Compressive Sensing Signal Reconstruction by Weighted Median Regression Estimates

Jose L. Paredes, *Senior Member, IEEE*, and Gonzalo R. Arce, *Fellow, IEEE*

Abstract—In this paper, we propose a simple and robust algorithm for compressive sensing (CS) signal reconstruction based on the weighted median (WM) operator. The proposed approach addresses the reconstruction problem by solving a l_0-regularized least absolute deviation (l_0-LAD) regression problem with a tunable regularization parameter, being suitable for applications where the underlying contamination follows a statistical model with heavier-than-Gaussian tails. The solution to this regularized LAD regression problem is efficiently computed, under a coordinate descent framework, by an iterative algorithm that comprises two stages. In the first stage, an estimation of the sparse signal is found by recasting the reconstruction problem as a parameter location estimation for each entry in the sparse vector leading to the minimization of a sum of weighted absolute deviations. The solution to this one-dimensional minimization problem turns out to be the WM operator acting on a shifted-and-scaled version of the measurement samples with weights taken from the entries in the measurement matrix. The resultant estimated value is then passed to a second stage that identifies whether the corresponding entry is relevant or not. This stage is achieved by a hard threshold operator with adaptable thresholding parameter that is suitably tuned as the algorithm progresses. This two-stage operation, WM operator followed by a hard threshold operator, adds the desired robustness to the estimation of the sparse signal and, at the same time, ensures the sparsity of the solution. Extensive simulations demonstrate the reconstruction capability of the proposed approach under different noise models. We compare the performance of the proposed approach to those yielded by state-of-the-art CS reconstruction algorithms showing that our approach achieves a better performance for different noise distributions. In particular, as the distribution tails become heavier the performance gain achieved by the proposed approach increases significantly.

Index Terms—Basis selection, compressive sensing, inverse problem, model selection, reconstruction algorithm, robust regression, sparse model, weighted median.

J. L. Paredes is with the Electrical Engineering Department, Universidad de Los Andes, Mérida, 5101 Venezuela (e-mail: paredesj@ula.ve).

G. R. Arce is with the Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716 USA (e-mail: arce@ee.udel.edu)..

I. INTRODUCTION

RECENTLY, the theory of compressive sensing (CS) has emerged as a promising approach that unifies signal sensing and signal compression into a single and simple task [1], [2]. The basic principle behind the CS framework is the use of nonadaptive linear projections to acquire an efficient, dimensionally reduced representation of a sparse signal. From that low-dimension representation, the original sparse signal can be recovered/recontructed by solving an inverse problem. Interestingly, the theory of compressive sensing has shown that by randomly projecting a sparse signal the most salient information is preserved in just a few measurements such that, with high probability, the sparse signal can be recovered from the measurements by solving the inverse problem $\mathbf{Y} = \mathbf{AX} + \boldsymbol{\xi}$ where $\mathbf{Y}$ is an M-dimensional measurement vector, $\mathbf{A}$ is the $M \times N$ measurement matrix with $M \ll N$, $\mathbf{X}$ is the target sparse signal, and $\boldsymbol{\xi}$ is the noise vector.

To this end, several algorithms have been proposed for signal reconstruction. A first class of reconstruction algorithms, tries to identify iteratively the column-vectors in $\mathbf{A}$ that are closer, in the Euclidean distance sense, to the measurement vector $\mathbf{Y}$. Matching pursuit (MP) [3], orthogonal matching pursuit (OMP) [4], stagewise OMP [5], regularized OMP [6], subspace MP [7], and compressive sensing MP (CoSaMP) [8] are examples of greedy-based algorithms that belong to this family. The basic idea behind these reconstruction methods is to find the location of the nonzero values of $\mathbf{X}$ by maximizing the correlation between the columns of $\mathbf{A}$ and a residual vector (measurement vector for the first iteration). Then, the nonzero values of $\mathbf{X}$ are estimated by minimizing the l_2-norm of the residual vector. These two stages (correlation maximization and l_2 minimization) emerge naturally as the optimal processing approaches under the assumption that the underlying contamination follows a Gaussian distribution.

A second class of reconstruction algorithms, known as convex-relaxation algorithms, recovers the sparse signal by solving a constrained l_p optimization problem. Reconstruction algorithms based on interior-point methods [9]–[11], projected gradient methods [12] and iterative thresholding [13], among others, belong to this class. One of these reconstruction algorithms that is particularly useful since it takes into account the fact that in a practical scenario the random measurements may be corrupted by noise is the basis pursuit denoising (BPDN) reconstruction algorithm [14], [15]. In this approach, the sparse signal $\mathbf{X}$ is found as the solution to the optimization problem $\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{AX}\|_{l_2}^2 + \tau \|\mathbf{X}\|_{l_1}$ where τ is a regularization

parameter that controls the influence of the data-fitting term (l_2-term) and the sparsity-inducing term (l_1-term) on the optimal solution. Note that the solution to this optimization problem is, indeed, the maximum *a posteriori* (MAP) estimate for $\mathbf{X}$ under the assumption that the underlying contamination follows a Gaussian model and using a Laplacian model as the sparseness-promoting prior model [16]–[19]. Interestingly, under a Bayesian framework the sparse signal is found by solving a l_1-regularized least square problem whose solution leads to a sparse vector $\mathbf{X}$ that minimizes the l_2-norm induced by the Gaussian assumption.

Despite its Bayesian optimality under a Gaussian noise assumption, the l_1-regularized least square reconstruction algorithm tends to be very sensitive to outliers or gross error present in the measurements. More precisely, if the underlying contamination no longer follows a Gaussian noise model and, instead, it is better characterized by a statistical model with heavier-than-Gaussian tails the performance of these reconstruction algorithms degrades notably due, mainly, to the poor robustness of the least square estimator. Moreover, greedy-based reconstruction algorithms which are optimum under the l_2-norm based assumptions are no longer effective in solving the inverse problem leading to spurious components in the recovered signal. This invites us to explore the rich theory of robust linear regression [20]–[23] as a plausible approach to address the CS reconstruction problem when the random projections are contaminated with heavy-tailed noise.

In this paper, a CS reconstruction algorithm is proposed that combines the robust properties of the least absolute deviation (LAD) regression with the sparsity forced by an l_0-norm. More precisely, we address the reconstruction problem by solving a l_0-regularized LAD regression problem, $\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_{l_1} + \tau\|\mathbf{X}\|_{l_0}$ where τ controls the influence of the l_1-term and the sparsity-inducing term in the optimum solution. Unlike l_1-regularized LS based reconstruction algorithms, the proposed l_0-regularized LAD algorithm offers robustness to a broad class of noise, in particular to heavy tail noise [23], being optimum under the maximum likelihood (ML) principle when the underlying contamination follows a Laplacian-distributed model. Furthermore, the use of l_1-norm in the data-fitting term has been shown to be a suitable approach for image denoising [24], image restoration [25] and sparse signal representation [18], [26], [27].

Solving this l_0-regularized LAD problem, however, is N_p-hard complex owing to the sparsity constraint imposed by the l_0-norm [28]. Therefore, a suboptimal approach based on the coordinate descent framework is adopted here to find the sparsest signal that minimizes the residual LAD. Interestingly, under the coordinate descent approach the solution to the multidimensional l_0-LAD problem can be efficiently computed by an iterative algorithm based on the weighted median operator followed by a hard threshold operator acting on a shifted and scaled version of the observation vector. The former operator which emerges naturally as the optimum solution of a scalar LAD regression problem adds the desired robustness to the estimation of the nonzero values of $\mathbf{X}$ while the latter operator, induced by the sparsity-promoting term, ensures the sparsity of the solution. Furthermore, the regularization parameter,

τ, of the multidimensional l_0-LAD problem becomes the hard-thresholding parameter of the second operation, in turn, it controls whether an entry of the sparse vector is significant or not.

The proposed approach can be thought of as a greedy-based algorithm since at each iteration one or several entries of $\mathbf{X}$, estimated by the WM operator, are tested for relevancy and added to or removed from the supporting set of the unknown sparse signal. As the iterative algorithm progresses, the hard-thresholding parameter is suitably tuned allowing thus to identify in order of descending signal strength the nonzero entries of the sparse vector. Furthermore, a refined estimate of the nonzero entries in $\mathbf{X}$ is carried out at each iteration based on a more reliable samples set—the sample set that results from removing the contribution of previous estimated entries from the measurement samples. Extensive simulations show that the proposed approach outperforms state-of-the-art CS signal reconstruction algorithms for noise model with heavier-than-Gaussian tail distribution while, for Gaussian noise, it yields a comparable performance to those outputted by CoSaMP [8] and convex-relaxation based algorithms [9].

II. PROBLEM FORMULATION

Consider an N-dimensional, K-sparse signal $\mathbf{X}$ that is remotely sensed by projecting it onto an $M \times N$ random basis $\mathbf{A}$. Furthermore, assume that the projections are noisy as described by

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \boldsymbol{\xi} \tag{1}$$

where $\mathbf{Y}$ is an M-dimensional measurement vector with $K < M \ll N$, $\mathbf{A}$ is the measurement matrix with i.i.d. random entries following a Gaussian or a Bernoulli distribution and $\boldsymbol{\xi} = [\xi_1, \xi_2, \ldots, \xi_M]^T$ is the noise vector with i.i.d. components obeying a common distribution $f_\xi(\xi)$. In this model, the noise vector comprises the noise introduced during the sensing and transmission processes including, among other sources, thermal noise, quantization, sensor system, and communication channel. Furthermore, the noise may contain outliers, hence, it is better characterized by a distribution with heavier-than-Gaussian tails.

The aim of a reconstruction algorithm is to recover the sparse signal $\mathbf{X} = [X_1, X_2, \ldots, X_N]^T$ from the reduced set of noisy projections $\{Y_1, Y_2, \ldots Y_M\}$. Since the number of measurements is in general much smaller than the signal dimension, the linear equation system (1) is underdetermined and has multiple solutions. Furthermore, since the measurement signal has been contaminated by noise, the recovery process reduces to finding an approximation of the signal, $\mathbf{X}$, such that a performance criterion is achieved under the constraint that the signal is sparse in some orthonormal basis, $\boldsymbol{\Psi}$.[1] A common criterion widely used in the Compressive Sensing literature is to reconstruct the sparse signal by minimizing a norm of a residual error subject to the constraint that the signal is sparse [1]. Formally [29]

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_{l_p} \quad \text{subject to} \quad \|\mathbf{X}\|_{l_0} \le K$$

---

[1]Without loss of generality, we assume that the signal is sparse in the canonical domain, therefore, the sparsity basis $\boldsymbol{\Psi}$ reduces to the identity matrix.

where $||\cdot||_{l_p}$ denotes the l_p-norm with $||\cdot||_{l_0}$ as the l_0 quasi-norm that outputs the number of nonzero components in its argument. This constrained minimal-norm optimization problem can be reformulated as [29]

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} ||\mathbf{Y} - \mathbf{AX}||_{l_p} + \tau ||\mathbf{X}||_{l_0} \qquad (2)$$

where τ is a regularization parameter that balances the conflicting goals of minimizing the l_p-norm of the residual errors while yielding, at the same time, a sparse solution on $\mathbf{X}$ [17].

The first term in (2), named as data-fitting term, gives us a measurement of a distance between the measurement vector and the projected reconstruction signal $\mathbf{A\hat{X}}$. The smaller the distance, the better the estimation is. The choice of the optimum norm for the data-fitting term is closely related to the characteristic of the noise. Thus, if the underlying contamination follows a Gaussian-distributed model, the Euclidean norm (l_2-norm) is the optimum choice under the Maximum Likelihood principle leading thus to solve a regularized least square (LS) regression problem for the recovery of the sparse signal $\mathbf{X}$. Most CS reconstruction algorithms rely on the assumption that the underlying contamination follows the Gaussian model and use the l_2-norm in the data fitting term [11], [14]–[16]. However, it is well known that the least square based estimators are highly sensitive to outliers present in the measurement vector leading to an unacceptable performance when the noise no longer follows the Gaussian assumption but, instead, it is better characterized by heavier-than Gaussian tail distributions [20].

To mitigate the effect of impulsive noise in the compressive measurements, a more robust norm for the data-fitting term has to be used. A natural choice will be to replace the l_2-norm by an l_1-norm [18], [24]–[27], reducing the reconstruction problem to solve an l_0-regularized LAD (l_0-LAD) regression problem. That is

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} ||\mathbf{Y} - \mathbf{AX}||_{l_1} + \tau ||\mathbf{X}||_{l_0}. \qquad (3)$$

LAD regression has been widely used in the statistical community as an alternative approach to solve linear regression problem when the observation vector is contaminated by outliers or gross errors, being optimum under the ML principle when the underlying contamination follows a Laplacian distribution (see [23] and references therein). We exploit these properties of the LAD estimator in the context of compressive sensing signal reconstruction.

The second term in (3) imposes sparsity in the solution. That is, we are interested in finding the signal $\mathbf{X}$ with the smallest possible number of nonzero components such that the absolute deviation between the observation vector, $\mathbf{Y}$ and the projected sparse signal, $\mathbf{AX}$, is minimized. This leads inevitably to an optimization problem that is N_p-hard whose direct solution even for modest-sized signal is unfeasible [28].

A. Iterative Coordinate Descent Approach for Solving the l_0-regularized LAD

To overcome the computational complexity of solving the N-dimensional optimization problem (3), a suboptimum approach based on the coordinate descent (CD) framework is used here. Under the CD framework, the solution of (3) is found by considering it as N scalar minimization subproblems, solve each single-variable minimization subproblem at once by keeping the other variables fixed to the value obtained in a previous iteration and repeat this procedure until a performance criterion is achieved. Coordinate descent methods have been widely used as a powerful optimization tool to solve multivariable optimization problems [30], [31]. More recently, it has been proposed for solving regularized linear regression problems in [32], [33] and for CS signal reconstruction in [34]. The key to the success of this framework relies on the fact that it is much easier to solve a one-dimensional minimization problem than a multidimensional one. Furthermore, if a closed form expression is available for the solution of the single-variable minimization subproblem, the multidimensional optimization problem reduces to update each component of the signal in an optimal fashion way. We exploit these features of the coordinate descent approach in deriving a robust CS signal reconstruction algorithm that tries to solve the l_0-LAD minimization problem.

Before ending this section, it should be pointed out that a second approach to reduce the computational complexity involved in solving the l_0-LAD optimization problem (3) is to approximate the l_0-norm in the sparsity-forcing term by a convex norm that at the same time induces sparsity. This convex-relaxation approach has been widely used as a feasible path to solve the l_0-regularized LS problem in [9]–[13]. Following this line of thought, the reconstruction problem reduces then to the following l_1-regularized LAD [18], [25], [26], [35], [36]: $\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} ||\mathbf{Y} - \mathbf{AX}||_{l_1} + \tau ||\mathbf{X}||_{l_1}$.

Interestingly, the solution to this optimization problem turns out to be the maximum *a posteriori* (MAP) estimate when the underlying contamination model and the sparseness-promoting prior model are both assumed to be Laplacian [37]–[39]. Although this l_1-LAD optimization problem can be efficiently solved by linear programming technics [18], [26], the use of Laplacian model as a sparsity-promoting model may not induce the necessary sparsity in the solution. Furthermore, as was shown in [40], to achieve the same performance in signal recovery, solving a convex-relaxation problem like l_1-regularized LAD requires more measurements than those needed by nonconvex optimization at the cost of more computational complexity.

III. THE ONE-DIMENSIONAL LAD PROBLEM

Before deriving a closed form expression for the solution of the regularized LAD problem, let us review briefly the weighted median operator since it emerges naturally as the underlying operation in the solution of the l_p-regularized LAD scalar problems.

A. Overview of Weighted Median Operator

The weighted median (WM) operator has deep roots in statistical estimation theory since it emerges as the maximum likelihood (ML) estimator of location derived from a set of independent samples obeying a Laplacian distribution. Consider an N-dimensional observation vector that follows the linear model: $\mathbf{z} = \alpha\mathbf{1} + \boldsymbol{\nu}$, where $\mathbf{z} = [z_1, z_2, \ldots, z_N]^T$ is the observation vector, $\mathbf{1} = [1, 1, \ldots, 1]^T$, $\alpha \in R$ is an unknown location parameter and $\boldsymbol{\nu}$ is an N-dimensional vector

that denotes additive white noise. Assume that the noise vector satisfies $\nu = \Lambda u$ where $\Lambda = \text{diag}\left(\frac{\sqrt{2}}{\lambda_1}, \frac{\sqrt{2}}{\lambda_2}, \ldots, \frac{\sqrt{2}}{\lambda_N}\right)$ is a diagonal matrix of known entries and u are independent and identically distributed (i.i.d.) samples obeying a zero-mean, unit-variance Laplacian distribution. This model allows for the possibility of heteroscedasticity in the noise vector, thus, each element in the observation vector follows a Laplacian distribution with a common location parameter, α and a (possibly) unique variance, $\sigma_i^2 = \frac{2}{\lambda_i^2}$.

Under the ML criterion the best estimate of α is the one for which the likelihood function reaches its maximum. That is,

$$\hat{\alpha} = \arg\max_{\alpha} f(\mathbf{z}; \alpha)$$
$$= \arg\max_{\alpha} \prod_{i=1}^{N} \frac{\lambda_i}{2} \exp\left(-\lambda_i |z_i - \alpha|\right). \tag{4}$$

This is equivalent to minimizing the sum of weighted absolute deviations

$$\hat{\alpha} = \arg\min_{\alpha} \sum_{i=1}^{N} W_i |z_i - \alpha| \tag{5}$$

where $W_i = \lambda_i$ denotes the weight associated to the ith observation sample. It represents a measurement of its influence on the estimation of location parameter. Thus, as $W_i \to 0$ the influence of the ith observation sample in the minimization problem becomes practically negligible, while as W_i becomes larger the value of α that minimizes (5) is pulled toward the ith observation sample.

Note, in (5), that the cost function to be minimized is a piecewise linear and convex function, therefore the value of α minimizing (5) is guaranteed to be one of the observation samples. It turns out that the solution to the minimization problem given by (5) is the weighted median operator originally introduced in [41] and defined as

$$\hat{\alpha} = \text{MEDIAN}\left(W_1 \diamond z_1, W_2 \diamond z_2, \ldots, W_N \diamond z_N\right) \tag{6}$$

where the symbol $\diamond$ denotes the replication operator defined as $W_i \diamond z_i = \overbrace{z_i, z_i, \ldots, z_i}^{W_i \text{ times}}$. Thus, the weighted median operator reduces to: replicating the ith observation sample W_i times, sorting the samples of the new expanded set and choosing the median value from the new set. Although this definition seems to constrain the weights to take on positive integer values, a more general definition of WM operator that admits positive real-valued weights exists which turns out to be computationally more efficient [42]–[44]. Appendix A presents a pseudocode of an efficient implementation [45] to compute the WM operator based on a fast computation of the sample median [46], [47].

It is worth mentioning that although the WM operator emerges naturally as the ML location estimation when the underlying contamination follows a Laplacian distribution, its inherent robustness makes it suitable for applications when the noise model has tails heavier than those of a normal distribution [42]–[44], [48]–[50].

B. Regularized LAD Problem: One-Dimensional Case

Consider the following one-dimensional minimization problem:

$$\arg\min_{\alpha} \sum_{i=1}^{M} |Y_i - a_i\alpha| + \tau_p |\alpha|_{l_p} \tag{7}$$

where $|\cdot|_{l_p}$ denotes the l_p-norm for scalar values,[2] Y_i is the ith observation sample, $\mathbf{a} = [a_1, a_2, \ldots, a_M]^T$ is a vector of known coefficients and τ_p is a regularization parameter whose effects in the minimization problem will be seen shortly. Note that in the context of linear regression theory, Y_i's are the measured variables, a_i's are the predictor variables and α is the unknown regression coefficient. We are interested in finding the value of α that minimizes the LAD subject to the constraints imposed by the l_p-term. The following theorems give closed form expressions for the solution of the optimization problem (7) for l_0 and l_1 constraint terms.

Theorem 1 [l_0-LAD]: The solution to the minimization problem

$$\hat{\alpha} = \arg\min_{\alpha} \sum_{i=1}^{M} |Y_i - a_i\alpha| + \tau_0 |\alpha|_{l_0} \tag{8}$$

is given by the equation at the bottom of the page, where $\mathbf{Y} = [Y_1, Y_2, \ldots, Y_M]^T$ and $\mathbf{a} = [a_1, a_2, \ldots, a_M]^T$.

Proof: Let $G_1(\alpha; \mathbf{Y}, \mathbf{a})$ and $G_2(\alpha; \mathbf{Y}, \mathbf{a})$ be, respectively, the cost functions $\sum_{i=1}^{M} |Y_i - a_i\alpha|$ and $\sum_{i=1}^{M} |Y_i - a_i\alpha| + \tau|\alpha|$. Note that $G_1(\alpha; \mathbf{Y}, \mathbf{a})$ can be rewritten as: $G_1(\alpha; \mathbf{Y}, \mathbf{a}) = \sum_{i=1}^{M} |a_i| \left|\frac{Y_i}{a_i} - \alpha\right|$. Since G_1 is a piecewise linear and convex function, it follows that it reaches its minimum at the weighted median sample [41], [42] of the scaled observation samples $\{\frac{Y_i}{a_i}|_{i=1}^{M}\}$ with weights $|a_i|, i = 1, 2, \ldots, M$. This is,

$$\tilde{\alpha} = \arg\min_{\alpha} G_1(\alpha; \mathbf{Y}, \mathbf{a}) = \text{MEDIAN}\left(|a_i| \diamond \left.\frac{Y_i}{a_i}\right|_{i=1}^{M}\right).$$

Furthermore, since $G_2 = G_1 + \tau_0$ for all $\alpha \neq 0$ and $G_2 = G_1$ for $\alpha = 0$, it is clear that the cost function G_2 has two local minimums, one at zero and one at $\tilde{\alpha}$. The global minimum is found by evaluating the cost function G_2 at these local minimums and chose the one that yields the smaller cost value. If a tie occurs,

[2] The l_0-norm for a scalar value is defined as $|\alpha|_{l_0} = 1$ if $\alpha \neq 0$, 0 otherwise.

$$\hat{\alpha} = \begin{cases} \tilde{\alpha} = \text{MEDIAN}\left(|a_i| \diamond \left.\frac{Y_i}{a_i}\right|_{i=1}^{M}\right) & \text{if } \|\mathbf{Y}\|_{l_1} - \|\mathbf{Y} - \tilde{\alpha}\mathbf{a}\|_{l_1} > \tau_0 \\ 0 & \text{Otherwise} \end{cases}$$
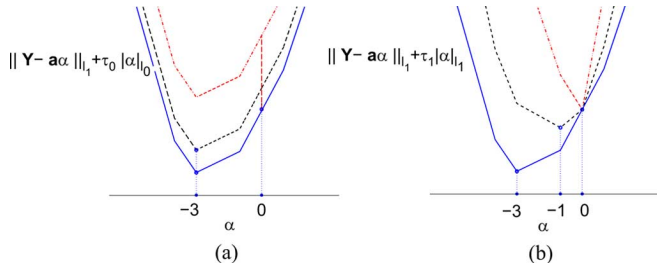
Fig. 1. Cost functions for the regularized optimization problems (a) l_0-LAD and (b) l_1-LAD as a function of α and the regularization parameter τ_p. The observation set $\{Y_i\}_{i=1}^{M}$ is $\{0.45, -3.00, 0.40, -0.50, -2.10, 0.10\}$ and the weight vector $\mathbf{a}$ is $[0.15, 0.75, 0.40, 0.50, 0.70, 0.05]^T$. Solid line $\tau_p = 0$, dashed line $\tau_p = 0.75$ and dash-dotted line $\tau_p = 2.50$. $p \in \{0, 1\}$.

i.e., $G_2(\tilde{\alpha}) = G_2(0)$, we favor the sparse solution. Thus, $\hat{\alpha} = \tilde{\alpha}$ if $G_2(\tilde{\alpha}) < G_2(0) = \|\mathbf{Y}\|_{l_1} - \|\mathbf{Y} - \tilde{\alpha}\mathbf{a}\|_{l_1} > \tau_0$, otherwise $\hat{\alpha} = 0$. ∎

Interestingly, Theorem 1 states that the solution to the l_0-regularized LAD problem can be thought of as a two-stage operation. First, an estimate of the parameter that minimizes the LAD is found by the weighted median operator and, then, a hard thresholding operator is applied on the estimated value. Note that the estimation stage has been reformulated as a location parameter estimation problem for which closed form solution exist. Note also that the hard thresholding parameter is equal to the regularization parameter, τ_0. Thus, large values for the regularization parameter force the solution of (8) to become zero while small value for τ_0 yields the weighted median sample as the solution of the l_0-LAD minimization problem.

For comparison purposes, let's consider the optimization problem (7) where the sparsity is induced by an l_1-term. The following theorem gives the solution to the one-dimension l_1-LAD problem.

Theorem 2 [l_1-LAD]: The solution to the minimization problem

$$\breve{\alpha} = \arg\min_{\alpha} \sum_{i}^{M} |Y_i - a_i\alpha| + \tau_1 |\alpha|_{l_1} \qquad (9)$$

is given by

$$\breve{\alpha} = \text{MEDIAN}\left(\tau_1 \diamond 0, |a_1| \diamond \frac{Y_1}{a_1}, \ldots, |a_M| \diamond \frac{Y_M}{a_M},\right)$$

Proof: It is easy to see that the regularization term $\tau_1|\alpha|$ can be merged into the summation term by augmenting the observation set with an additional observation that takes on a zero value and whose weight is the regularization parameter τ_1. The proof follows since the weighted median operator minimizes the weighted LAD of the augmented data set. ∎

Note that the regularization process induced by the l_1-term leads to a weighting operation on a zero-valued sample in the WM operation. Thus, large values for the regularization parameter implies a large value for the weight corresponding to the zero-valued sample. This, in turns, pulls the WM estimation toward zero favoring sparsity. On the other hand, small values for the regularization parameter implies less influence of the zero-

valued sample on the estimation of α leading to a WM output driven by the observation vector, $\mathbf{Y} = [Y_1, Y_2, \ldots, Y_M]^T$ and the weight vector $\mathbf{a} = [a_1, a_2, \ldots, a_M]^T$.

To better illustrate the effect of the regularization term on the optimization problems (8) and (9), Fig. 1 depicts the cost functions as a function of the parameter α for several values for the regularization parameter τ_p. As can be seen in Fig. 1(a), for $\tau \neq 0$ the cost function has two local minimums one at $\alpha = 0$ and one at $\alpha = \tilde{\alpha}$, that is, the value at which the cost function $G_1(\alpha; \mathbf{Y}, \mathbf{a})$ reaches its minimum. The global minimum is selected from these two local minimums to yield the solution to the minimization problem (8). Note also that the l_0-term in (8) shifts vertically the cost function without changing the location of the local minimum.

Fig. 1(b) shows the effects of the regularization parameter on the cost function of the optimization problem (9). As is shown in Fig. 1(b), the estimate is pulled to zero as the regularization parameter that, in turns, defines the weight for the zero-value sample, is increased. Note also that the regularization term produces a shift of the cost function along the horizontal direction changing gradually the location where it reaches its minimum as τ_1 increases. This shifted effect introduces a bias on the estimation of α much like the bias observed in a median based estimator when one of the samples is contaminated by an outlier, the middle order statistic will move in the direction of the outlier [51, p. 90].

IV. CS SIGNAL RECONSTRUCTION BY SOLVING l_0-LAD: A COORDINATE DESCENT APPROACH

Consider the CS reconstruction algorithm where the sparse signal is recovered by minimizing the sum of absolute deviation of the residual $(\mathbf{Y} - \mathbf{AX})_i$, $i = 1, 2, \ldots, N$ subject to the constraint imposed by the sparsity-inducing term. More precisely, we recover the sparse signal by solving

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \sum_{i=1}^{M} |(\mathbf{Y} - \mathbf{AX})_i| + \tau_0 \|\mathbf{X}\|_{l_0}. \qquad (10)$$

Note that having the LAD as the data-fitting term, it is expected that the resulting estimator be robust against noise [20], [23] and also enjoy a sparse representation since it combines the LAD criterion and the sparsity-forced regularization.

To solve this optimization problem under the coordinate descent framework, let's proceed as follows. Assume that we want to estimate the nth entry of the sparse vector $\mathbf{X}$ while keeping the other entries fixed. Furthermore, assume for now that the other entries are known or have been previously estimated. Thus, the N-dimensional optimization problem of (10) reduces to the following single variable minimization problem

$$\hat{X}_n = \arg\min_{X_n} \sum_{i=1}^{M} \left| Y_i - \sum_{\substack{j=1 \\ j \neq n}}^{N} A_{ij}X_j - A_{in}X_n \right|$$

$$+ \tau_0 |X_n|_{l_0} + \tau_0 \sum_{\substack{j=1 \\ j \neq n}}^{N} |X_j|_{l_0}, \qquad (11)$$

which can be rewritten as

$$\hat{X}_n = \arg\min_{X_n} \sum_{i=1}^{M} |A_{in}| \left| \frac{Y_i - \sum_{j=1,j\neq n}^{N} A_{ij}X_j}{A_{in}} - X_n \right| + \tau_0 |X_n|_{l_0} \quad (12)$$

provided that none of the entries in the nth column of the measurement matrix $\mathbf{A}$ is zero. Note that if one of these entries is zero, then the corresponding summand in (12) can be dropped since it becomes a constant (independent of X_n). Note also that the last summation of (11) has been dropped since it does not depend on X_n.

Upon closer examination of (12), it can be noticed that the first term on the right-hand side is a sum of the weighted absolute deviations, where $\frac{Y_i - \sum_{j=1,j\neq n}^{N} A_{ij}X_j}{A_{in}}$ for $i = 1, 2, \ldots M$ are the data samples, $|A_{in}|$, $i = 1, 2, \ldots, M$ are the weights and X_n plays the role of a location parameter. From Theorem 1, the solution to this optimization problem is found by computing a median based estimator

$$\tilde{X}_n = \text{MEDIAN} \left(|A_{in}| \diamond \left| \frac{Y_i - \sum_{j=1,j\neq n}^{N} A_{ij}X_j}{A_{in}} \right|_{i=1}^{M} \right) \quad (13)$$

followed by the hard threshold operator

$$\hat{X}_n = \begin{cases} \tilde{X}_n & \text{if } \|\mathbf{r}_n\|_{l_1} - \|\mathbf{r}_n - \mathbf{A}_n \tilde{X}_n\|_{l_1} > \tau_0 \\ 0 & \text{otherwise} \end{cases}$$

where $\mathbf{r}_n = \mathbf{Y} - \sum_{j=1,j\neq n}^{N} \mathbf{A}_j X_j$ is the nth residual term that remains after subtracting from the measurement vector the contribution of all the components but the nth one. In (13), $\mathbf{A}_n$ denotes the nth column-vector of the measurement matrix $\mathbf{A}$.

Interestingly, the estimation of the nth component of the sparse vector reduces to the weighted median operator where the weights are the entries of the nth column of the measurement matrix and the observation samples are a shifted-and-scaled version of the measurement vector. Furthermore, the regularization process induced by the sparsity term leads to a hard thresholding operation on the estimated value.

More interesting, the nth entry is considered relevant and, hence, not forced to zero, if it leads to a significant reduction in the l_1-norm of the residual signal along the nth coordinate. That reduction in the nth residual term has to be larger than the regularization parameter, τ_0. Further simplifications of (13) lead us to an interesting observation. First, note that

$$\begin{aligned} \tau_0 &< \|\mathbf{r}_n\|_{l_1} - \|\mathbf{r}_n - \mathbf{A}_n \tilde{X}_n\|_{l_1} \\ &= \|\mathbf{A}_n X_n\|_{l_1} - \|\mathbf{A}_n X_n - \mathbf{A}_n \tilde{X}_n\|_{l_1} \\ &= \|\mathbf{A}_n\|_{l_1} \left(|X_n| - |X_n - \tilde{X}_n| \right) \\ &\leq \|\mathbf{A}_n\|_{l_1} |\tilde{X}_n| \\ &\approx M\sigma\sqrt{\frac{2}{\pi}} |\tilde{X}_n| \end{aligned} \quad (14)$$

where we have assumed that the entries $A_{i,j}$ of the measurement matrix are random realizations of a zero-mean Gaussian r.v. with variance σ^2. Therefore, $|A_{ij}|$ follows a half-normal distribution with mean $\sigma\sqrt{\frac{2}{\pi}}$ leading to the approximation on the l_1-norm of vector $\mathbf{A}_n$. In deriving (14), we have assumed that all X_i's, $i \neq n$, are known and the measurements are considered noise-free.

Upon closer examination of (14), it can be seen that the nth entry of the sparse signal $\mathbf{X}$ is considered relevant, hence, a nonzero-value, if its magnitude is greater than $c_1\tau_0$, otherwise it is forced to zero since its contribution to the signal formation is negligible. Furthermore, (14) tells us that the regularization parameter controls whether the nth entry is considered relevant or not based on an estimate of its magnitude value. Clearly, if a good estimate of the entry is found and with a suitable choice of the regularization parameter, the thresholding operation will identify correctly the nonzero values of $\mathbf{X}$. Note that $\tau_0 < M\sigma\sqrt{\frac{2}{\pi}} \min\{|X_i|, i = 1, 2, \ldots, N\}$ in order to identify all the nonzero values of $\mathbf{X}$.

Note also that the robustness inherent in the WM operation and the sparsity induced by the l_0 term are jointly combined into a single estimation-basis-selection task. More precisely, the selection of the few significant entries of the sparse vector $\mathbf{X}$ and the estimation of their values are jointly achieved by the WM operation followed by the hard-thresholding operator. The former operation adds the desired robustness on the estimation of the nth entry whereas the latter detects if $\hat{X}_n$ is one of the relevant component of $\mathbf{X}$. Furthermore, the regularization parameter acts like a tunable parameter that control which component is zeroed according to its magnitude value.

A. Iterative WM Regression Estimate

In deriving (12)–(13) it was assumed that X_j for $j = 1, 2, \ldots, N, j \neq n$ are known, or have been previously estimated somehow. Thus, in (12) the term $\sum_{j=1,j\neq n}^{N} A_{ij}X_j$ can be subtracted from each observation sample removing partially the contribution of the other nonzero-valued entries of the sparse vector $\mathbf{X}$ from the observation vector. This results in a new data set recentered around of X_n from which an estimation of location is carried out by the WM operation.

This suggests a very intuitive way of solving the l_0-regularized LAD regression problem given in (10) as follows. First, hold constant all but one of the entries of $\mathbf{X}$, find the entry that is allowed to vary by minimizing the weighted LAD, then alternate the role of the varying entry and repeat this process until a performance criterion is achieved. Thus, each entry of the sparse vector is iteratively estimated based on previous estimated value of the other entries. Interestingly, this approach to solve LAD has been historically referred to as the Edgeworth's algorithm and has been recently studied and refined in [52] with further applications to normalization of CDNA microarray data in [53], [54]. Most recently, this approach has been applied to solve l_1-regularized LAD regression problem under the framework of coordinate-wise descent [33], [35].

Before introducing the iterative algorithm for sparse signal reconstruction, the following remark will be used in the formulation of the reconstruction algorithm.

TABLE I
ITERATIVE WM REGRESSION ALGORITHM

Input	Observation vector $\mathbf{Y}$
	Measurement matrix $\mathbf{A}$
	Number of iterations K_0 and target residual energy ϵ
Initialization	
	Initial hard-thresholding value: $T_h = T_{h_i} = \|\mathbf{A}^T\mathbf{Y}\|_\infty$
	Iteration counter: $k = 1$
	Estimation at $k = 1$: $\hat{\mathbf{X}}^{(1)} = \mathbf{0}_N$
Iteration	
Step A	For the n-th entry of $\hat{\mathbf{X}}$, $n = 1, 2, \ldots, N$, compute
	$$\tilde{X} = \text{MEDIAN}\left(|A_{in}| \diamond \left. \frac{Y_i - \sum_{j=1, j\neq n}^{N} A_{ij}\hat{X}_j^{(k)}}{A_{in}} \right|_{i=1}^{M} \right)$$
	$$\hat{X}_n^{(k)} = \begin{cases} \tilde{X} & \text{if} \quad \|\mathbf{r}_n\|_{l_1} - \|\mathbf{r}_n - \tilde{X}\mathbf{A}_n\|_{l_1} > T_h \\ 0 & \text{Otherwise} \end{cases}$$
Step B	Update the hard-thresholding parameter and the estimation of $\mathbf{X}$
	$T_h = T_{h_i}\beta^k$
	$\hat{\mathbf{X}}^{(k+1)} = \hat{\mathbf{X}}^{(k)}$
Step C	Check stopping criterion
	If ($\frac{\|\mathbf{Y} - \mathbf{A}\hat{\mathbf{X}}^{(k)}\|_2^2}{\|\mathbf{Y}\|_2^2} > \epsilon$) and $k \leq K_0$ then set $k = k + 1$ and go to step A; otherwise, end
Output	Recovered sparse signal $\hat{\mathbf{X}}$

Remark: To implement the iterative reconstruction algorithm, we need to determine the regularization parameter from the measurement sample set $\{Y_i|_{i=1}^N\}$. This step has been shown to be critical in solving regularized linear regression problems [9], [17], [19], [35], [36] as this parameter governs the sparsity of the solution. In our approach, the regularization parameter also plays a vital role in the solution of the l_0-LAD problem since it becomes the threshold parameter[3] of the two-stage operator (13). Larger values for τ_0 leads to selecting only those entries that overcome the threshold value, leaving out nonzero entries with small values and, thus attaining a sparser solution. On the other hand, small values for τ_0 may lead to a wrong estimate of the support of $\mathbf{X}$ leading to spurious noise in the reconstruction signal.

The commonly used cross-validation [19] and generalized cross-validation methods [55] may be suitably adapted to determine the optimal regularization parameter as it was done in [56] for l_1-LS based CS signal reconstruction and in [33] and [35] for l_1-LAD regression. However, since this parameter strongly influences the accuracy of the reconstruction signal [29], [33], special attention has to be given to its estimation which may increase the computational complexity of the proposed algorithm. In our method, we follow a continuation approach, similar to that used in [57] to solve a l_1-regularized least square optimization problem. That is, we treat the regularization parameter (threshold of the hard thresholding operation) as a tuning parameter whose value changes as the iterative algorithm progresses. More precisely, start with a relative large value of T_h favoring sparsity during the first iterations, where T_h denotes the threshold parameter. Then, as the iterative algorithm progresses, the hard-thresholding parameter is slowly reduced. By doing so, it is expected that only those entries in the sparse vector $\mathbf{X}$ that have the most significant values (higher magnitude values) are identified in the first iterations. Subsequently, as the hard-thresholding parameter is reduced, those entries that exceed the new threshold value are identified next. The algorithm continues until the T_h reaches a final value or until a target residual energy is reached. The threshold value thus changes with each iteration as detailed in Table I that describes the iterative WM regression algorithm for sparse signal reconstruction. Appendix B shows a pseudocode of the proposed algorithm using the notation introduced in [58].

Several interesting observations should be pointed out about the proposed reconstruction algorithm. First, at the initial stage, all the entries of the estimated sparse vector are set to zero as in [59] and [60]. This initialization for the unknown variables is motivated by the fact that only a few components of the sparse vector $\mathbf{X}$ are expected to take on nonzero values. Second, at each iteration a weighted median based estimate is computed for each entry and the resultant estimated value is passed through a hard-thresholding operation. Note, in particular in step A, that the most recent estimated value for each entry is used for the estimation of subsequent entries in the same iteration step. More precisely, to compute the nth entry of $\hat{\mathbf{X}}$ at the kth iteration, the samples set $\{\hat{X}_1^{(k)}, \hat{X}_2^{(k)}, \ldots, \hat{X}_{n-1}^{(k)}, \hat{X}_{n+1}^{(k-1)}, \hat{X}_{n+2}^{(k-1)}, \ldots, \hat{X}_N^{(k-1)}\}$ that contains $(n-1)$ previous estimated values done at the kth iteration and $N - n$ estimated values obtained at the $(k-1)$th iteration is used in the computation of $\hat{X}_n^{(k)}$. It turns out that replacing the previous estimates by the most recent ones in a recursive fashion increases the convergence speed of the reconstruction algorithm.

Third, the updating operation of step B changes the hard-thresholding parameter for the next iteration. As mentioned above, T_h is dynamically adjusted as the iterative algorithm progresses. Hence, starting at the initial value T_{h_i}, the hard-thresholding parameter decays slowly as k increases. We set the hard-thresholding parameter to follow an

[3]Hereafter, the regularization parameter, τ_0 and the threshold parameter, T_h, of the hard thresholding operation are treated indistinctly.
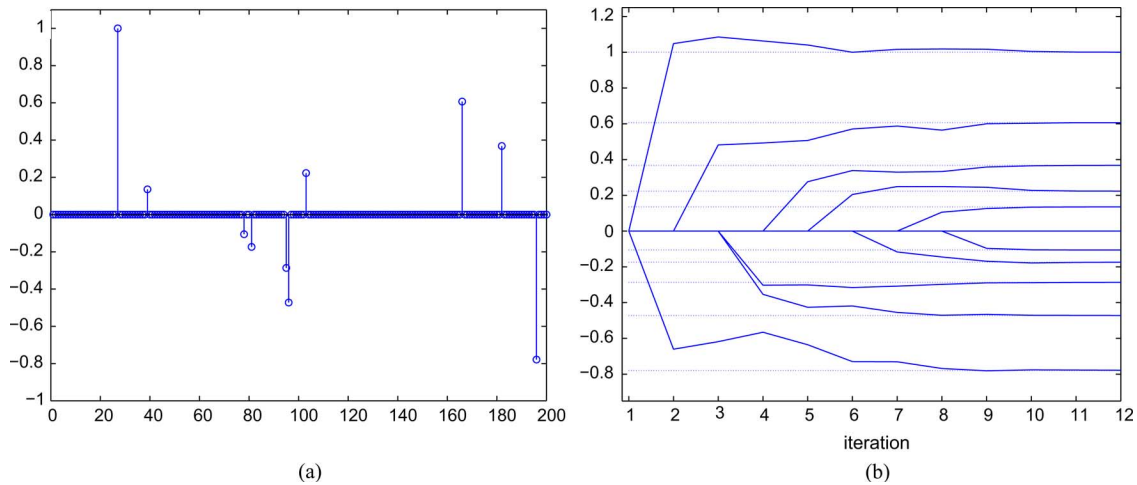
Fig. 2.   (a) Sparse test signal. (b) Nonzero entries of sparse vector as the iterative algorithm progresses. Dotted lines: true values. Solid lines: estimated values.

exponential decay function, i.e., $T_h = \tau = T_{h_i}\beta^k$ where $k = 1, 2, 3, \ldots, K_0$ and $0 < \beta < 1$ is a tuning parameter that controls the decreasing speed of T_h. This particular setting allows us to decrease the hard-thesholding parameter rapidly for the first few iterations, then slowly as k increases and, ultimately, approaching to zero as $k \rightarrow \infty$. As will be seen shortly, this decreasing behavior is helpful in detecting the few significant nonzero values of the sparse vector $\mathbf{X}$. Furthermore, the setting of the initial value for the threshold parameter, $T_{h_i} = \|\mathbf{A}^T\mathbf{X}\|_\infty$, ensures that T_h covers the all dynamic range of the sparse signal providing that enough iterations are allowed.

Finally, much like the stopping rule suggested in [3], [5]–[8], and [59], our stopping criterion to end the iterations is determined by the normalized energy of the residual error and a maximum number of algorithm iteration, K_0. Selecting K_0 and ϵ is a tradeoff among the desired target accuracy, the speed of the algorithm and the signal-to-noise ratio (SNR). It is worth mentioning that any other stopping criterion that adds robustness to impulsive noise can be readily adapted. For instance, the use of Geometric signal to noise ratio (G-SNR) [61] of the residual signal as stopping criterion allow us to use the proposed algorithm with noise following a heavy tail distribution.

Fig. 2 shows an illustrative example depicting how the nonzero values of the sparse vector are detected and estimated iteratively by the proposed algorithm. For this example, the tuning parameter β is set to 0.75. Note that it takes less than 10 iterations for the proposed algorithm to detect and estimate all the nonzero values of the sparse signal. Note also that the nonzero values are detecting in order of descending magnitude values, although this does not necessarily always occur.

It should be pointed out that an entry which is considered significant in a previous iteration but shown to be wrong at a later iteration can be removed from the estimated sparse vector. Furthermore, at each iteration several entries may be added to or removed from the estimated sparse vector as is shown in Fig. 2 where at iterations one and three two nonzero values are simultaneously detected for the first time.

The exponential decay behavior of the hard-thresholding parameter allows us to quickly detect, during the initial iterations,
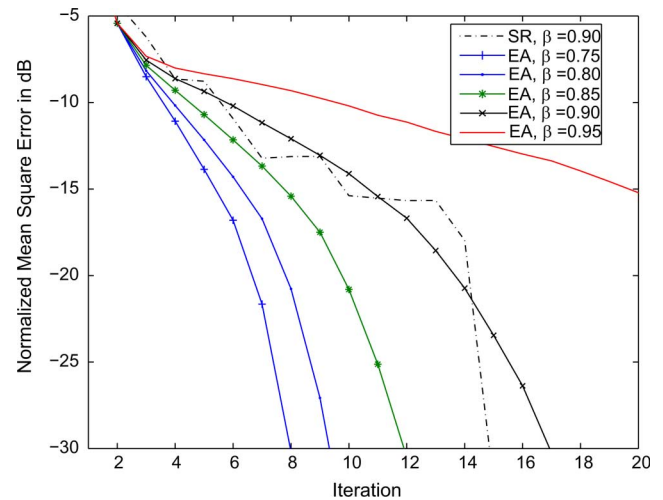

Fig. 3.   Normalized mean square error as a function of the iteration. Dash-dotted line represents a single realization (SR) of A while solid lines the ensemble average (EA) of 1000 realizations of A for each value of the tuning parameter β.

those entries of $\mathbf{X}$ that have large magnitude values. Furthermore, their corresponding WM estimated values are refined at each iteration. Therefore, with just a few iterations a good approximation to the sparse signal can be achieved by the iterative algorithm. As the iteration counter increases, the hard-thresholding parameter decreases slower allowing us to detect those entries in $\mathbf{X}$ with small magnitude values since the strongest entries have been partially removed from the observation vector.

Fig. 3 depicts the normalized mean square error (NMSE) in dB as the iterative algorithm proceeds for the reconstruction of the noiseless sparse signal shown in Fig. 2(a). As can be seen in Fig. 3, instead of decaying exponentially with a fixed rate, the NMSE of the iterative WM regression algorithm decays in a piecewise lineal fashion way. More precisely, a single realization of the NMSE follows an non-uniform staircase-shaped curve where abrupt jumps occur at each iteration where a new nonzero entry is detected. The magnitude of the step depends on the amplitude of the nonzero entry detected and the re-estimated values of the entries previously found. Note also that

the convergence speed of the proposed algorithm depends on the selection of the tuning parameter. At first glance, it seems that as the decaying speed of the hard-thresholding parameter becomes faster (small values for β) the proposed reconstruction algorithm converges much faster. However, small values for the threshold parameter may drive the algorithm to wrong estimates of the entries of the sparse signal, leading to spurious components (wrong basis selection) and wrong estimated values for the nonzero entries in the recovered signal. This, in turns, leads to an error floor in the NMSE as the iteration count becomes larger.

Selection of β and Number of Iterations: In the scenario where the sparsity level and the noise variance are unknown, the selection of β and the number of iterations which, in turn, sets the final value for the hard-thresholding parameter is not an easy task. Suitable values for these two parameters depend on the noise variance, the desired accuracy and the problem setup (N, M and K). Therefore, one has to resort to cross validation (trial and error) in order to find the best values of β and K_0 for a desirable performance for the problem at hand. It has been found that selecting a β in the interval $[0.75, 0.95]$, in general, yields good performance while the selection of the number of iterations is a tradeoff between algorithm speed and desired reconstruction accuracy. However, one has to be aware that increasing the number of iteration doesn't necessarily leads to improvement in NMSE since it is possible that as β^{K_0} becomes too small, spurious components start to appear in the reconstructed signal due to additive noise.

If an estimated noise level is available, however, the target residual energy, ϵ, can be set to $\frac{\tilde{\sigma}_\xi^2 M}{\|\mathbf{Y}\|_2^2}$, where $\tilde{\sigma}_\xi^2$ is the estimated value for the noise variance. With this setting, it turns out that the stopping criterion reduces to comparing the variance of the residual signal to the noise variance. Furthermore, the selection of the decaying speed (β) for the regularization parameter becomes less critical since any value in the interval $[0.75, 0.95]$ yields almost the same performance.

B. Comparison Analysis

In order to place the proposed algorithm in the context of previous related work, it is worth to compare the proposed algorithm to other iterative algorithms used in the literature for CS signal reconstruction. Specifically, we compare the proposed algorithm to the Matching Pursuit (MP) algorithm [3] and the Linearized Bregman Iterative (LBI) algorithm [59], [60] due to their similarities to our approach.

In the MP algorithm, the column-vector in $\mathbf{A}$ that best matches the residual signal in the Euclidean distance sense is selected as a relevant component. In contrast, our approach can be thought of as an iterative selection of the column-vectors in $\mathbf{A}$ that leads to the largest reductions in the l_1-norm of the residual signal. Furthermore, MP updating rules aims to minimize the l_2-norm of the residual signal, while our approach applies a hard-thresholding operation on the WM estimated value that minimizes the weighted l_1-norm of the residual signal. Moreover, since MP relies on the minimization of l_2-norm for basis selection and parameter estimation, its performance deteriorates if the observation vector is contaminated

with noise with heavier-than-Gaussian tailed distribution. In contrast, the proposed approach uses a robust estimator for the estimation of each entry of the sparse vector and an adaptive hard threshold operator for basis selection.

Comparing WM regression to the LBI algorithm, we see some resemblance. First, both algorithms rely on threshold based operators for basis selection. While the LBI algorithm uses a soft-thresholding operator for model selection, the WM regression algorithm uses a hard-thresholding operation with an adaptable threshold value. Secondly, the updating rule of the LBI algorithm involves a weighted mean operation on the residual samples while our approach uses a weighted median operation on the residual samples as a estimation of the corresponding entry. Interestingly, it is well-known that the WM operator is considered as the analogous operator to the weighted mean operator [42], [44]. This analogy also emerges in the updating rule of both algorithm. Since weighted mean operators offer poor robustness to outliers in the data sample [42], the updated sparse vector in the LBI algorithm is severely affected by the presence of impulsive noise in the measurements.

C. Computation Complexity Per Iteration

The proposed reconstruction algorithm is simple to implement and requires minimum memory resources. It involves just shifting and scaling operations, WM operators and comparisons. The per-iteration complexity is as follows. The data shifting and scaling operations can be efficiently implemented at a complexity cost of $\mathcal{O}(M)$ while the WM operation boils down to sorting and thus can be efficiently implemented by quicksort, mergesort or heapsort algorithms [46], [62] at a complexity cost of $\mathcal{O}(M \log(M))$. Since the proposed algorithm performs these operations for each entry in $\mathbf{X}$, the total per-iteration complexity reduces to $\mathcal{O}(N(M + M \log(M)))$. At first glance, it seems that the complexity of the proposed approach may be considered high compared to the per-iteration complexity of other CS iterative algorithms. However, this complexity can be notably reduced if an efficient implementation of the WM operator that avoids the sorting operation is used. This can be achieved by extending the concepts used in the QuickSelect algorithm [46], [47], [62] for the median operator to the weighted median operator leading to a complexity of order $\mathcal{O}(M)$ for the WM computation [45]. Appendix A shows the pseudocode of such extension. The overall computation complexity per iteration of the proposed algorithm reduces thus to $\mathcal{O}(NM)$ which is in the same order than MP and LBI algorithms. As will be shown in the simulations, the number of iterations required by the proposed algorithm is significantly fewer compared to those required by other reconstruction algorithms.

Finally, apart from storage the measurement vector $\mathbf{Y}$ and the projection matrix $\mathbf{A}$, the memory requirement is in the order of the dimension of the sparse signal to be recovered.

Before ending this section it should be pointed out that a formal convergence proof of the proposed algorithm is unavailable at the moment and remains as an open research problem. The nonlinear nature of the estimation stage makes a convergence analysis intractable. Furthermore, classical convergence analysis for coordinate descent algorithms relies

TABLE II
NUMBER OF ITERATIONS NEEDED TO REACH A NORMALIZED RESIDUAL ENERGY OF 10^{-10}. RESULTANT NORMALIZED MSE AND COMPUTATION TIME

N	M	K	Number of Iterations			Normalized MSE in dB			Computation time in secs.		
			WMR	MP	LBI	WMR	MP	LBI	WMR	MP	LBI
300	75	6	9.2	27.0	722.7	-106	-100	-96	0.0800	0.0022	0.0490
600	150	12	9.7	62.4	302.3	-104	-99	-96	0.2038	0.0138	0.0782
900	225	18	10.7	91.1	782.7	-105	-99	-97	0.3647	0.0457	0.8822
1200	300	24	10.6	123.1	984.2	-103	-99	-97	0.5518	0.2367	2.4936
2500	625	50	11.1	266.3	1007.4	-104	-99	-97	1.8788	1.9400	10.4298
5000	1250	100	12.2	528.5	4850.4	-103	-98	-98	5.8918	14.8658	197.3714
300	120	15	14.1	122.5	461.1	-102	-68	-94	0.1348	0.0115	0.0433
600	240	30	15.3	183.4	592.0	-103	-98	-95	0.3528	0.0529	0.4381
900	360	45	15.2	284.6	607.0	-102	-98	-95	0.6283	0.4848	1.3790
1200	480	60	15.4	396.9	921.3	-102	-98	-95	0.9369	1.0908	3.5379
2500	1000	125	16.2	816.0	2342.7	-103	-98	-96	3.5945	9.6535	38.8690
5000	2000	250	16.6	1653.9	3475.3	-101	-97	-96	11.1431	73.0798	223.4094

on the assumption that the cost function is continually differentiable [30], [31], in such case the minimum reached by a coordinate descendent algorithm is guaranteed to be the global minimum of the multidimensional cost function [30, pp. 277]. In our algorithm, however, the cost function for l_0-LAD is a nonconvex and a nondifferentiable function making useless the classical convergence analysis approach. Nevertheless, the coordinate descendent framework applied to l_1-LAD minimization problem has shown to be a reliable approach [32], [33] even though the cost function is nondifferentiable.

V. SIMULATIONS

In this section, the performance of the WM regression reconstruction algorithm is tested in several problems commonly considered in the CS literature. First, the proposed algorithm is tested in the recovery of a sparse signal from a reduced set of noise-free projections. Next, we text the robustness of the proposed approach when the measurements are contaminated with noise obeying statistical models with different distribution tails. Finally, the proposed algorithm is used in solving a LAD regression problem with a sparse parameter vector where the sparse signal as well as the contamination are both modeled by Laplacian distributions [33], [35]. In all the simulations, unless otherwise stated, the N-sample point sparse signal, $\mathbf{X}$, is generated by randomly locating the nonzero entries and then by assigning to the nonzero entries of $\mathbf{X}$ random values obeying a specific distribution. We use a zero-mean, unit-variance Gaussian distribution, a uniform distribution in the interval $(-1, 1)$ and a Laplacian distribution for the amplitude of the nonzero values of $\mathbf{X}$. Furthermore, the projection matrix $\mathbf{A}$ is created by first generating an $M \times N$ matrix with i.i.d. draws of a Normal distribution $\mathcal{N}(0, 1)$ and then by normalizing $\mathbf{A}$ such that each row has unit l_2-norm as in [9], [11], [16]. For the proposed algorithm, the initial threshold value is set to $\|\mathbf{A}^T\mathbf{Y}\|_\infty$ unless otherwise stated.

As a performance measure, we use the normalized mean-square error (MSE) (in dB) of the reconstructed signal, defined as Normalized MSE (dB) $= 10\log_{10}\left\{\frac{1}{T}\sum_{i=1}^{T}\frac{\|\hat{\mathbf{X}}^{[i]}-\mathbf{X}^{[i]}\|_2^2}{\|\mathbf{X}^{[i]}\|_2^2}\right\}$, where T is the number

of random trials. $\hat{\mathbf{X}}^{[i]}$ and $\mathbf{X}^{[i]}$ are, respectively, the recovered signal and the targeted signal for the ith realization. For each random trial, a new projection matrix, a new sparse signal, and a random realization of the noise (if applied) are generated.

A. Number of Iterations Needed to Achieve a Performance Residual Energy

In the first simulation, we are interested in finding the number of iterations needed for the proposed algorithms to achieve a given residual energy and compared it to those attained by the MP and LBI algorithms. To this end, two sets of sparse signals are generated as in [59]. In the first set, the number of nonzero entries in the sparse signal is set to $0.02N$, whereas in the second set, it is equal to $0.05N$ where N is the dimension on the sparse signal. Furthermore, the number of random projections, M, is set to $0.25N$ and $0.40N$ for the first and second set, respectively. For both sets of testing signals, the nonzero values of the sparse signal follow a $(-1, 1)$-uniform distribution. As a stopping rule for all the iterative algorithms, we use the normalized energy of the residual signal to end the iterations. Thus, each iterative algorithm finishes as soon as $\frac{\|\mathbf{Y}-\mathbf{A}\hat{\mathbf{X}}^{(k)}\|_2^2}{\|\mathbf{Y}\|_2^2} < 10^{-10}$ is reached. For this simulation, β is set to 0.75. Table II shows the results yielded by the various iterative algorithms where each entry reported is the ensemble average over 20 trials. As can be noticed in Table II, the proposed algorithm needs just a few iterations to achieve the performance target. Indeed, it took nearly 7 (15) times fewer iterations than the MP algorithm and about 68 (42) times fewer iterations than the LBI algorithm for the first(second) set of sparse signals. Note that the number of iterations required by the WMR algorithm remains practically constant for the reconstruction of all the sparse signals in the same set. For this simulation, the parameters of the Linearized Bregman algorithm are set to one. Note also in Table II that WM regression algorithm achieves the lowest normalized MSE. Furthermore, the computation times achieved by the proposed algorithm are smaller than those yielded by the MP and LBI for relatively large-scale reconstruction problems ($N > 1000$) while for a small-scale problem it yields competitive results compare to the LBI algorithm being the MP algorithm the fastest one. The running-time were obtained on a Core 2 Duo CPU, @ 2 GHz and 2 GB of RAM running the XP OS.
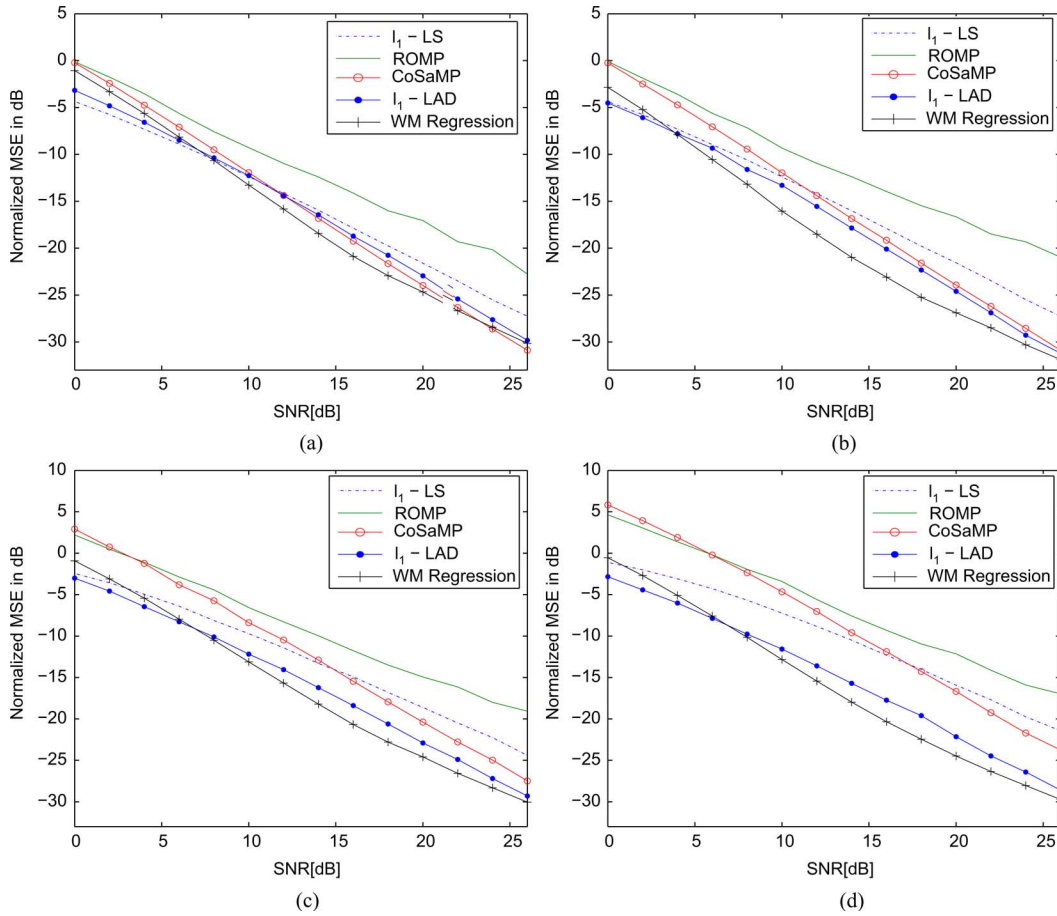
Fig. 4. Performance of the reconstruction algorithms under additive noise obeying: (a) Gaussian distribution; (b) Laplacian distribution; (c) ϵ-contaminated normal with $\epsilon = 1\%$; and (d) ϵ-contaminated normal with $\epsilon = 3\%$. $N = 512$, $M = 256$, $K = 25$.

B. Robustness to Impulsive Noise

In order to test the robustness of the proposed algorithm to noise, a 25-sparse signal of length 512 is generated with the nonzero entries drawn from a zero-mean, unit-variance Gaussian distribution. The projected signal with 250 samples is then contaminated with noise obeying statistical models with different distribution tails. We used Gaussian, Laplacian, ϵ-contaminated normal and Cauchy distributions for the noise model. For the ϵ-contaminated normal, $f_\xi(\xi) = (1 - \epsilon)\mathcal{N}(0, \sigma_1^2) + \epsilon\mathcal{N}(0, \sigma_2^2)$, σ_1^2 is set according to the desired SNR whereas $\sigma_2^2 = 100\sigma_1^2$. For all these simulations, we define the signal-to-noise ratio as SNR $= \frac{(\mathbf{AX})^T \mathbf{AX}}{M\sigma^2}$.

We compare the performance of the proposed approach in the noisy case to those yielded by two greedy based iterative algorithms [compressive sensing MP (CoSaMP) [8] and regularized orthogonal MP (ROMP) [6]] and two reconstruction algorithms belonging to the class of convex-relaxation based algorithms. More specifically, we recover the signal by solving the l_1-regularized optimization problems

$$\arg\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{AX}\|_{l_p} + \tau_p\|\mathbf{X}\|_{l_1} \quad \text{for} \quad p \in \{1, 2\}$$

where the regularization parameter τ_p is found as the one that yields the smallest normalized MSE for each experimental setup (N, M, K, SNR, and noise statistics).

The interior-point based method proposed in [9] is used to solve the l_1-regularized least square problem that results if $p = 2$. While for $p = 1$ we proceed as follows. First, an approximate solution, denoted as $\hat{\mathbf{X}}$, is found by solving the l_1-regularized LAD regression problem $\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{AX}\|_{l_1} + \tau_1\|\mathbf{X}\|_{l_1}$. Once an approximate solution is obtained, we perform debaising and denoising steps to reduce the attenuation of signal magnitude due to the presence of the regularization term [33] and to eliminate the spurious components in the approximate solution. To this end, the K entries of $\hat{\mathbf{X}}$ with largest magnitude are re-estimated by solving $\min_{\mathbf{X}_K} \|\mathbf{Y} - \mathbf{A}_K\mathbf{X}_K\|_{l_1} + \tau_1\|\mathbf{X}_K\|_{l_1}$, whereas the $N - K$ remaining entries of $\hat{\mathbf{X}}$ are set to zero. In this later optimization problem, $\mathbf{X}_K$ denotes a K-dimensional vector and $\mathbf{A}_K$ is an $M \times K$ matrix that contains the K column-vectors of $\mathbf{A}$ related to the location of the K largest values of $\hat{\mathbf{X}}$. Notice that, in this approach, we assume that the sparsity level is known in advance and that the optimal regularization parameter has been found by intensive simulations for each experimental setup. We use the Matlab's fminunc function to solve these optimization problems. For short notation, the results obtained with these convex-relaxation based approaches are denoted as l_1-LS and l_1-LAD for $p = 2$ and $p = 1$, respectively.

Fig. 4 depicts the curves of normalized MSE versus SNR yielded by the various reconstruction algorithms for different noise statistics. Each point in the curves is obtained by aver-
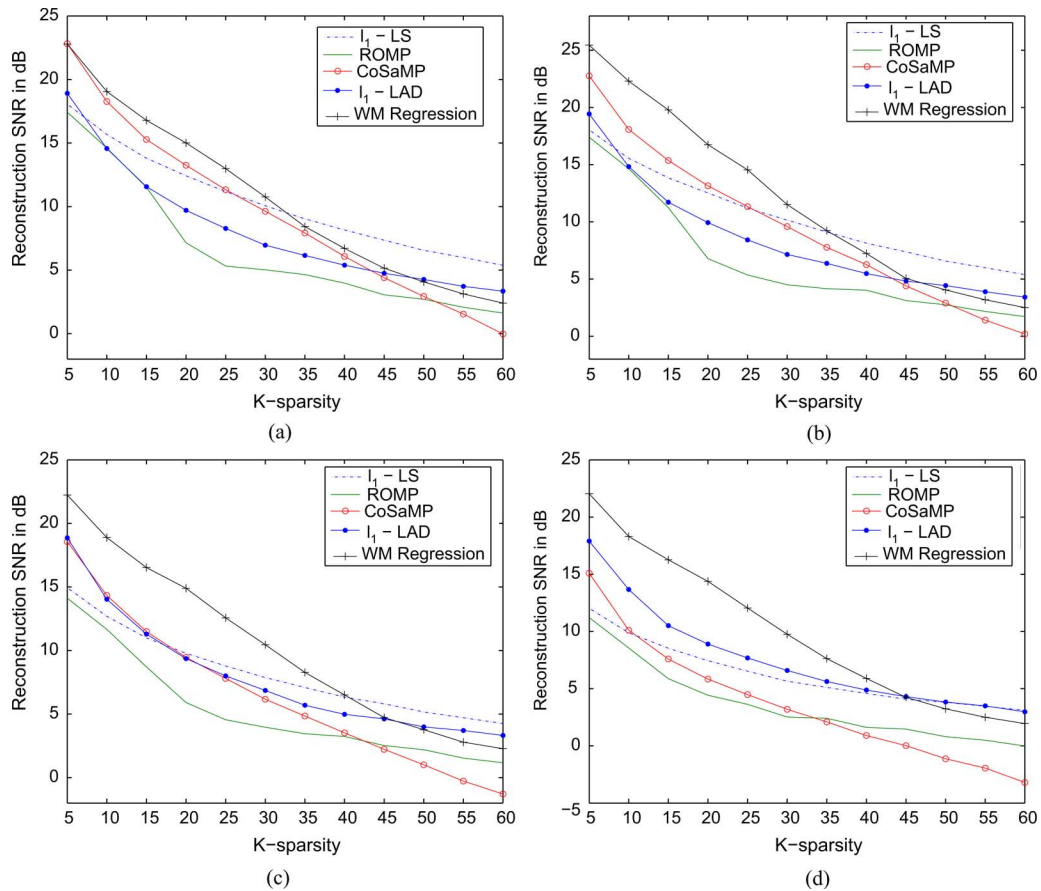
Fig. 5. Reconstruction SNR for the various algorithms under additive noise: (a) Gaussian contamination; (b) Laplacian contamination; (c) ϵ-contaminated normal with $\epsilon = 1\%$; and (d) ϵ-contaminated normal with $\epsilon = 3\%$. $N = 1024, M = 200, \mathrm{SNR} = 12\,\mathrm{dB}$.

aging over 1000 realizations of the linear model. For the proposed algorithm, the hard thresholding parameter β was set to 0.95 and the number of iterations is fixed to 40. For the l_1-LS algorithm the relative tolerance is fixed to 0.01 whereas for the greedy based algorithms ROMP and CoSaMP, the sparsity level is given as an input parameter [6], [8].

As can be seen in Fig. 4(a), when the underlying contamination is Gaussian noise, the proposed algorithm outperforms CoSaMP almost everywhere but at high SNR $(> 24\,\mathrm{dB})$ and, it yields better performance than that yielded by the ROMP. Furthermore, comparing the performance of the proposed approach to those yielded by the convex-relaxation algorithms described above, it can be seen that WM regression yields better results for SNR greater than 8 dB. For low SNR, however, the algorithms based on convex-relaxation outperforms our approach at cost of having previously optimized the regularization parameter for each SNR and noise statistics. Furthermore, for the l_1-LAD the sparsity level, $K = 25$, is also assumed to be known in advance. Note that this information is more than what we use in our approach.

It can be noticed in Fig. 4(d) that for a contamination fraction of 3% and a SNR of 12 dB, the proposed reconstruction algorithm achieves improvements in the Normalized MSE of about 1.8 dB over l_1-LAD, 6.6 dB over l_1-LS, 8.4 dB over CoSaMP and more than 9.5 dB with respect to ROMP. This improvement is due, in part, to the inherent robustness of the WM operator

in the estimation of the nonzero entries of $\mathbf{X}$ and, in part, to the robust detection of the their locations achieved by the adaptive threshold operator. Furthermore, as expected using LAD as the minimization criterion for data-fitting term leads to a much better performance under heavy-tail noise compared to those found by least square based estimator. Moreover, having l_0-term as sparsity-inducing term forces the desired sparsity in the solution more than that found by convex-relaxation based algorithms, like l_1-LAD and l_1-LS.

To further illustrate the performance of the proposed reconstruction algorithm, Fig. 5 depicts the reconstruction SNR in dB as the sparsity level changes for a 1024-dimensional sparse signal using 200 measurements contaminated with different noise distributions at a $\mathrm{SNR} = 12\,\mathrm{dB}$, where the reconstruction SNR in dB is just the negative of the normalized MSE in dB. Thus, the larger the reconstruction SNR is, the better the algorithm performs. As can be seen in this figure, the WM regression reconstruction algorithm yields a reconstruction error of the same order of the noise level for a sparsity of around 30 while for the other reconstruction algorithms the sparsity level is significantly smaller. More precisely, upon closer observation of Fig. 5(b), for instance, it can be seen that in order to have a reconstruction SNR greater than the Laplacian noise level the sparsity for ROMP, l_1-LAD, l_1-LS and CoSaMP, must be smaller than 14, 15, 22, and 23, respectively, while for the proposed algorithm it is around 30.
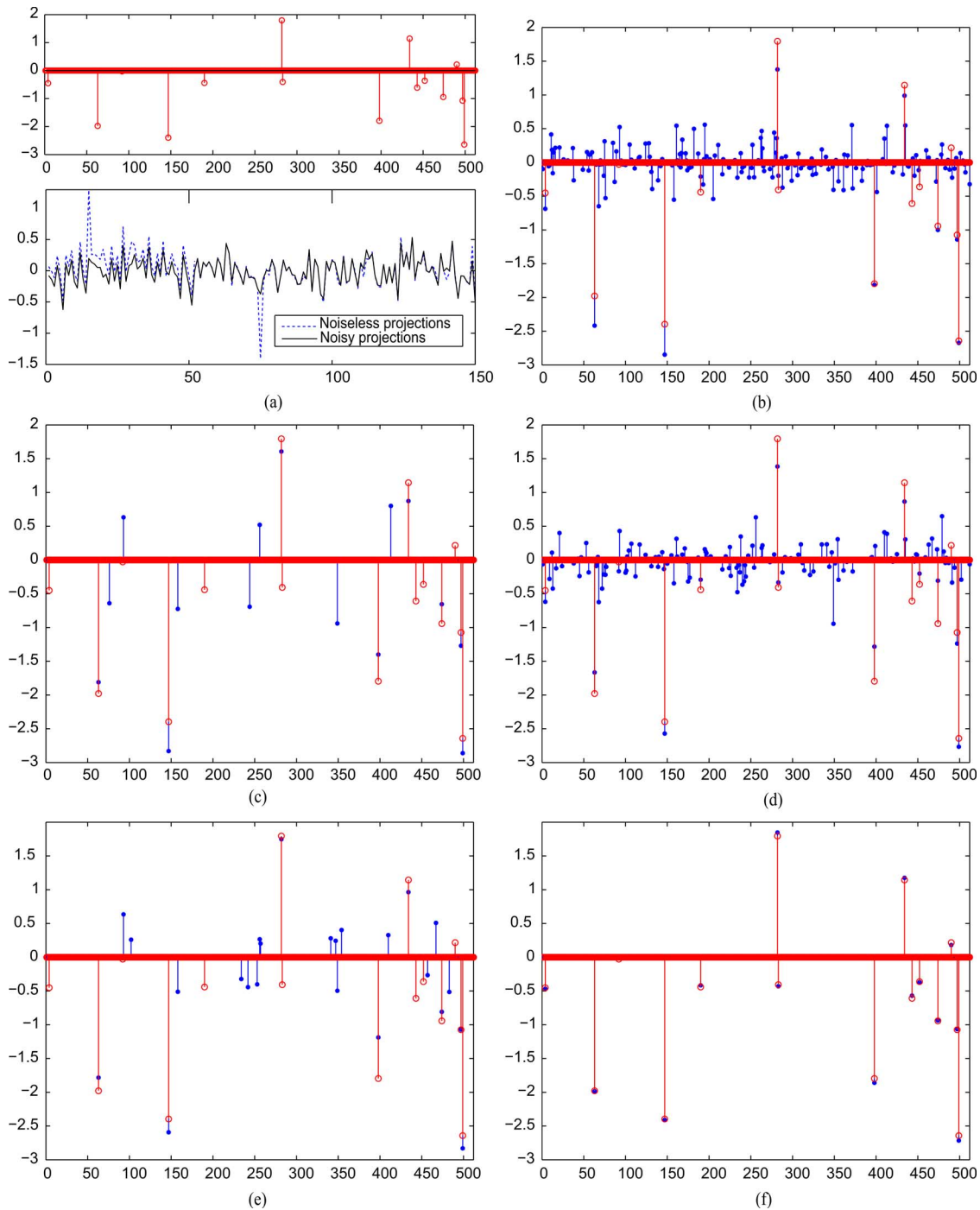
Fig. 6. Reconstruction of a 15-sparse, 512-long signal from a set of 150 contaminated measurements. (a) Top: Sparse signal. Bottom: noiseless and noisy projections. Recovered signals yield by: (b) MP; (c) CoSaMP; (d) l_1-LS; (e) LBI; and (f) WM regression algorithms. ○ denotes original signal, ● denotes reconstructed signal.

To illustrate the robustness of the proposed approach to the presence of noise with heavier distribution tails, consider the 15-sparse, 512-long signal shown in Fig. 6(a) (top) that has been projected and contaminated with impulsive noise obeying the standard (0,1)-Cauchy distribution. The noise is scaled by a factor of 10^{-2} and then added to the projected signal. Fig. 6(a) (bottom) shows both the noiseless and noisy projections. Note that the noise level is such that the noisy projections approximate closely the noiseless projections almost

everywhere but at the entries 15 and 75 where two outliers are present. Fig. 6(b)–(f) show the recovered signals achieved by the various reconstruction algorithms using 150 measurements. Note that the MP and l_1-LS reconstruction algorithms are very sensitive to the presence of outliers introducing several spurious components in the reconstructed signal. As can be seen in Fig. 6(f), our approach is able not only to identify correctly the most significant nonzero entries of the sparse signal but to output good estimated values for the nonzero entries as well.
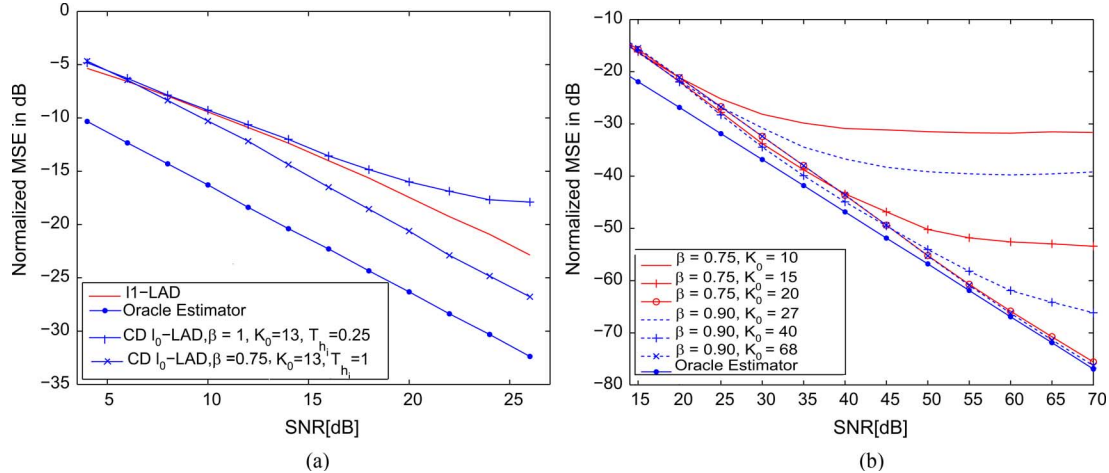
Fig. 7. Performance of the proposed approach for different values of SNR, number of iterations (K_0), threshold parameter (β) and initial threshold value, T_{h_i}. $N = 200$, $M = 100$, $K = 20$.

C. Solving a Linear Regression Problem With a Sparse Parameter Vector

As a final simulation, we test the performance of the proposed approach in solving a linear regression problem with a sparse parameter vector and compare it to the optimum solution yielded by an oracle estimator. We follow a similar simulation setup than the one used in [33], [35]. That is, in the linear model $\mathbf{Y} = \mathbf{A}\mathbf{X} + \sigma\boldsymbol{\xi}$, the first 20 entries of $\mathbf{X}$ are realizations of a (0,1)-Laplacian distributed r.v. while the other 180 entries are set to zero. The entries of the predictor matrix, A_{ij}, follow a standard normal distribution and $\boldsymbol{\xi}$ obeys the multivariate Laplacian distribution with zero mean and covariance $\mathbf{I}$. The scalar parameter σ is set according to the desired SNR. As in [33], we are interested in the underdetermined case, hence we use 100 observations to estimate the entries of the sparse vector.

We compare the performance of the proposed approach to the performance yielded by an oracle estimator that exploits the *a priori* information about the location of the nonzero values of the sparse vector and gives as estimated values of the nonzero entries of $\mathbf{X}$ the solution to the LAD problem $\min_{\mathbf{X}_\Omega} \|\mathbf{Y} - \mathbf{A}_\Omega\mathbf{X}_\Omega\|_{l_1}$ solved using a convex-optimization algorithm [63], where $\mathbf{X}_\Omega$ is a 20-dimensional vector and $\mathbf{A}_\Omega$ denotes the first 20 column-vector of $\mathbf{A}$. Furthermore, a convex-relaxation approach is also used to estimate the sparse parameter vector. More precisely, the sparse vector is found as the solution to the l_1-regularized LAD optimization problem $\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_{l_1} + \tau_1\|\mathbf{X}\|_{l_1}$ where again the regularization parameter, τ_1, is optimally found for each SNR. To solve this l_1-LAD regression problem, we use the following approach. As in [32] and [35], we reformulate the l_1-LAD problem as an unregularized LAD regression problem by defining the augmented observation vector $\mathbf{Y}' \in R^{M+N}$ where $Y_i' = Y_i$ for $i = 1, 2, \ldots, M$, $Y_i' = 0$ for $i = M + 1, \ldots N + M$ and the expanded predictor matrix $\mathbf{A}' = [\mathbf{A}^T \,\vdots\, \tau_1\mathbf{I}_N]^T$, where $\mathbf{I}_N$ denotes the $N \times N$ identity matrix. Thus, the sparse parameter vector is found by solving $\min_{\mathbf{X}} \|\mathbf{Y}' - \mathbf{A}'\mathbf{X}\|_{l_1}$ using the *Econometric Toolbox* developed in [63].

Fig. 7(a) depicts the performance achieved by the various estimators. Each point in the curves is the average over 2000 random realizations for the linear model. For short notation the proposed WM regression algorithm is denoted as CD l_0-LAD. For comparative purposes, we show the performance yielded by our proposed approach for two different sets of parameters. In the first one, the threshold parameter is held fixed to T_{h_i} during the algorithm iterations while, in the second parameter set, it changes according to β^k. As expected, finding the sparse vector by solving an l_0-regularized LAD regression problem leads to a much better performance than the one found by solving a l_1-regularized LAD. Note that the performance of the proposed approach is the one closest to the oracle estimator, observing a performance loss of about 5 dB for low SNR and close to 0 dB for high SNR [see Fig. 7(b)]. Note also that changing the threshold parameter as the algorithm progresses leads to a performance gain compared to holding it fix. This performance gain tends to increase as the SNR becomes higher. Finally, in Fig. 7(b), we compare the performance of the proposed approach to the oracle estimator for different sets of parameters. For comparative purposes, the hard-thresholding parameter starts at $T_{h_i} = 1$ and ends approximately at the same value for $\beta \in \{0.75, 0.90\}$. Interestingly, for high SNR the performance of the proposed algorithm improves notably for a relative large number of iterations. In fact, its performance approaches that of the oracle estimator for SNR > 50 dB. Furthermore, if the number of algorithm iterations is low, there is an evident normalized MSE floor which tends to be higher for low values of β. Note also that the decaying speed of the hard-thresholding parameter affects the performance of the proposed approach for SNR greater than 25 dB, observing an improvement in performance as the threshold parameter decays slower. While, for lower SNR, the proposed algorithm performs practically the same for $\beta = 0.75$ and $\beta = 0.90$. Thus, for high SNR, having a low value for β ($\beta \to 0.75$) and just a few algorithm iterations lead to a fast signal estimation at expensive of a relatively high normalized MSE, while selecting a higher value for β, the algorithm needs more iterations but achieves much lower normalized MSE.

VI. CONCLUSION

In this paper, we present a robust reconstruction algorithm for CS signal recovery. The proposed approach is able to mitigate the influence of impulsive noise in the measurement vector leading to a sparse reconstructed signal without spurious components. For the noiseless case, the proposed approach yields a similar reconstructed accuracy compared to those attained by other CS reconstruction algorithms, while a notable performance gain is observed as the noise becomes more impulsive. Although we have focused our attention on CS signal reconstruction, the concepts introduced in this paper can be extended to other signal and communications applications that involves solving an inverse problem. Signal representation on overcomplete dictionaries [15], sparse channel estimation and equalization [64], [65] and sparse system identification [66] are just three examples of applications where the proposed algorithm can be used. Furthermore, if more robustness is desired a weighted Myriad operator [67], [68] can be used in place of the WM operator [69].

APPENDIX A
WEIGHTED MEDIAN COMPUTATION

Given a sample set $X = \{x_1, x_2, \ldots, x_M\}$ and a set of weights $W = \{w_1, w_2, \ldots, w_M\}$, $w_i \in \mathcal{R}$, the MEDIAN $(w_1 \diamond x_1, w_2 \diamond x_2, \ldots, w_M \diamond x_M)$ is the kth signed sample that satisfies [43], [44]

$$\sum_{\text{sgn}(w_i)x_i < \text{sgn}(w_k)x_k} |w_i| < w_0 \text{ and}$$

$$\sum_{\text{sgn}(w_i)x_i > \text{sgn}(w_k)x_k} |w_i| > w_0$$

where $w_0 = \frac{\sum_{i=1}^{M} |w_i|}{2}$. Notice that if we resort to a sorting operation, the computation of the WM operation can be attained at a running-time of $\mathcal{O}(M \log(M))$ [46]. However, since we just want to find the kth signed sample that satisfied the above condition sorting is not needed. To find the WM sample, we proceed as follows:

1) Define the signed sample set by passing the signs of the weights to the corresponding samples, i.e., $Z = \{z_i = \text{sgn}(w_i)x_i \text{ for } i = 1, 2, \ldots, M\}$.
2) Redefine the set of weights by taken the magnitude of each weight, i.e., $W = \{w_i = |w_i|, i = 1, 2, \ldots, M\}$.
3) Compute the threshold value, $w_0 = 0.5 \cdot \sum_{i=1}^{M} w_i$
4) Run the weighted-median function on $\{z_1, z_2, \ldots, z_M\}$.

Next, we present the pseudocode of the weighted-median function adapted from [46]. For the sake of simplicity we have assumed that the singed samples, z_i, are not repeated.

Weighted-median (z)

---

if $M > 2$

Compute the median of the input sample set: $z_k = \text{median}(z_1, z_2, \ldots, z_M)$

Compute $w_L = \sum_{z_i < z_k} w_i$ and $w_R = \sum_{z_i > z_k} w_i$

if $w_L < w_0$ and $w_R < w_0$

then return z_k

elseif $w_L > w_0$

then

$w_k = w_k + w_R$

$Z' \leftarrow \{z_i : z_i \leq z_k\}$

return **Weighted-median** (Z')

else

$w_k = w_k + w_L$

$Z' \leftarrow \{z_i : z_i \geq z_k\}$

return **Weighted-median** (Z')

elseif $M = 1$

then return z_1

else

if $w_1 \geq w_2$

then return z_1

else

return z_2

Notice that the computation of the sample median can be performed in $\mathcal{O}(M)$ time using a Quickselect algorithm like the one introduced in [47] leading to an overall computation time of $\mathcal{O}(M)$ [45].

APPENDIX B
PSEUDOCODE OF THE WM REGRESSION ALGORITHM

$\hat{\mathbf{X}} = \mathbf{0}_N$

$\hat{\mathbf{Y}} = \mathbf{0}_M$

$k = 0$

$T_h = T_{h_i}$

while $((k <= K_0) \text{ and } (\frac{\|\mathbf{Y} - \hat{\mathbf{Y}}\|_2^2}{\|\mathbf{Y}\|_2^2} > \epsilon))$

for n = 1 : N

$\mathbf{r} = \mathbf{Y} - \hat{\mathbf{Y}} + \mathbf{A}(:, n) * \hat{X}(n)$

Find the weighted median $\hat{X}$ of the sample set $\left\{ \left. \frac{\mathbf{r}(i)}{A(i,n)} \right|_{i=1}^{M} \right\}$

with weights $\left\{ \left. |A(i,n)| \right|_{i=1}^{M} \right\}$

Compute $skt = \|\mathbf{r}\|_{l_1} - \left\| \mathbf{r} - \mathbf{A}(:, n) * \hat{X} \right\|_{l_1}$

$temp = \hat{X}(n)$

if $(skt > T_h)$

then $\hat{X}(n) = \hat{X}$

else

then $\hat{X}(n) = 0$

$$\hat{\mathbf{Y}} = \hat{\mathbf{Y}} + (\hat{X}(n) - temp) * \mathbf{A}(:, n)$$

end

$$T_h = \beta * T_h$$

$$k++$$

end

where $\mathbf{0}_N$ denotes the N-dimensional all-zero vector and $\mathbf{A}(:, n)$ is the nth column-vector of $\mathbf{A}$.

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[2] E. J. Candes, "Compressive sampling," in *Proc. Int. Cong. Mathemat.*, Madrid, Spain, 2006, vol. 3, pp. 1433–1452.

[3] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, 1993.

[4] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[5] D. L. Donoho, Y. Tsaig, I. Drori, and J. L. Starck, Sparse Solution of Underdetermined Linear Equations by Stagewise Orthogonal Matching Pursuit, Dep. Statist. Stanford Univ., Stanford, CA, 2006 [Online]. Available: sparselab.stanford.edu/SparseLab_files/local_files/StOMP.pdf

[6] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Found. Computat. Math.*, vol. 9, no. 3, pp. 317–334, Jun. 2009.

[7] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.

[8] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, pp. 301–321, 2008.

[9] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale l_1-regularized least squares," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 606–617, Dec. 2007.

[10] E. Candés, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, pp. 489–509, 2006.

[11] E. Candès and J. Romberg, ℓ_1-Magic: A Collection of MATLAB Routines for Solving the Convex Optimization Programs Central to Compressive Sampling. 2006 [Online]. Available: www.l1magic.org/

[12] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Applications to compressive sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–598, 2007.

[13] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Aug. 2004.

[14] E. J. Candés, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[15] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[16] S. Ji, Y. Xeu, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.

[17] E. G. Larsson and Y. Selén, "Linear regression with a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 451–460, Feb. 2007.

[18] J.-J. Fuchs, "Fast implementation of a l1-l1 regularized sparse representations algorithm," in *Proc. 2009 IEEE Int. Conf. Acoust., Speech Signal Process.*, Wash., DC, 2009, pp. 3329–3332.

[19] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Statist. Soc.*, vol. 58, no. 1, pp. 267–288, 1996.

[20] P. J. Huber, *Robust statistics*, J. Wiley, Ed. New York: Wiley, 1981.

[21] W. Li and J. J. Swetits, "The linear ℓ_1 estimator and the huber M-estimator," *SIAM J. Optimiz.*, vol. 8, no. 2, pp. 457–475, May 1998.

[22] A. Giloni, J. S. Simonoff, and B. Segupta, "Robust weighted LAD regression," *Computat. Statist. Data Anal.*, vol. 50, pp. 3124–3140, Jul. 2005.

[23] T. E. Dielman, "Least absolute value regression: Recent contributions," *J. Statist. Computat. Simul.*, vol. 75, no. 4, pp. 263–286, Apr. 2005.

[24] M. Nikolova, "A variational approach to remove outliers and impulse noise," *J. Math. Imaging Vis.*, vol. 20, no. 1–2, pp. 99–120, 2004.

[25] H. Fu, M. K. Ng, M. Nikolova, and J. L. Barlow, "Efficient minimization methods of mixed l2-l1 and l1-l1 norms for image restoration," *SIAM J. Sci. Comput.*, vol. 27, no. 6, pp. 1881–1902, 2006.

[26] L. Granai and P. Vandergheynst, "Sparse approximation by linear programming using an L1 data-fidelity term," in *Proc. Workshop on Signal Process. Adapt. Sparse Structur. Represent.*, 2005.

[27] P. Rodríguez and B. Wohlberg, "An efficient algorithm for sparse representations with l^p data fidelity term," in *Proc. 4th IEEE Andean Tech. Conf. (ANDESCON)*, Cusco, Perú, Oct. 2008.

[28] H. Xiaoming and C. Jie, Hardness on Numerical Realization of Some Penalized Likelihood Estimators School of Indust. Syst. Eng., Georgia Inst. Technol., 2006 [Online]. Available: http://www2.isye.gatech.edu/statistics/papers/06-08.pdf

[29] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, Jun. 2010.

[30] A. Ruszczynski, *Nonlinear Optimization*. Princeton, NJ: Princeton Univ., 2006.

[31] Z. Q. Luo and P. Tseng, "On the convergence of the coordinate descent method for convex differentiable minimization," *J. Optimiz. Theory Appl.*, vol. 72, no. 1, pp. 7–35, Jan. 1992.

[32] J. Friedman, T. Hastie, H. Hofling, and R. Tibishirani, "Pathwise coordinate optimization," *Ann. Appl. Statist.*, vol. 1, no. 2, pp. 302–332, 2007.

[33] T. T. Wu and K. Lange, "Coordinate descent algorithm for lasso penalized regression," *Ann. Appl. Statist.*, vol. 2, no. 1, pp. 224–244, 2008.

[34] Y. Li and S. Osher, "Coordinate descent optimization for ℓ^1 minimization with application to compressed sensing; a greedy algorithm," *Inverse Probl. Imag. (IPI)*, vol. 3, no. 3, pp. 487–503, Aug. 2009.

[35] H. Wang, G. Li, and G. Jiang, "Robust regression shrinkage and consistent variable selection through the LAD-Lasso," *J. Business Econom. Statist.*, vol. 25, no. 3, pp. 347–355, Jul. 2007.

[36] L. Wang, M. D. Gordon, and J. Zhu, "Regularized least absolute deviation regression and an efficient algorithm for parameter tuning," in *Proc. 6th Int. Conf. Data Mining (ICDM'06)*, 2006.

[37] M. Seeger, S. Gerwinn1, and M. Bethge1, in *Machine Learning: ECML 2007, ser. Lecture Notes in Computer Science*, Berlin/Heidelberg, 2007, vol. 4701, pp. 298–309, Bayesian Inference for Sparse Generalized Linear Models.

[38] A. Kabán, "On bayesian classification with Laplace priors," *Pattern Recogn. Lett.*, vol. 28, no. 10, pp. 1271–1282, 2007.

[39] M. A. T. Figueiredo, "Adaptive sparseness for supervised learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1150–1159, Sep. 2003.

[40] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, 2008.

[41] D. R. K. Brownrigg, "The weighted median filter," *Commun. ACM*, vol. 27, no. 8, pp. 807–818, Aug. 1984.

[42] L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo, "Weighted median filters: A tutorial," *IEEE Trans. Circuits Syst.*, vol. 43, no. 3, pp. 157–192, Mar. 1996.

[43] G. R. Arce, *Nonlinear Signal Processing. A Statistical Approach*. New York: Wiley Intersci., 2005.

[44] G. R. Arce, "A general weighted median filter structure admitting negative weights," *IEEE Trans. Signal Process.*, vol. 46, no. 12, pp. 3195–3205, Dec. 1998.

[45] A. Rauh and G. R. Arce, "A fast weighted median algorithm based on quick select," in *Proc. 2010 IEEE Int. Conf. Image Process.*, Sep. 2010.

[46] T. Cormen, C. Leiserson, R. Rivest, and C. Stein, *Introduction to Algorithms, ser. MIT Electrical Engineering and Computer Science*. Cambridge, MA: MIT Press, 2001.

[47] W. Cunto and I. Munro, "Average case selection," *J. ACM*, vol. 36, no. 2, Apr. 1989.

[48] G. R. Arce and J. L. Paredes, "Recursive weighted median filters admitting negative weights and their optimization," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 768–779, Mar. 2000.

[49] M. McLoughlin and G. R. Arce, "Deterministic properties of the recursive separable median filter," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 1, pp. 98–106, Jan. 1987.

[50] K. Barner and G. R. Arce, "Permutation filters—a class of nonlinear filters based on set permutations," *IEEE Trans. Signal Process.*, vol. 42, no. 4, pp. 782–798, Apr. 1994.

[51] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions, ser. Wiley Series in Probability and Statistics.* New York: Wiley, 1986.

[52] Y. Li and G. R. Arce, "A maximum likelihood approach to least absolute deviation regression," *EURASIP J. Appl. Signal Process.*, no. 12, pp. 1762–1769, 2004.

[53] J. M. Ramírez, J. L. Paredes, and G. Arce, "Normalization of CDNA microarray data based on least absolute," in *Proc. 2006 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP'2006)*, Toulouse, France, May 2006, vol. II, pp. 1016–1019.

[54] J. M. Ramírez and J. L. Paredes, "A fast normalization method of CDNA microarray data based on LAD," in *Proc. IV Latin-Amer. Congress on Biomed. Eng. (CLAIB 2007)*, Margarita, Venezuela, Sep. 2007.

[55] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. Statist. Assoc.*, vol. 96, no. 456, pp. 1348–1360, Dec. 2001.

[56] P. Boufounos, M. F. Duarte, and R. G. Baraniuk, "Sparse signal reconstruction from noisy compressive measurements using cross validation," in *Proc. IEEE Workshop on Statist. Signal Process.*, Madison, WI, Aug. 2007, pp. 1–5.

[57] E. T. Hale, W. Yin, and Y. Zhang, "A fixed-point continuation method for ℓ_1-regularized minimization with applications to compressed sensing," Rice Univ., Houston, TX, CAAM Tech. Rep. TR07-07, 2007.

[58] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.

[59] S. Osher, Y. Mao, B. Dong, and W. Yin, "Fast linearized Bregman iteration for compressive sensing and sparse denoising," *Commun. Math. Sci.*, vol. 8, no. 1, pp. 93–111, 2010.

[60] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman iterative algorithm for ℓ_1-minimization with applications to compressive sensing," *SIAM J. Imaging Sci.*, vol. 1, no. 1, pp. 143–168, 2008.

[61] J. G. Gonzalez, J. L. Paredes, and G. R. Arce, "Zero-order statistics: A mathematical framework for the processing and characterization of very impulsive signals," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3839–3851, Oct. 2006.

[62] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C*, 2nd ed. New York: Cambridge Univ. Press, 1992.

[63] J. P. LeSage, Applied Econometrics Using MATLAB 1999 [Online]. Available: http://www.spatial-econometrics.com/

[64] J. L. Paredes, G. R. Arce, and Z. Wang, "Ultra-wideband compressed sensing: Channel estimation," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 3, pp. 383–395, Oct. 2007.

[65] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374–377, Mar. 2002.

[66] F. O'Regan and C. Heneghan, "A low power algorithm for sparse system identification using cross-correlation," *J. VLSI Signal Process. Syst.*, vol. 40, no. 3, pp. 311–333, Jul. 2005.

[67] J. G. Gonzalez and G. R. Arce, "Optimality of the myriad filter in practical impulsive-noise environments," *IEEE Trans. Signal Process.*, vol. 49, no. 2, pp. 438–441, Feb. 2001.

[68] S. Kalluri and G. R. Arce, "Adaptive weighted myriad filter algorithms for robust signal processing in alpha-stable noise environments," *IEEE Trans. Signal Process.*, vol. 46, no. 2, pp. 322–334, Feb. 1998.

[69] G. R. Arce, D. Otero, A. B. Ramirez, and J. L. Paredes, "Reconstruction of sparse signals from ℓ_1 dimensionality-reduced Cauchy random-projections," in *Proc. 2010 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Mar. 2010, pp. 4014–4017.

Jose L. Paredes (S'00–M'01–SM'10) received the Diploma in electrical engineering from the Universidad de Los Andes, Mérida, Venezuela, in 1995, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Delaware (UD), Newark, in 1999 and 2001, respectively.

From 1996 to 2001, he was a Research Assistant with the Department of Electrical and Computer Engineering, UD. Since 2001, he has been a full time Professor with the Electrical Engineering Department, Universidad de Los Andes. In 2009, he was a visiting professor with the Electrical and Computer Engineering Department, UD. He has consulted with industry in the areas of digital communication and signal processing. His research interests include robust signal and image processing, nonlinear filter theory, adaptive signal processing, and digital communications. He holds five U.S. patents.

Dr. Paredes is the recipient of several awards. In 1995, he received several medals and awards for being the first Electrical Engineering student to graduate from the Universidad de Los Andes with the highest honors. He also was awarded three scholarships from the Government of Venezuela (in 1988, 1992, and 1996). In 1997, he won the Best Paper award from the U.S. Army's ATIRP Federated Laboratory Consortium. He received the Outstanding Advisor Award granted by IEEE-INELECTRA as supervisor of the Best National Research Undergraduate Project in 2005. He is also the recipient of the 2006 ISA Transaction Best Paper Award.

Gonzalo R. Arce (F'00) received the Ph.D. degree from Purdue University, West Lafayette, IN.

He joined the University of Delaware, Newark, where he is the Charles Black Evans Distinguished Professor of Electrical and Computer Engineering. He served as Department Chairman from 1999 to 2009. He held the 2010 Fulbright—Nokia Distinguished Chair in Information and Communications Technologies at the Aalto University, Helsinki, Finland. He has held Visiting Professor appointments with the Tampere University of Technology and the Unisys Corporate Technology Center. His research interests include statistical signal processing, nonlinear signal processing, and computational imaging. He is the author or coauthor of the books *Nonlinear Signal Processing* (New York: Wiley, 2004), *Modern Digital Halftoning* (Boca Raton, FL: CRC, 2008), and *Computational Lithography* (New York: Wiley, 2010). He is a frequent consultant to industry in the areas of communications and imaging.

Dr. Arce is the recipient of several awards including the DuPont Young Investigator Award, the NSF Research Initiation Grant, and a fellowship from the Center for Advanced Studies, University of Delaware. He has supervised four Colburn Price Award-winning Ph.D. dissertations at the University of Delaware. He has served as an Associate Editor for several journals of the IEEE and the Optical Society of America.