ROBUST SPARSE SIGNAL RECOVERY BASED ON WEIGHTED MEDIAN OPERATOR

Juan Marcos Ramírez and José Luis Paredes

Department of Electrical Engineering, Universidad de Los Andes, Mérida 5101, Venezuela

ABSTRACT

In this paper, a coordinate descent algorithm for robust sparse signal representation in redundant dictionaries is proposed. Under the coordinate descent framework, each target coefficient is robustly estimated applying the weighted median to a scaled-and-shifted version of the input signal weighted by the magnitude of an atom associated to the underlying coefficient. Sparsity is induced by appending, in the weighted median operation, a zero-valued sample weighted by an adaptive parameter. This leads to a generalized thresholding function over each target coefficient minimizing, thus, both the bias on the nonzero-value estimates and the sensitivity to small levels of noise. Furthermore, a continuation approach is included in order to set a suitable value of the regularization parameter that leads to the best representation at a current noise level. Numerical simulations are presented, in the context of compressive sensing, to compare the performance of the proposed algorithm to those yielded by state-of-the-art methods.

Index Terms— Coordinate descent, sparse signal representation in overcomplete dictionaries, weighted median

1. INTRODUCTION

A recovery algorithm, in the context of sparse representations in overcomplete dictionaries, aims at estimating the sparsest coefficient vector $\mathbf{x} \in \mathbb{R}^N$ that best expands the signal-of-interest $\mathbf{z} \in \mathbb{R}^M$, with $\mathbf{z} = \mathbf{A}\mathbf{x}$ and $M \leq N$, from a set of noisy samples $\mathbf{y} = \mathbf{A}\mathbf{x}+\boldsymbol{\nu}$. In this context, $\mathbf{A} \in \mathbb{R}^{M\times N}$ represents an overcomplete (or redundant) dictionary, in other words, a collection of specially designed or get trained waveforms tailored to the application at hand that best expands a set of target signals. Furthermore, $\boldsymbol{\nu}$ is the noise vector whose characterization is assumed as additive samples that follow a common statistical distribution. Finally, an estimate of the desired noiseless signal can be built —from the recovered sparse coefficient vector $\hat{\mathbf{x}}$— as a linear combination of a few waveforms in the overcomplete dictionary, i.e. $\hat{\mathbf{z}} = \mathbf{A}\hat{\mathbf{x}}$.

The problem of finding a sparse representation in an redundant dictionary can be reformulated under the *maximum-a-posteriori-probability* (MAP) estimation framework that searches an estimate of the target coefficient vector $\hat{\mathbf{x}}$ by solving

$$\hat{\mathbf{x}} \quad = \quad \arg\max_{\mathbf{x}} \left[\log p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) + \log p_{\mathbf{x}}(\mathbf{x})\right], \qquad (1)$$

where $\log p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})$ is the *log-likelihood* function that encloses the relationship between the the noisy observations $\mathbf{y}$ and target vector $\mathbf{x}$, which is, in general, associated to the statistical model that best describes the additive noise $\boldsymbol{\nu}$ embedded in the signal acquisition environment; and $p_{\mathbf{x}}(\mathbf{x})$ is the prior probability density function (pdf) that is derived from the statistical characterization of a previously known feature in the target coefficient vector $\mathbf{x}$ [1].

Most recovery methods obtain an estimate of the desired sparse vector, $\hat{\mathbf{x}}$, by solving an optimization problem that is tightly attached to the specifications of the additive contamination. Thus, if $\boldsymbol{\nu}$ follows a zero-mean Gaussian distribution with variance σ^2, $p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) \propto \exp(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2)$, leading to the ℓ_2-norm of the data fitting error as a performance criterion to be minimized. Furthermore, since the goal is to find the sparsest representation of the desired vector, a sparsity constraint on the elements in $\mathbf{x}$ is imposed, reducing the recovery task to solve an ℓ_0-regularized least square (ℓ_0-LS) regression problem [2]. However, when the underlying noise is better characterized by heavy-tailed distributions, the performance of the ℓ_0-LS approach tends to degrade notably, leading to incorporate incorrect atoms in the recovered representation [3]. This is due to the fact that the ℓ_2-norm is highly sensitive to outliers or gross errors in the observed data, raising the need to develop robust sparse signal representations in overcomplete dictionaries that recover reliable versions of the target signal in the presence of impulsive noise.

In this paper, a coordinate descent algorithm for robust sparse signal representation in overcomplete dictionaries is proposed. This algorithm estimates the underlying coefficient vector addressing the recovering issue as an ℓ_1-regularized least absolute deviation (ℓ_1-LAD) regression problem, leading to the weighted median (WM) as the optimal estimator for computing each single coefficient, where the regularization effect reduces to append, in the WM operation, a zero-valued sample weighted by an adaptive regularization term. Furthermore, a continuation strategy is incorporated in order to find the best regularization parameter at the current level of noise, avoiding, thus, the use of cross-validation methods commonly used in regularized optimization problems. The paper is organized as follows. Section 2 describes the ℓ_1-LAD regression problem, and a coordinate descent algorithm for robust sparse signal recovery is outlined in Section 3. Section 4 shows some simulation results and concluding remarks are summarized in Section 5.

2. THE ℓ_1-LAD REGRESSION PROBLEM

A first approach to develop robust sparse signal recovery algorithms in redundant dictionaries consists in replacing the ℓ_2-norm by the ℓ_1-norm in the data fitting term. This approach naturally emerges when the additive noise is modeled as i.i.d. samples that follow a zero-mean Laplacian distribution with scale parameter $\sigma_{\boldsymbol{\nu}}$. Therefore, the sparse signal representation reduces to the solution of the following optimization problem,

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_1 + \lambda\|\mathbf{x}\|_0\right\}, \qquad (2)$$

where $\|\cdot\|_0$ is the ℓ_0-pseudonorm that counts the number of nonzero elements in the target vector. However, since ℓ_0-pseudonorm is a noncovex function, minimizing (2) becomes in a combinatorial NP-hard optimization problem [4]. To overcome this drawback, convex approximations to the ℓ_0-pseudonorm can be also considered in order to yield a more tractable mathematical analysis and fast recovering algorithms, much like it is done in the ℓ_2-norm based recovery algorithms [5, 6]. To this end, the sparsity inducing term can be relaxed using the ℓ_1-norm of the desired coefficient vector $\mathbf{x}$, leading, thus, to

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_1 + \lambda\|\mathbf{x}\|_1\right\}, \qquad (3)$$

This research is supported by the Council of Scientific, Humanistic, Technological, and Artistic Development of the University of Los Andes (CDCHTA-ULA) under the project I-1336-12-02-B

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

which is commonly referred to as ℓ_1-regularized LAD regression problem (ℓ_1-LAD) [7]. It can be easily shown that, under the MAP estimation perspective, the regularization parameter λ can be specified as the ratio of the dispersion parameter of the noise distribution and the dispersion constant of the prior model, i.e. $\lambda = \frac{\sigma_{\boldsymbol{\nu}}}{\sigma_{\mathbf{x}}}$, and as will be detailed later, a more realistic prior model, suitable to the statistics of each transformed coefficient, will be required for determining an accurate estimation.

Several algorithms have been proposed recently that solve the ℓ_1-LAD regression problem, where LAD-Lasso [8] and ℓ_1-ℓ_1 matching pursuit [9] are just two optimization strategies for robust recovering. However, the effect of the regularization term on determining each target coefficient is not clear. To gain some insight into the effect of the regularization term on each component of the target vector $\mathbf{x}$, let's use the coordinate descent framework to minimize (3). To this end, assume that the k-th coefficient of the target vector, x_k, is iteratively updated keeping fix the remained entries of $\mathbf{x}^{(m)}$, where m is an iteration index. Furthermore, assume for now that the others entries of the target vector $\mathbf{x}^{(m)}$ are somehow known. Therefore, the N-dimensional ℓ_1-LAD optimization problem reduces to the following 1-D minimization problem:

$$\begin{aligned}\hat{x}_k^{(m+1)} &= \arg\min_{x_k} \left\|\mathbf{y} - \mathbf{A}\mathbf{x}^{(m)} + \mathbf{a}_k x_k^{(m)} - \mathbf{a}_k x_k\right\|_1 \\ &\cdots + \lambda \left\|\mathbf{x}^{(m)}\right\|_1, \end{aligned} \tag{4}$$

where $\mathbf{a}_k$ is the k-th column-vector in $\mathbf{A}$; and the expression $\mathbf{A}\mathbf{x}^{(m)} - \mathbf{a}_k x_k^{(m)}$ cancels out the contribution of the k-th entry, obtained at the previous iteration, in the current estimation. The minimization of Eq. (4) can be rewritten as follows

$$\begin{aligned}\hat{x}_k^{(m+1)} &= \arg\min_{x_k} \sum_{i=1}^{M} |a_{k,i}| \left| \frac{\left(\mathbf{y} - \mathbf{A}\mathbf{x}^{(m)} + \mathbf{a}_k x_k^{(m)}\right)_i}{a_{k,i}} - x_k \right| \\ &\cdots + \lambda |x_k| . \end{aligned} \tag{5}$$

Upon closer examination of Eq. (5), it can be noticed that the first term on the right-hand side is a sum of weighted absolute deviations, where $\frac{\left(\mathbf{y}-\mathbf{A}\mathbf{x}^{(m)}+\mathbf{a}_k x_k^{(m)}\right)_i}{a_{k,i}}$ for $i = 1, 2, \ldots, M$ are the samples, $|a_{k,i}|$ is the weight vector, and x_k plays the role of a location parameter under the maximum likelihood estimation (MLE) approach. Further simplication can be achieved by noticing that the regularization term $\lambda |x_k|$ can be merged into the summation of weighted absolute deviations as follows

$$\hat{x}_k^{(m+1)} = \arg\min_{x_k} \sum_{i=1}^{M+1} |w_i| \, |Y_i - x_k| , \tag{6}$$

where

$$Y_i = \begin{cases} \frac{\left(\mathbf{y}-\mathbf{A}\mathbf{x}^{(m)}+\mathbf{a}_k x_k^{(m)}\right)_i}{a_{k,i}} & \text{for } i = 1, 2, \ldots, M \\ 0 & i = M+1 \end{cases} \tag{7}$$

and

$$w_i = \begin{cases} |a_{k,i}| & \text{for } i = 1, 2, \ldots, M \\ \lambda & i = M+1 \end{cases} . \tag{8}$$

The solution to this minimization problem turns out to be the WM operator, where $Y_i|_{i=1}^{M+1}$ are the data samples; $w_i|_{i=1}^{M+1}$ are the weights, and $\hat{x}_k^{(m+1)}$ is the weighted median output, in other words,

$$\hat{x}_k^{(m+1)} = \text{MEDIAN}\left(w_i \diamond Y_i|_{i=1}^{M+1}\right) \tag{9}$$

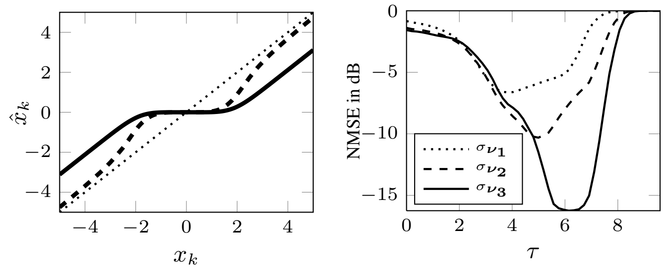

Fig. 1. (a) LUT curves with fixed λ (solid), and with adaptive λ (dashed); (b) NMSE in dB versus τ for three different noise dispersion parameters.

where $\diamond$ is defined as $w_i \diamond Y_i = \overbrace{Y_i, Y_i, \ldots, Y_i}^{w_i \text{ times}}$ [10]. Note in (8) that the weights are the entries of the k-th column of the dictionary expanded by a weight that takes the value of the regularization parameter, which is, in turns, associated to a zero-valued sample. Note also that, the sample data is a shifted-and-scaled version of the observed noisy signal. Thus, the regularization process reduces to appending a zero-valued sample weighting by the regularization term λ, in the WM operation; where a large value for λ implies a large weighting on the zero, favoring the sparsity of the solution, on the other hand, small values for λ implies less influence of the zero-valued sample on the estimation of $\hat{x}_k^{(m+1)}$, leading to a WM output driven by the shifted-and-scaled version of measurements weighted by the k-th column of the dictionary, $\mathbf{a}_k$.

3. THE ROBUST REPRESENTATION ALGORITHM

An intuitive recovering procedure can be developed based on the coordinate descent approach for solving the ℓ_1-LAD regression problem. Basically, this procedure iteratively estimates each entry of the target vector, $\hat{\mathbf{x}}$, applying the WM operation over the augmented vector defined in (7) using the weights given by (8), taking into account, in the definition of Y_i, the previously estimated signal coefficients. However, this approach has a zero attracting effect on the estimated nonzero coefficients $\hat{x}_k$, much like the effect of the soft thresholding operation has in the context of signal denoising. In order illustrate this behavior, look-up-table (LUT) curves are built from simulations. A sparse signal in the discrete cosine transform (DCT) dictionary is generated at each simulation trial, with length $N = 256$ and sparsity $S = 8$; where seven nonzero entries are drawn from a zero-mean Gaussian distribution with unit variance, and the remained nonzero coefficient is controlled and adjusted in a dynamic range. The 8-sparse DCT vector is projected to the discrete-time domain, and then, target coefficients of the projected noiseless signal are estimated using the heuristic coordinate descent approach, and the value of the corresponding 8-th controlled component of the sparse vector is recorded. For each controlled value, the procedure described above is performed 1000 trials. The averaged LUT plot for a fixed regularization parameter, $\lambda = 8$, is depicted in the solid curve of Fig. 1(a). As can be observed, the solution of the ℓ_1-LAD optimization problem, using the coordinate descend approach with the same fixed regularization parameter for all target coefficients, behaves in similar way than a soft-thresholding function, where large coefficient values are biased whereas small values are set to zero.

Upon a closer examination of Eq. (5), the same fix value of the regularization parameter responds to the assumption that each target coefficient x_k follows a Laplacian distribution with common dispersion parameter $\sigma_{\mathbf{x}}$. However, recent work in the context of DCT coefficient distribution has reported extensive numerical examples that show a high variability in the scaling constant for different DCT coefficients in a large image database [11], leading to the specification of an adjustable dispersion parameter σ_{x_k} for each desired coefficient x_k, and consequently, a particular regularization parameter. Exploiting this fact and under the MAP principles, the estimation of the scale parameter σ_{x_k} reduces to solve the following optimization problem,

$$\hat{\sigma}_{x_k} = \arg\min_{\sigma_{x_k}} \left\{ \frac{1}{\sigma_{\boldsymbol{\nu}}} \sum_{i=1}^{M} |(\mathbf{y}-\mathbf{A}\mathbf{x})_i| - M \log\left(\frac{1}{2\sigma_{\boldsymbol{\nu}}}\right) \cdots + \sum_{j=1}^{N} \frac{|x_j|}{\sigma_{x_j}} - \sum_{j=1}^{N} \log\left(\frac{1}{2\sigma_{x_j}}\right) \right\} \quad (10)$$

leading to $\hat{\sigma}_{x_k} = |x_k|$. Notice that the estimation of σ_{x_k} is subject to the previous knowledge of the target coefficient x_k, whose prediction is the main objective of this work. However, in the context of robust regression methods, robust estimators have been proposed, such as unpenalized LAD estimator [8], and unpenalizaed Huber estimator [12], for estimating the adaptive regularization term λ_k that is associated with each target coefficient x_k. Since λ_k depends on previous estimations of the target coefficient x_k, it is coherent to resort to a nested iterative procedure, where the inner loop updates both each desired coefficient x_k and its corresponding regularization parameter λ_k; and the outer loop refines the overall vector estimate. In other words, the adjustable regularization term for the k-th entry can be obtained as

$$\hat{\lambda}_k^{(m+1)} = \frac{\sigma_{\boldsymbol{\nu}}}{\hat{\sigma}_{x_k}} = \frac{\sigma_{\boldsymbol{\nu}}}{\left|\hat{x}_k^{(m)}\right|}. \quad (11)$$

where m is the outer iteration index. Therefore, an estimate of each coefficient that solves the 1-D optimization problem reduces to apply the weighted median operator over the same sample vector specified in Eq. (7) weighted by

$$w_i = \begin{cases} |a_{k,i}| & \text{for } i = 1, 2, \ldots, M \\ \hat{\lambda}_k^{(m+1)} & i = M+1 \end{cases}, \quad (12)$$

where $\hat{\lambda}_k^{(m+1)} = \frac{\tau}{\left|x_k^{(m)}\right|}$, and $\tau \in \mathbb{R}^+$ is a parameter that is closely related to the common dispersion of the noise distribution. Notice that, since the overall objective vector has a high degree of sparsity, a roughly update of the regularization parameter is specified using $\lambda_k^{(m+1)} = \frac{\tau}{\varepsilon + \left|x_k^{(m)}\right|}$, where the constant $\varepsilon \leq 1 \in \mathbb{R}^+$ is included to avoid division by zero.

Including an adaptive regularization parameter in the estimation of each desired coefficient has a similar effect than a generalized thresholding function, where each distinct regularization term has different influence depending on the target coefficient estimate at the previous iteration $x_k^{(m)}$. For instance, a large values of λ_k —related to a small absolute values of $x_k^{(m)}$— favors a zero-valued output in the estimation since the zero-valued sample is largely influence by the corresponding weight λ_k; on the other hand, a small value of λ_k —associated with a large value of $|x_k^{(m)}|$— reduces the zero attracting effect approaching the estimation to the unpenalized LAD. In order to show this behavior, LUT curves are built using an adaptive regularization term λ_k, in similar conditions than those used for fix λ, with the additional outer iteration parameter $itmax = 4$. The averaged LUT plot for an adaptive regularization parameter λ_k is displayed in the dashed curve of Fig. 1(a). As can be seen in this curve, the estimation using an adjustable λ_k behaves in similar way than a generalized thresholding function, where large coefficients have less bias than the solution obtained using a fixed regularization parameter, approaching the estimated coefficient to the unpenalized LAD solution. However, this thresholding behavior is less sensitive than the hard-thresholding operation, where small variations of noisy coefficients does not yield a significant changes in the estimation [13].

```
Input     : y, A, τ^(0), itmax, tolerance, β
x̂^(0) ⟵ 0_N;
m = 0;
error = ||y − A x̂^(0)||_2^2 / ||y||_2^2;
while m < itmax and error > tolerance do
    τ^(m) = τ^(0) β^m;
    for k = 1 to N do
        λ_k^(m) = τ^(m) / (ε + |x̂_k^(m)|);
        Y_i = { (y − A x̂ + a_k x̂_k^(m))_i / a_{k,i}   for i = 1, 2, ..., M
              { 0                                   i = M + 1              ;
        w_i = { |a_{k,i}|    for i = 1, 2, ..., M
              { λ_k^(m)      i = M + 1              ;
        x̂_k^(m+1) = MEDIAN( w_i ◇ Y_i |_{i=1}^{M+1} );
    end
    error = ||y − A x̂^(m+1)||_2^2 / ||y||_2^2;
    m = m + 1;
end
output    : x̂ = [x̂_1, x̂_2, ..., x̂_N]^T
```

Algorithm 1: Coordinate descent approach with an adaptive λ and continuation strategy.

An additional parameter to be considered, for recovering the optimal version of target coefficients, is the common constant τ. As mentioned above, τ is closely related to the scale parameter $\sigma_{\boldsymbol{\nu}}$ of the Laplacian distribution of the additive noise in the sensing model, and therefore, an appropriate selection of this constant is subject to the noise levels that corrupts the underlying signal, that, in general, is unknown. In order to observe the influence of τ on the accuracy of the recovered entries, the ensemble average of the normalized mean square error (NMSE), in dB, of the recovered coefficients versus τ is built from simulations. More precisely, an 8-sparse signal of length 512 is generated, where the nonzero coefficients are drawn from a zero-mean Gaussian distribution with unit variance, and this sparse representation is then projected using a redundant DCT dictionary of size 256×512. At each realization, the projected signal is corrupted with additive noise, whose samples follow a zero-mean Laplacian distribution with a specific value of the dispersion parameter $\sigma_{\boldsymbol{\nu}}$. The nested iterative procedure —with a specific value of the regularization constant τ and the iteration parameter fixed to $itmax = 10$— is implemented for recovering an estimate of the target coefficients from noisy measurements. After that, the mean square error of the recovered coefficient vector normalized by the original vector energy is obtained as a performance measure. For each simulation trial, at a specific value of τ, a new noise vector is added to the projected noiseless signal; and for each fixed value of τ, 1000 simulation trials are performed. Figure 1(b) shows the NMSE, in dB, as τ changes for three different scale parameters ($\sigma_{\boldsymbol{\nu}_1} = 1.0 \times 10^{-1}$, $\sigma_{\boldsymbol{\nu}_2} = 5.0 \times 10^{-2}$, $\sigma_{\boldsymbol{\nu}_3} = 2.5 \times 10^{-2}$). Note that each curve, at a specific noise level, reaches the smallest value of the NMSE at a different value of τ. This confirms the functional relationship between the noise variance and τ. Thus, to obtain an accurate estimate of the target vector, the noise variance should be estimated and a suitable value for τ must be selected, alternatively, as described next, we use a continuation approach to define τ.

In order to adapt the recovering algorithm to the noise levels,

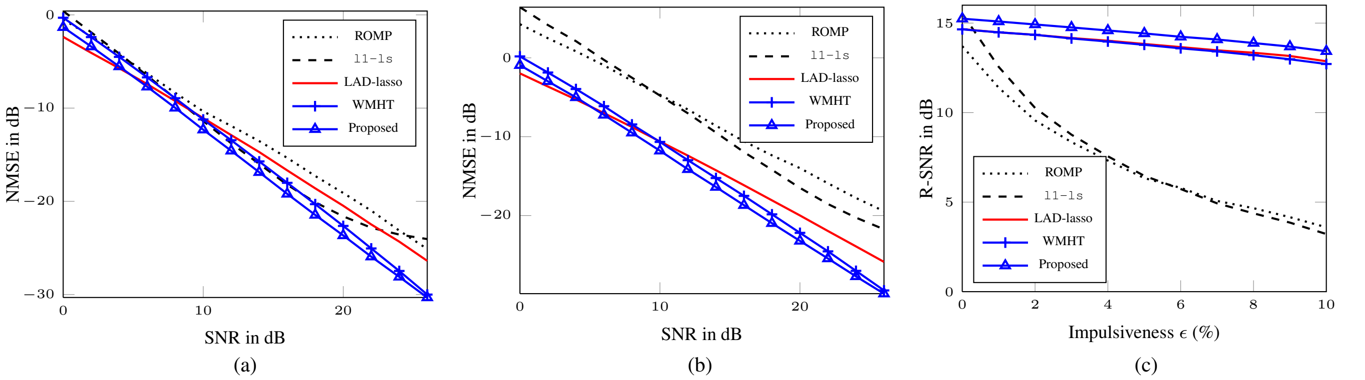

Fig. 2. NMSE, in dB, of the recovered coefficient vector versus SNR of measurements corrupted with additive noise using: (a) Gaussian distribution, (b) e-contaminated normal with $\epsilon = 3\%$; (c) R-SNR of the target signal versus the impulsive level of e-contaminated additive noise.

we treat τ as a tuning parameter whose value decrease as the iterative algorithm progresses. More precisely, the algorithm starts with a relative large value of $\tau^{(0)}$ favoring the sparsity in the solution, and identifying only the most significant nonzero coefficients in the first iteration, but possibly with large bias. Subsequently, as the iterative algorithm progresses, $\tau^{(m)}$ is slowly reduced, new smaller entries are identified, and the values of the previous identified nonzero coefficients are refined. The algorithm continues until a total number of iterations is reached or until an error value is achieved. A pseudo-code of this procedure is shown in Algorithm 1. Note that, for a new value of $\tau^{(m)}$ we are solving a new optimization problem taking into account the solution at the previous iteration, with $\tau^{(m-1)}$, as initial starting point for the new outer iteration loop.

4. SIMULATION RESULTS

We test the proposed method as a signal reconstruction algorithm in the compressive sensing framework. For all simulation sets, each target noiseless signal has sparse coefficients in the canonical domain; the nonzero coefficients are i.i.d. samples that follow a zero-mean Gaussian distribution with unit variance; and the components of the measurement matrix $\mathbf{A}$ are drawn from zero-mean Gaussian model with $\|\mathbf{a}_k\|_2^2 = 1$ for $k = 1, \ldots, N$. The setting for executing the proposed algorithm is $\varepsilon = 0.01$, $itmax = 100$, and $tol = 10^{-6}$. Also, we compare the performance of the proposed approach to those yielded by two methods that minimize the ℓ_2-norm of the data fitting term [regularized orthogonal matching pursuit (ROMP) [14] and ℓ_1-regularized least square (`l1-ls`) [6]] and two algorithms that solve ℓ_1-LAD regression problem [LAD-lasso [8], and a previously proposed algorithm denoted as weighted median hard thresholding (WMHT) [15]].

First, a 25-sparse signal of length $N = 512$ is built, and the projected signal of length $M = 256$ is corrupted with additive noise that obeys statistical models with two different distribution tails: Gaussian and e-contaminated normal distributions; where e-contaminated normal is a mixed noise model that merges Gaussian distribution and sparse gross errors. More precisely, e-contaminated normal obeys the following model $f_\eta(\eta) = (1-\epsilon)\mathcal{N}(0,\sigma_1) + \epsilon\mathcal{N}(0,\sigma_2)$, where σ_1 is set to the desired signal-to-noise ratio (SNR) whereas $\sigma_2 = 100\sigma_1$; ϵ is the amount of gross errors; and SNR is related to σ_1 by means of SNR $= \frac{(\mathbf{Ax})^{\mathbf{T}}(\mathbf{Ax})}{(M\sigma_1)}$.

Fig. 2(a) depicts the curves of NMSE of the recovered coefficient vector versus SNR of measurements yielded by the various reconstruction methods for the additive Gaussian noise. Each point in the curves are obtained by averaging 1000 trials of the respective experiment. Note that, the proposed algorithm outperforms other methods in almost all values of SNR, only at small values of SNR (SNR < 4) LAD-lasso yields better performance than our approach. Fig. 2(b) displays the curves of NMSE of the recovered coefficient vector versus SNR of measurements corrupted with e-contaminated normal distribution, setting the amount of gross errors at a 3%. As can be observed, the methods that solve the ℓ_1-LAD regression problem outperform those yielded by algorithms that minimize the ℓ_2-norm of the data fitting error, where robust methods achieve a gap of improvement of about 3 dB. Also, LAD-lasso exhibits less errors for small values of SNR (SNR < 4), and, as can be noticed in Fig. 2(a) and Fig. 2(b), the proposed algorithm outperforms WMHT for the entire of the depicted SNR interval. This improvement is due to our approach induces a generalized thresholding function on each nonzero estimate, yielding an unbiased coefficient whose output is less sensitive to small levels of noise than the hard thresholding operation.

Finally, to illustrate the robustness of the proposed method to outliers, Fig. 2(c) shows reconstruction SNR (R-SNR) obtained for the projected signal using an increasing set of impulsive noise levels; where the R-SNR is just the negative of NMSE in dB. A 25-sparse signal of length $N = 1024$ is generated, and then, the projected signal of length $M = 256$ is corrupted with additive noise that obeys an e-contaminated normal model, at a specific level of impulsive noise. For each value of impulsive noise level ϵ, each point in the curves is obtained by averaging 1000 realizations of the linear model. As can be observed, the proposed algorithm outperforms the others recovery techniques in presence of impulsive noise, where the recovered methods based on the minimization of the ℓ_2-norm notably degrades this performance as the impulsive noise increase.

5. CONCLUSIONS

In this paper, a coordinate descent approach for robust sparse signal representation has been presented. More specifically, this approach addresses the recovering of the target coefficient vector as a ℓ_1-LAD optimization problem; where each coefficient is robustly estimated as the WM of a scaled-and-shifted version of noisy observations weighted by the magnitude of the corresponding atom components. An adaptive regularization parameter is included in the WM as an appended weight that influences the zero-valued sample, inducing a generalized thrseholding operation on the estimation of each target coefficient. A continuation strategy is also incorporated in order to adapt the the estimation of the target vector to a current noise level. In the context of compressive sensing, extensive simulations show that the proposed algorithm outperforms other approach in the presence of impulsive noise. Furthermore, our robust approach exhibits a competitive performance when the underlying additive noise follows a Gaussian distribution.

6. REFERENCES

[1] M. Elad, "Why simple shrinkage is still relevant for redundant representations," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5559–5569, Dec. 2006.

[2] D. Donoho and M. Elad and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[3] R. and K. Barner and T. C. Aysal, "Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise," *IEEE J. Sel. Topics Signal Process.*, , no. 2, pp. 392–408, Apr. 2010.

[4] X. Huo and J. Chen, "Hardness on numerical realization of some penalized likelihood estimators," *Industrial and Technology Statistics Group, Georgia Inst. Technol., Atlanta, GA, Tech. Paper*, vol. 8, pp. 2006, 2006.

[5] M. Zibulevsky and M. Elad, "L1-l2 optimization in signal and image processing," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 76–88, May 2010.

[6] S-J Kim and K. Koh and M. Lustig and S. Boyd and D. Gorinevsky, "An interior-point method for large-scale l1-regularized least squares," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 606–617, Dec. 2007.

[7] L. Wang and M. D. Gordon and J. Zhu, "Regularized least absolute deviations regression and an efficient algorithm for parameter tuning," in *Data Mining, 2006. ICDM'06. Sixth International Conference on*. IEEE, pp., 690–700.

[8] H. Wang and G. Li and G. Jiang, "Robust regression shrinkage and consistent variable selection through the lad-lasso," *Journal of Business & Economic Statistics*, vol. 25, no. 3, pp. 347–355, 2007.

[9] J-J Fuchs, "Fast implementation of a $\ell_1 - \ell_1$ regularized sparse representations algorithm.," in *Proceedings of the 2009 IEEE International Conference on Acoustic, Speech, and Signal Processing*, Washinton, DC, USA, pp., 3329–3332.

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

[11] I. Ramírez and G. Sapiro, "Universal regularizers for robust sparse coding and modeling," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3850–3864, Sept. 2012.

[12] S. Lambert-Lacroix, and L. Zwald, "Robust regression through the hubers criterion and adaptive lasso penalty," *Electronic Journal of Statistics*, vol. 5, pp. 1015–1053, 2011.

[13] A. Gholami and S. M. Hosseini, "A general framework for sparsity-based denoising and inversion," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5202–5211, Nov. 2011.

[14] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Foundations of computational mathematics*, vol. 9, no. 3, pp. 317–334, 2009.

[15] J. L. Paredes and G. Arce, "Compressive sensing signal reconstruction by weighted median regression estimates," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2585–2601, June 2011.