (ℓ_0-term) on the optimal solution. The coordinate-descent approach updates the estimate of each element of the sparse vector $\mathbf{x}$, while keeping the others fixed. Without loss of generality, the solution for the one-dimensional version of (37) is given by the following theorem.

Theorem 5. ([43]) *Let the function $Q(\mathbf{z}_j; x_j)$, with $\mathbf{z}_j = [z_{1,j},\ldots,z_{m,j}]$, be the Lorentzian norm, for the one-dimensional case, defined as*

$$Q(\mathbf{z}_j; x_j) = \sum_{i=1}^{m}\log\left[\kappa^2 + W_{i,j}(z_{i,j}-x_j)^2\right] \tag{38}$$

where κ is a linearity parameter and $W_{i,j} = \frac{\kappa^2}{\eta_{i,j}^2}$ are the weights having the parameter $\eta_{i,j}$ given by $\eta_{i,j} = \frac{\sum_{i=1,i\neq j}^{m}|y_i|}{\phi_{i,j}}$. The elements $z_{i,j}$ correspond to the i-th observation sample weighted by the element (i,j) of the sampling matrix, i.e., $z_{i,j} = \frac{y_i}{\phi_{i,j}}$. The solution to the ℓ_0-Regularized Lorentzian problem in (37) *is given by*

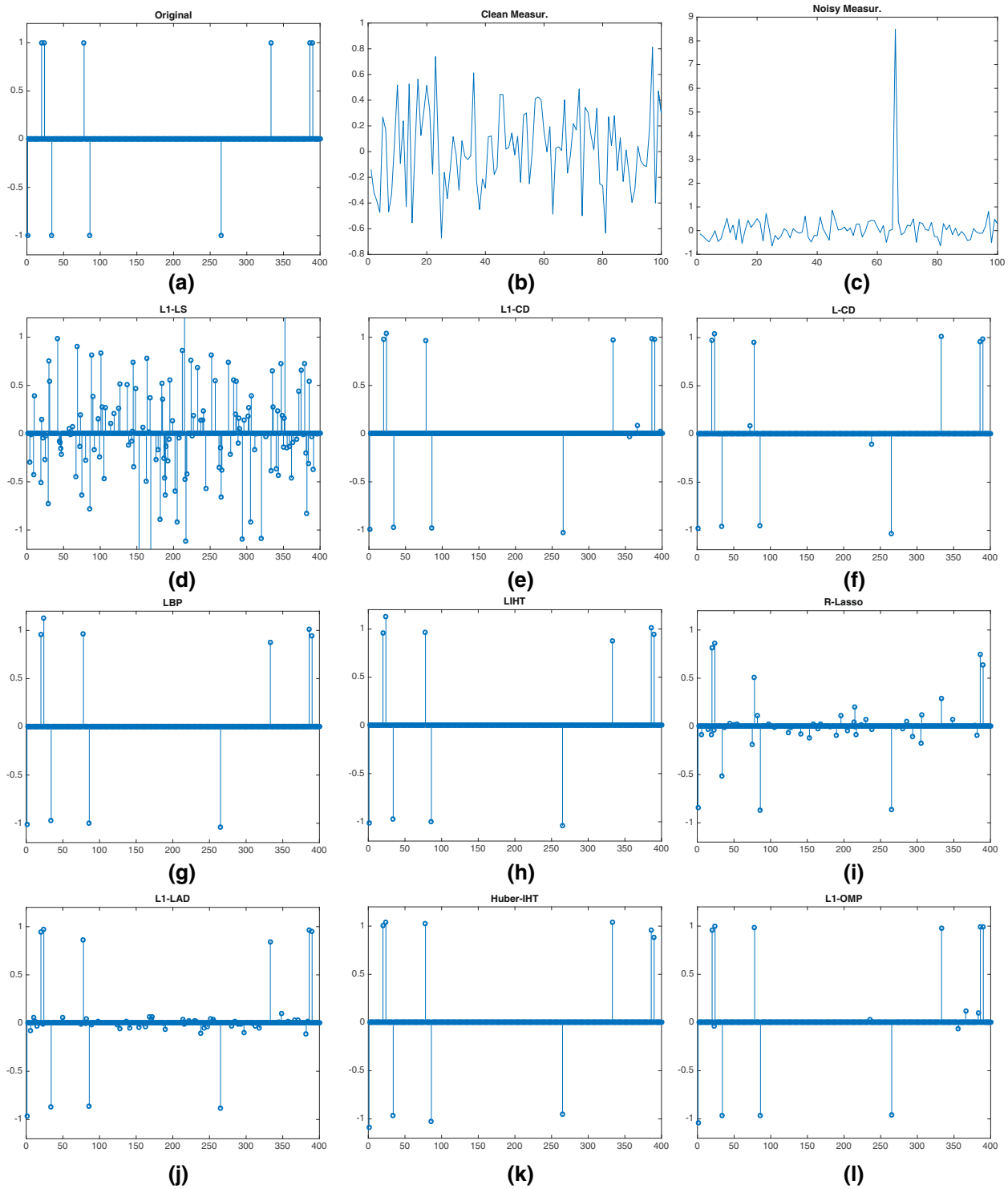

Fig. 3 Sparse signal reconstruction example from α-stable corrupted measurements ($s = 10$, $m = 100$, $n = 400$, and $\alpha = 1$). **a** True signal. **b** Clean measurements. **c** Noisy measurements. **d** Reconstructed signal using ℓ_1-LS (SER = −6.5 dB). **e** Reconstructed signal using ℓ_1-CD (SER = 28.2 dB). **f** Reconstructed signal using L-CD (SER = 25.1 dB). **g** Reconstructed signal using LBP (SER = 24.0 dB). **h** Reconstructed signal using LIHT (SER = 24.0 dB). **i** Reconstructed signal using R-Lasso (SER = 8.9 dB). **j** Reconstructed signal using ℓ_1-LAD (SER = 16.9 dB). **k** Reconstructed signal using Huber-IHT (SER = 25.1 dB). **l** Reconstructed signal using ℓ_1-OMP (SER = 24.1 dB)

$$\hat{x}_j = \begin{cases} \tilde{x}_j \ ; \textit{if } Q(\mathbf{z}_j; 0) > Q(\mathbf{z}_j; \tilde{x}_j) + \tau \\ 0 \ \ ; \textit{otherwise,} \end{cases}$$

where $\tilde{x}_j = \arg\min_{x_j} Q(\mathbf{z}_j; x_j)$ *and* τ *is the regularization parameter that governs the sparsity of the solution.*

Since this method requires the estimation of one coordinate at a time per iteration, the method is computationally expensive. A modified version, that accelerates the reconstruction of sparse signals by determining which coordinates are allowed to be estimated at each iteration, was proposed in [83].

5 Illustrative numerical examples

In this section, we present numerical experiments that illustrate the robustness and effectiveness of the reviewed methods, for the recovery of a sparse signal from noisy compressive samples. In particular, we compare the performance of the following robust methods: ℓ_1-based coordinate descent (ℓ_1-CD) [44], the ℓ_1-LAD method solved by ADMM, Lorentzian-based basis pursuit (LBP) [41], Lorentzian-based iterative hard thresholding (LITH) [42], Lorentzian-based coordinate descent (L-CD) [43], the robust lasso (R-Lasso) method [29], the Huber iterative hard thresholding (HIHT) method [50], and the ℓ_1-OMP method [48]. In order to evaluate the susceptiveness to outliers of traditional CS methods, we also include the performance of the ℓ_1-LS method [58].

First, all methods are tested in the reconstruction of a synthetic 10-sparse signal in the canonical basis, with length $n = 400$. The nonzero coefficients have equal amplitude, equiprobable sign, and randomly chosen position. Gaussian sensing matrices are employed with $m = 100$, and the measurements are then contaminated with α-stable noise (with $\alpha = 1$). Figure 3a shows the true signal and Fig. 3b, c shows the clean and contaminated measurements, respectively.

Figure 3d–l depicts the reconstructed signal obtained with all nine methods. Performance is measured using the signal-to-error ratio (SER), defined by

$$\text{SER(dB)} = 10 \log_{10} \left\{ \frac{\sum_{i=0}^{n} x_i^2}{\sum_{i=0}^{n} (x_i - \hat{x}_i)^2} \right\}. \tag{39}$$

For the proposed experiment (see Fig. 3), all methods perform relatively well at reconstructing the sparse signal from a small number of random measurements, except the ℓ_2-based ℓ_1-LS method. It is clear that the traditional ℓ_1-LS method for CS fails at estimating the sparse signal when gross errors are present in the compressed measurements. R-Lasso and ℓ_1-LAD are slightly more robust than ℓ_1-LS because the true support is correctly estimated although some components outside the true support also have strong amplitude. On the other hand, the coordinate descent approaches ℓ_1-CD and L-CD are greedy methods that correctly identify the true support with correct amplitudes. A few components appear at wrong coordinates but with small amplitude values. LIHT, LBP, HIHT, and ℓ_1-OMP methods can also correctly identify the components but the amplitudes are not completely correct. A summary of the reconstruction of the sparse one-dimensional signal for all methods is given in the third column of Table 1.

The second experiment explores the behavior of all nine methods in the reconstruction of a sparse signal from measurements acquired in different impulsive environments. We compare all methods in the reconstruction of a sparse signal having the same characteristics as in the first experiment, i.e., $s = 10, m = 100, n = 400$. However, now the random projections are contaminated with α-stable noise, with the tail parameter, α, varying from 0.2 to 2, i.e., from very impulsive to the Gaussian case. The scale parameter of the noise is set to $\sigma = 0.01$ for all cases. The results are depicted in Fig. 4. All results are averaged over 100 realizations of the sensing matrix, noise, and the sparse signals.

Table 1 Summary of sparse reconstruction methods

Method	Optimization problem	SER for signal [dB]	SER for image [dB]	Time [s]
LBP	$\|\mathbf{y} - \Phi\mathbf{x}\|_{\text{LL}_{2,\gamma}} + \lambda\|\mathbf{x}\|_1$	24.0	20.7	10.58
LIHT	$\|\mathbf{y} - \Phi\mathbf{x}\|_{\text{LL}_{2,\gamma}}$ s.t. $\|\mathbf{x}\|_0 \leq s$	24.0	19.4	2.13
R-Lasso	$\frac{1}{2}\|\mathbf{y} - \Phi\mathbf{x} - \mathbf{r}\|_2^2 + \tau_x\|\mathbf{x}\|_1 + \tau_r\|\mathbf{r}\|_1$	8.9	18.1	7.23
L-CD	$\|\mathbf{y} - \Phi\mathbf{x}\|_{\text{LL}_{2,\gamma}} + \tau\|\mathbf{x}\|_0$	25.1	19.2	6522.7
ℓ_1-CD	$\|\mathbf{y} - \Phi\mathbf{x}\|_1 + \tau\|\mathbf{x}\|_0$	28.2	20.3	3814.2
ℓ_1-LS	$\frac{1}{2}\|\mathbf{y} - \Phi\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$	–6.5	7.3	4.73
ℓ_1-LAD	$\|\mathbf{y} - \Phi\mathbf{x}\|_1 + \tau\|\mathbf{x}\|_1$	16.9	19.6	7.05
HIHT	$\sum_{i=1}^{M} \rho\left(\frac{y_i - \phi_i^T\mathbf{x}}{\sigma}\right)$ s.t. $\|\mathbf{x}\|_0 \leq s$	25.1	19.4	90.78
ℓ_1-OMP	$\|\mathbf{y} - \Phi\mathbf{x}\|_1$ s.t. $\|\mathbf{x}\|_0 \leq s$	24.1	–	–

First column acronym for the method. *Second column* optimization problem. *Third column* SER in dB obtained in the reconstruction of a sparse signal. *Fourth column* SER in dB obtained in the reconstruction of the cameraman image. *Fifth column* execution time required to reconstruct the cameraman image

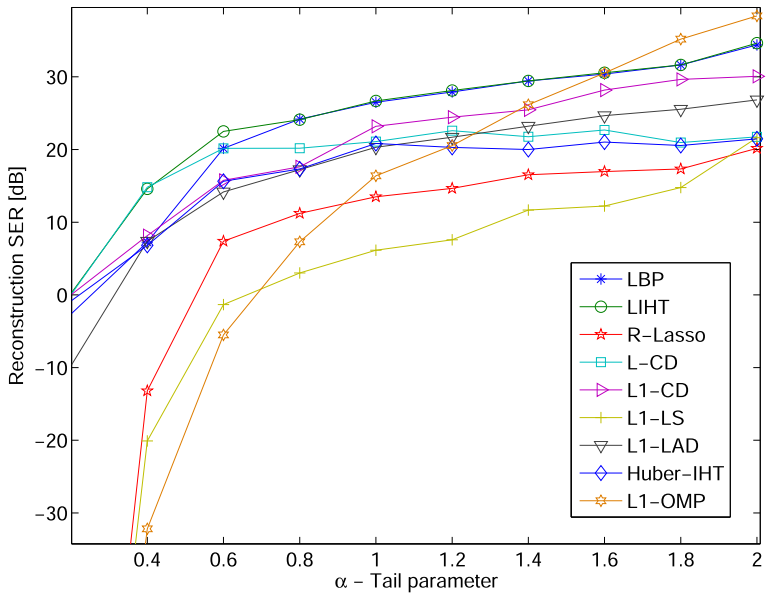

Fig. 4 Sparse signals reconstruction from α-stable corrupted measurements, for α varying from 0.2 to 2 ($s = 10$, $m = 100$, $n = 400$)

It can again be noticed that the ℓ_1-LS and ℓ_1-OMP methods fail at reconstructing the signals in very impulsive noise (for $\alpha < 1$). As the α tail parameter increases, these methods improves the average SER, giving the best result for the Gaussian case. The ℓ_1-OMP method yields faithful reconstructions for $\alpha > 1.2$. The robust-lasso (R-Lasso) is able to reconstruct sparse signals when the noise tail parameter is larger than 0.8. For very impulse noise ($\alpha < 0.8$), the reconstruction SER is highly degraded. All other robust methods are able to reconstruct the sparse signals, even in noise environments with tail parameters of $\alpha > 0.4$. Figure 4 shows that the robust methods, not only work well in impulsive environments but also when the noise is Gaussian.

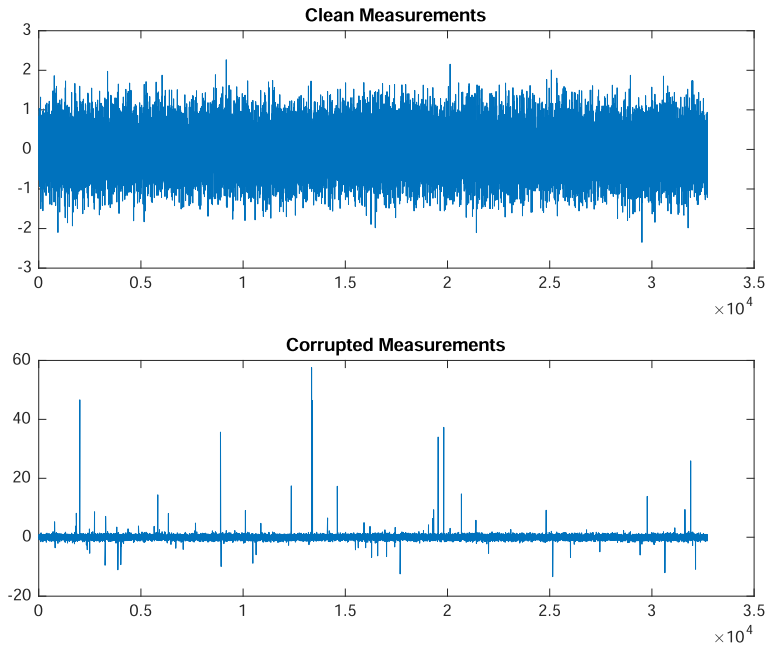

Fig. 5 Example of a 256 × 256 image sampled by a random DCT ensemble with $m = 32718$. The measurements are corrupted by α-stable noise with $\alpha = 1$ and $\sigma = 0.01$. *Top* clean measurements. *Bottom* corrupted measurements

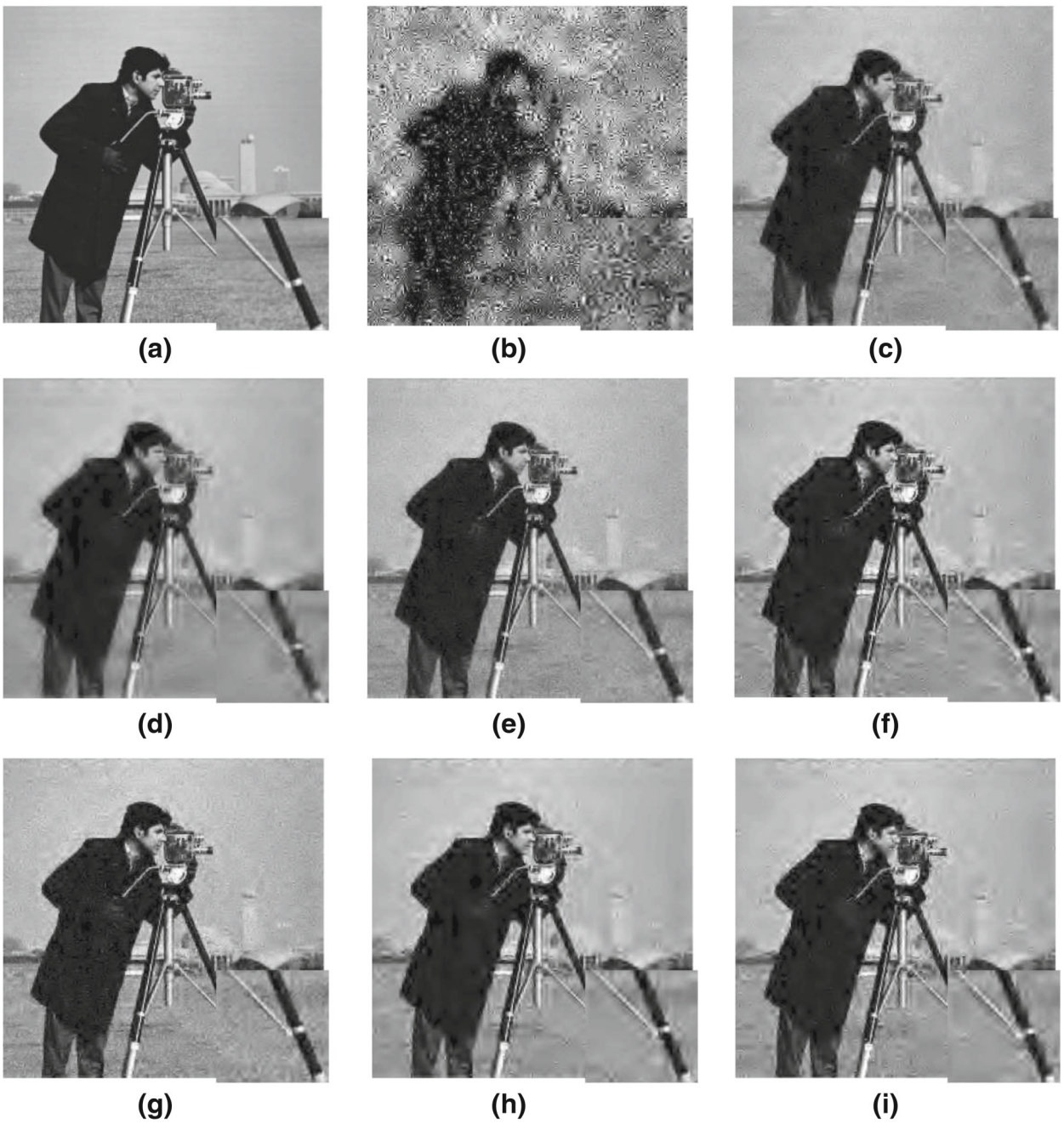

Fig. 6 Cameraman image reconstruction (with zoomed of the bottom right area) example from Cauchy corrupted measurements. **a** Original image. **b** Reconstructed image using ℓ_1-LS (SER = 7.3 dB). **c** Reconstructed image using ℓ_1-LAD (SER = 19.6 dB). **d** Reconstructed image using R-Lasso (SER = 18.1 dB). **e** Reconstructed image using LBP (SER = 20.7 dB). **f** Reconstructed image using LIHT (SER = 19.4 dB). **g** Reconstructed image using ℓ_1-CD (SER = 20.3 dB). **h** Reconstructed image using L-CD (SER = 19.2 dB). **i** Reconstructed image using Huber-IHT (SER = 19.4 dB)

The last experiment shows the performance of the reviewed methods in the reconstruction of the cameraman image of size 256×256 from a set contaminated measurements. We take $m = 32,768$ measurements, i.e., 50 % undersampling, acquired using a random DCT ensemble, and we used the Daubechies *db*4 wavelet as sparsity representation basis. The measurements are contaminated with α-stable noise, having tail parameter $\alpha = 1$ and scale parameter $\sigma = 0.01$. The ℓ_1-OMP method is not included in this experiment due to its high computational cost when dealing with compressible high dimensional signals.

In Fig. 5, the top shows the clean random measurements obtained with the DCT ensemble and the bottom

shows the measurements contaminated with heavy-tailed noise. Results of the image reconstruction are depicted in Fig. 6. Figure 6a shows the original image, and Fig. 6b–i shows the reconstructions for eight methods: ℓ_1-LS, ℓ_1-LAD via ADMM, R-Lasso, LBP, LIHT, ℓ_1-CD, L-CD, and HIHT. Note that the ℓ_1-LS generates several artifacts, and the image is not correctly reconstructed. The ℓ_1-LAD via ADMM, ℓ_1-CD, and LBP methods generate images with better quality than R-Lasso, LIHT, L-CD, and HIHT such that even small details are preserved. A summary of the performance of the methods for this experiment is given in terms of SER (in dB) and execution times (in s), in columns 4 and 5 of Table 1, respectively.

Note that the convex methods, namely, ℓ_1-LAD and R-Lasso, are fast and offer a good computational efficiency since there have been a lot of recent efforts in solving large-scale convex problems [84]. Also, these methods enjoy the rich theoretical guarantees for convex problems. The rest of the methods are based either on nonconvex cost functions or nonconvex constraint sets, thus only convergence to a local minimum can be guaranteed. Also note that all methods, except the coordinate descent methods, ℓ_1-CD and L-CD, and the ℓ_1-OMP method, do not need to explicitly form the sensing matrix Φ but only need functions that implement the matrix-vector multiplication by Φ and Φ^T at each iteration. Thus, if fast implementations are available for such functions, the computational complexity of the algorithms can be largely reduced. On the other hand, the coordinate descent methods are not computationally efficient because only one coordinate is estimated at each iteration and an explicit representation of the matrix Φ is needed. However, these methods offer scalability when the sensing matrix is very large and can only be accessed one row per iteration. Also, fast methods have been proposed where only those coordinates with larger influence in the residuals are estimated at each iteration [83]. The ℓ_1-OMP method is not computationally efficient for high dimensional signals because it needs an explicit representation of the matrix Φ in order to perform the ℓ_1 correlation with every column of the sensing matrix at each iteration of the algorithm. Recall that computing an ℓ_1 correlation between two vectors involves solving an scalar regression problem.

Regarding implementation issues, most methods have free parameters to tune in order to yield a good performance. The greedy methods, such as LIHT, HIHT, and ℓ_1-OMP, are sensitive to know the correct sparsity level a priori. The other methods do not require prior assumptions of the degree of sparsity. The Lorentzian-based methods, namely LBP and LIHT, are sensitive to finding a good initial estimate of the scale parameter whereas the HIHT method estimates both the signal and the scale parameter of the cost function. The ℓ_1-CD method depends a lot on the number of iterations and a rate decay parameter that has to be fixed beforehand. The HIHT method relies on a good tuning of the c constant to get a good performance.

6 Conclusions

We presented a review of robust sparse reconstruction strategies in CS when the compressive measurements are corrupted by outliers or impulsive noise. The reviewed methods are based on employing M-estimators as data fitting terms and include greedy and optimization-based approaches to solve the inverse problems. The robust methods are shown to outperform existing CS techniques (that traditionally use ℓ_2 norms for data fitting) when the measurements have gross errors, while having similar performance in light-tailed environments.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Signal Processing Laboratory (LTS5), Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland. [2]Universidad Industrial de Santander, Bucaramanga, Colombia. [3]Department of Electrical and Computer Engineering, University of Delaware, 19716 Newark, DE, USA. [4]US Army Research Laboratory, 20783 Adelphi, MD, USA.

Received: 30 April 2016 Accepted: 22 September 2016
Published online: 19 October 2016

References

1. DL Donoho, Compressed sensing. IEEE Trans. Inf. Theory. **52**(4), 1289–1306 (2006)
2. EJ Candès, T Tao, Near-optimal signal recovery from random projections: universal encoding strategies? IEEE Trans. Inf. Theory. **52**(12), 5406–5425 (2006)
3. EJ Candès, in *Proceedings, Int. Congress of Mathematics*. Compressive sampling (European Mathematical Society, Madrid, 2006), pp. 1433–1452
4. EJ Candès, MB Wakin, An introduction to compressive sampling. IEEE Signal Proc. Mag. **25**(2), 21–30 (2008)
5. M Fornasier, H Rauhut, in *Handbook of Mathematical Methods in Imaging*, ed. by O Scherzer. Compressed sensing (Springer, New York, 2011)
6. CG Graff, EY Sidky, Compressive sensing in medical imaging. Appl. Opt. **54**(8), 23–44 (2015)
7. RE Carrillo, JD McEwen, Y Wiaux, PURIFY: a new approach to radio-interferometric imaging. Mon. Not. R. Astron. Soc. **439**(4), 3591–3604 (2014)
8. LC Potter, E Ertin, JT Parker, M Cetin, Sparsity and compressed sensing in radar imaging. Proc. IEEE. **98**(6), 1006–1020 (2010)
9. GR Arce, DJ Brady, L Carin, H Arguello, DS Kittle, Compressive coded aperture spectral imaging: an introduction. IEEE Signal Proc. Mag. **31**(1), 105–115 (2014)
10. YC Eldar, G Kutyniok, *Compressed sensing: theory and applications*. (Cambridge University Press, Cambridge, 2012)
11. SA Kassam, HV Poor, Robust techniques for signal processing: a survey. Proc. IEEE. **73**(3), 433–481 (1985)
12. A Swami, B Sadler, On some detection and estimation problems in heavy-tailed noise. Signal Process. **82**(12), 1829–1846 (2002)
13. JG Gonzales, J-L Paredes, GR Arce, Zero-order statistics: a mathematical framework for the processing and characterization of very impulsive signals. IEEE Trans. Signal Process. **54**(10), 3839–3851 (2006)
14. A Zoubir, V Koivunen, Y Chakhchoukh, M Muma, Robust estimation in signal processing. IEEE Signal Proc. Mag. **29**(4), 61–80 (2012)
15. KE Barner, GR Arce (eds.), *Nonlinear signal and image processing: theory, methods, and applications* (CRC Press, Boca Raton, 2004)
16. GR Arce, *Nonlinear signal processing: a statistical approach*. (Wiley, New York, 2005)

17. PJ Huber, *Robust statistics*. (Wiley, New York, 1981)
18. F Hampel, E Ronchetti, P Rousseeuw, W Stahel, *Robust statistics: the approach based on influence functions*. (Wiley, New York, 1986)
19. RA Maronna, RD Martin, VJ Yohai, WA Stahel, *Robust statistics: theory and methods*. (Wiley, New York, 2006)
20. H Wang, G Li, G Jiang, Robust regression shrinkage and consistent variable selection through the LAD-Lasso. J. Bus. Econ. Stat. **25**(3), 347–355 (2007)
21. B Popilka, S Setzer, G Steidl, Signal recovery from incomplete measurements in the presence of outliers. Inverse Probl. Imaging. **1**(4), 661–672 (2007)
22. J Wright, AY Yang, A Ganesh, S Sastryand, Y Ma, Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. **31**(2), 210–227 (2009)
23. J Laska, M Davenport, RG Baraniuk, in *Proc. IEEE Asilomar Conference on Signals, Systems and Computers*. Exact signal recovery from sparsely corrupted measurements through the pursuit of justice (IEEE, Pacific Grove, 2009)
24. J Wright, Y Ma, Dense error correction via ℓ_1-minimization. IEEE Trans. on Inf. Theory. **56**(7), 3540–3560 (2010)
25. Z Li, F Wu, J Wright, in *Proc. Data Compression Conference*. On the systematic measurement matrix for compressed sensing in the presence of gross errors (IEEE, Snowbird, 2010)
26. C Studer, P Kuppinger, G Pope, H Bolcskei, Recovery of sparsely corrupted signals. IEEE Trans. Inf. Theory. **58**(5), 3115–3130 (2012)
27. X Li, Compressed sensing and matrix completion with constant proportion of corruptions. Constr. Approx. **37**(1), 73–99 (2013)
28. NH Nguyen, TD Tran, Exact recoverability from dense corrupted observations via ℓ_1-minimization. IEEE Trans. Inf. Theory. **59**(4), 3540–3560 (2013)
29. NH Nguyen, TD Tran, Robust lasso with missing and grossly corrupted observations. IEEE Trans. Inf. Theory. **59**(4), 2036–2058 (2013)
30. C Studer, R Baraniuk, Stable restoration and separation of approximately sparse signals. Appl. Comput. Harmon. Anal. **37**(1), 12–35 (2014)
31. R Foygel, L Mackey, Corrupted sensing: novel guarantees for separating structured signals. IEEE Trans. Inf. Theory. **60**(2), 1223–1247 (2014)
32. M McCoy, J Tropp, Sharp recovery bounds for convex demixing, with applications. Found. Comput. Math. **14**(3), 503–567 (2014)
33. EJ Candès, X Li, Y Ma, J Wright, Robust principal component analysis? J. ACM. **58**(3), 11–11137 (2011)
34. Y Jin, BD Rao, in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* Algorithms for robust linear regression by exploiting the connection to sparse signal recovery (IEEE, Dallas, TX, 2010), pp. 3830–3833
35. G Mateos, G Giannakis, Robust nonparametric regression via sparsity control with application to load curve data cleansing. IEEE Trans. Signal Process. **60**(4), 1571–1584 (2012)
36. K Mitra, A Veeraraghavan, R Chellappa, Analysis of sparse regularization based robust regression approaches. IEEE Trans. Signal Process. **61**(5), 1249–1257 (2013)
37. G Papageorgiou, P Bouboulis, S Theodoridis, K Themelis, in *Proc. Eur. Signal Processing Conf. (EUSIPCO)*. Robust linear regression analysis: the greedy way (IEEE, Lisbon, 2014)
38. G Papageorgiou, P Bouboulis, S Theodoridis, Robust linear regression analysis—a greedy approach. IEEE Trans. Signal Process. **63**(15), 3872–3887 (2015)
39. EJ Candès, T Tao, Decoding by linear programming. IEEE Trans. Inf. Theory. **51**(12), 4203–4215 (2005)
40. EJ Candès, P Randall, Highly robust error correction by convex programming. IEEE Trans. Inf. Theory. **54**(7), 2829–2840 (2008)
41. RE Carrillo, KE Barner, TC Aysal, Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise. IEEE J. Sel. Topic Signal Process. **4**(2), 392–408 (2010)
42. RE Carrillo, KE Barner, Lorentzian iterative hard thresholding: robust compressed sensing with prior information. IEEE Trans. Signal Process. **61**(19), 4822–4833 (2013)
43. A Ramirez, GR Arce, D Otero, J-L Paredes, B Sadler, Reconstruction of sparse signals from l1 dimensionality-reduced Cauchy random projections. IEEE Trans. Signal Process. **60**(11), 5725–5737 (2012)
44. J-L Paredes, GR Arce, Compressive sensing signal reconstruction by weighted median regression estimates. IEEE Trans. Signal Process. **59**(3), 2585–2601 (2011)
45. H Zhang, Y Chi, Y Liang, in *Proc. International Conference on Machine Learning*. Provable non-convex phase retrieval with outliers: median truncated Wirtinger flow (International Machine Learning Society, New York, 2016). http://jmlr.org/proceedings/papers/v48/
46. M Filipovic, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process*. Reconstruction of sparse signals from highly corrupted measurements by nonconvex minimization (IEEE, Florence, 2014)
47. F Wen, Y Liu, RC Qui, W Yu, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process*. Robust sparse recovery for compressive sensing in impulsive noise using ℓ_p-norm model fitting, (Shangai, 2016)
48. W-J Zeng, HC So, X Jiang, Outlier-robust greedy pursuit algorithms in ℓ_p-space for sparse approximation. IEEE Trans. Signal Process. **64**(1), 60–75 (2016)
49. SA Razavi, E Ollila, V Koivunen, in *Proc. Eur. Signal Processing Conf. (EUSIPCO)*. Robust greedy algorithms for compressed sensing (IEEE, Bucharest, 2012)
50. E Ollila, HJ Kim, V Koivunen, in *Proc. Int. Symp. Comm., Control and Signal Processing*. Robust iterative hard thresholding for compressed sensing (IEEE, Athens, 2014)
51. DS Pham, S Venkatesh, Improved image recovery from compressed data contaminated with impulsive noise. IEEE Trans. Image Process. **21**(1), 397–405 (2012)
52. DS Pham, S Venkatesh, Efficient algorithms for robust recovery of images from compressed data. IEEE Trans. Image Process. **22**(12), 4724–4737 (2013)
53. RE Carrillo, TC Aysal, KE Barner, in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process*. Bayesian compressed sensing using generalized Cauchy priors (IEEE, Dallas, 2010)
54. J Shang, Z Wang, Q Huang, A robust algorithm for joint sparse recovery in presence of impulsive noise. IEEE Signal Proc. Lett. **22**(8), 1166–1170 (2015)
55. G Mateos, G Giannakis, Robust PCA as bilinear decomposition with outlier-sparsity regularization. IEEE Trans. on Signal Process. **60**(10), 5176–5190 (2012)
56. EJ Candès, The restricted isometry property and its implications for compressed sensing. Compte Rendus de l'Academie des Sciences, Paris, Series I. **346**, 589–592 (2008)
57. R Baraniuk, M Davenport, R DeVore, M Walkin, A simple proof of the restricted isometry property for random matrices. Constr. Approx. **28**(3), 253–263 (2008)
58. R Tibshirani, Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B Methodol. **58**(1), 267–288 (1996)
59. A Beck, M Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009)
60. JA Tropp, AC Gilbert, Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans. Inf. Theory. **53**(12), 4655–4666 (2007)
61. T Tony, L Wang, Orthogonal matching pursuit for sparse signal recovery with noise. IEEE Trans. Inf. Theory. **57**(7), 4680–4688 (2011)
62. D Needell, R Vershynin, Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. Found. Computat. Math. **09**(3), 317–334 (2009)
63. T Blumensath, ME Davies, Iterative hard thresholding for compressed sensing. Appl. Comp. Harm. Analys. **27**(3), 265–274 (2009)
64. W Li, JJ Swetits, The linear l1-estimator and the Huber M-estimator. SIAM J. Optimiz. **8**(2), 457–475 (1998)
65. TE Dielman, Least absolute value regression: recent contributions. J. Statist. Computat. Simul. **75**(4), 263–286 (2005)
66. DRK Brownrigg, The weighted median filter. Commun. ACM. **27**(8), 807–818 (1984)
67. GR Arce, A general weighted median filter structure admitting negative weights. IEEE Trans. Signal Process. **46**(12), 3195–3205 (1999)
68. M Shao, CL Nikias, Signal processing with fractional lower order moments: stable processes and their applications. Proc. IEEE. **81**(7), 986–1010 (1993)
69. MA Arcones, Lp-estimators as estimates of a parameter of location for a sharp-pointed symmetric density. Scand. J. Stat. **25**(4), 693–715 (1998)
70. W-J Zeng, H-C So, L Huang, ℓ_p-MUSIC: Robust direction-of-arrival estimator for impulsive noise enviroments. IEEE Trans. Signal Process. **61**(17), 4296–4308 (2013)
71. W-J Zeng, H-C So, AM Zoubir, An ℓ_p-norm minimization approach to time delay estimation in impulsive noise. Digital Signal Process. **23**(4), 1247–1254 (2013)

72. JG Gonzalez, GR Arce, Statistically-efficient filtering in impulsive environments: weighted myriad filters. EURASIP J. Appl. Sign. Proc. **1**, 4–20 (2002)
73. S Kalluri, GR Arce, Adaptive weighted myriad filter algorithms for robust signal processing in α-stable noise environments. IEEE Trans. Signal Process. **46**(2), 322–334 (1998)
74. S Kalluri, GR Arce, Fast algorithms for weighted myriad computation by fixed-point search. IEEE Trans. Signal Process. **48**(1), 159–171 (2000)
75. RC Nunez, JG Gonzalez, GR Arce, JP Nolan, Fast and accurate computation of the myriad filter via branch-and-bound search. IEEE Trans. Signal Process. **56**(7), 3340–3346 (2008)
76. JG Gonzales, GR Arce, Optimality of the myriad filter in practical impulsive-noise environments. IEEE Trans. Signal Process. **49**(2), 438–441 (2001)
77. RE Carrillo, TC Aysal, KE Barner, A generalized Cauchy distribution framework for problems requiring robust behavior. EURASIP J. Adv. Signal Process. **2010**(Article ID 312989), 19 (2010)
78. TC Aysal, KE Barner, Meridian filtering for robust signal processing. IEEE Trans. Signal Process. **55**(8), 39449–3962 (2007)
79. J Yang, Y Zhang, Alternating direction algorithms for l1-problems in compressive sensing. SIAM J. Sci. Comput. **33**(1), 250–278 (2011)
80. Y Xiao, H Zhu, S-Y Wu, Primal and dual alternating direction algorithms for ℓ_1-ℓ_1 norm minimization problems in compressive sensing. Comput. Optim. Appl. **54**(2), 441–459 (2013)
81. P Gong, C Zhang, Z Lu, J Huang, J Ye, in *Proc. International Conference on Machine Learning*. A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems (International Machine Learning Society, Atlanta, 2013)
82. D Needell, JA Tropp, Cosamp: Iterative signal recovery from incomplete and inaccurate samples. Appl. Compos. Harmon. Anal. **26**(3), 301–321 (2008)
83. AB Ramirez, GR Arce, BM Sadler, in *Proc. 18th European Signal Process. Conf.* Fast algorithms for reconstruction of sparse signals from Cauchy random projections (IEEE, Aalborg, 2010), pp. 432–436
84. V Cevher, S Becker, M Schmidt, Convex optimization for big data: scalable, randomized, and parallel algorithms for big data analytics. IEEE Signal Proc. Mag. **31**(5), 32–43 (2014)

Submit your manuscript to a SpringerOpen journal and benefit from:

- ► Convenient online submission
- ► Rigorous peer review
- ► Immediate publication on acceptance
- ► Open access: articles freely available online
- ► High visibility within the field
- ► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com