ROBUST GREEDY ALGORITHMS FOR COMPRESSED SENSING

S. Alireza Razavi *Esa Ollila* *Visa Koivunen*

Department of Signal Processing and Acoustics, SMARAD CoE, Aalto University, Espoo, Finland

ABSTRACT

The problem of sparse signal reconstruction in the presence of possibly impulsive noise is studied. The state-of-the-art greedy algorithms, Iterative Hard Thresholding (IHT), Orthogonal Matching Pursuit (OMP), and Compressive Sampling Matching Pursuit (CoSaMP) are robustified in order to cope with impulsive noise environments and outliers. We employ robust weighting of the residuals and replace the least-squares estimates by M-estimates of regression. Also a robust M-estimation based ridge regression is considered and shown to possess high potential when utilized in CS algorithms.

***Index Terms*—** compressive sensing, matching pursuit, greedy algorithms

1. INTRODUCTION

Compressed sensing (CS) has attracted a lot of research interest recently; see [3] and references therein. CS exploits the fact that many signals that arise in nature are sparse, *i.e.* they can be represented as a linear combination of relatively a small number of elementary signals called *atoms*. The problem may be formulated as follows. Let $\mathbf{s} \in \mathbb{R}^N$ be a K-*sparse* signal with $K \ll N$ nonzero elements, $\mathbf{\Phi}$ be an $M \times N$ *measurement matrix*, with $M < N$, and

$$\mathbf{y} = \mathbf{\Phi}\mathbf{s} + \boldsymbol{\epsilon}, \tag{1}$$

be the measurements corrupted by noise $\boldsymbol{\epsilon}$. The objective is to reconstruct the signal vector $\mathbf{s}$ knowing the measurements $\mathbf{y}$, the measurement matrix $\mathbf{\Phi}$, and the sparsity K. The following optimization problem gives a K-sparse solution $\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{s}\|_2$ subject to $\|\mathbf{s}\|_0 \leq K$, where $\|\cdot\|_p$ denotes the ℓ_p norm. This optimization is known to be NP-hard. Hence suboptimal reduced complexity CS reconstruction algorithms have been proposed [16]. These can be roughly divided into two classes, *convex-relaxation algorithms* and *greedy algorithms*. Methods in the convex-relaxation class are based on replacing the nonconvex ℓ_0 norm of the signal by ℓ_1 norm (or some other suitable norm or pseudo-norm) and solving a convex optimization problem instead of a NP-hard problem; see, e.g., [2, 7]. Greedy methods are iterative methods that contain a step in which the column(s) of $\mathbf{\Phi}$ which are most correlated with the current residual vector are identified. CoSaMP [12], OMP [15], Subspace Pursuit [6], and IHT [1] all belong to this class.

Recently CS in the presence of impulsive noise has been addressed by several authors. In [4], a reconstruction algorithm was proposed based on the minimization of the ℓ_1 norm subject to a nonlinear constraint on the signal based on the Lorentzian pseudo-norm, but the algorithm suffers from high complexity. Later in [5], the authors proposed a computationally less demanding approach in which IHT was robustified by introducing the Lorentzian pseudo-norm of residuals into each iteration; called Lorentzian IHT (LIHT) hereafter. In [13], the authors used ℓ_0-regularized least absolute deviation (LAD) regression model and proposed an algorithm utilizing weighted median regression to obtain an approximate solution. The method is computationally demanding. Moreover, the selection of parameters of the algorithm, namely number of iterations and decaying speed, is not an easy task and is done by trial and error.

In this paper, we propose robust versions of CoSaMP, IHT and OMP. We utilize robust weighting of the residuals and robust regression in place of least squares (LS) regression.

The proposed methods are compared with LIHT [5] under two different noise models. Section 2 reviews the robust M-estimation of regression coefficients and ridge regression approach. Then, in Section 3, we review the CoSaMP, IHT and OMP algorithms and propose their robust modifications. In Section 4, we study the performance of the proposed methods by simulations. Finally, Section 5 concludes the paper.

Notations: For a vector $\mathbf{u}$, a matrix $\mathbf{X}$, and a support set A, u_i denotes the i-th component of $\mathbf{u}$, $\mathbf{x}_i$ denotes the (transposed) ith row vector of $\mathbf{X}$, vector $\mathbf{u}_A$ (matrix $\mathbf{X}_A$) denotes the components (columns) of $\mathbf{u}$ ($\mathbf{X}$) corresponding to support set A, $H_K(\cdot)$ denotes the operator which sets all but the largest (in magnitude) K elements of its vector-valued argument to zero and $\mathrm{I}(\cdot)$ denotes the indicator function.

2. M-ESTIMATES OF REGRESSION

Suppose now that a set T of size L (where $K \leq L << M$) has been selected with an aim that it should include the signal support, $\mathrm{supp}(\mathbf{s}) \subseteq T$. Hence we can consider the linear regression model $\mathbf{y} = \mathbf{\Phi}_T\mathbf{s}_T + \boldsymbol{\epsilon}$ with more measurements than regressors and obtain a robust estimate $\hat{\mathbf{s}}_T$ via M-regression method. Then an estimate of the signal support are the indices of the K-largest elements of $\mathbf{s}_T$. For notational convenience, we suppress the subscript T from $\mathbf{\Phi}$ and $\mathbf{s}$. The error terms are assumed to be i.i.d. from a symmetric continuous distribution with an unknown scale parameter $\sigma > 0$.

Let $e_i = e_i(\mathbf{s}) = y_i - \boldsymbol{\phi}_i^\top \mathbf{s}$ denote the ith residual for a candidate vector $\mathbf{s}$. At this point assume that the scale paremeter σ is known. Then the M-estimate of regression coefficient $\hat{\mathbf{s}}$ solves the *M-estimating equation*

$$\sum_{i=1}^{M} \psi(e_i/\sigma)\boldsymbol{\phi}_i = \mathbf{0}, \tag{2}$$

where ψ is continuous, odd function and preferably (for robustness) bounded.The obtained estimate often possesses a maximum likelihood (ML) interpretation if $\psi(e) \propto -\log f_0(e)$, where $f_0(\cdot)$ denotes the standard form ($\sigma = 1$) of the error density. Express $\psi(e) = e \cdot w(e)$ and write $w_i \triangleq w(e_i/\sigma)$ and $\mathbf{W} = \mathrm{diag}(w_1, \ldots, w_M)$. In case of zero residual ($e = 0$), we set $w(e) = 0$. Then (2) rewrites as

$$\hat{\mathbf{s}} = (\mathbf{\Phi}^\top \mathbf{W}\mathbf{\Phi})^{-1}\mathbf{\Phi}^\top \mathbf{W}\mathbf{y}. \tag{3}$$

In the case of normal (Gaussian) errors, the unique ML-estimate of $\mathbf{s}$ is obtained with $\psi(e) = e$ (so $w(e) = 1$), which correspond to the LS-estimates. Excluding the normal case, the estimating equation (3) is implicit since the weights w_i on the right hand side also depend on $\mathbf{s}$ through the residual $e_i = e_i(\mathbf{s})$. For example,

$$\psi_{\mathrm{C}}(e) = 2e/(1 + e^2), \quad w_{\mathrm{C}}(e) = 2/(1 + e^2), \tag{4}$$

are the respective psi- and weight functions of the Cauchy error distribution. Hence an algorithm to find the solution $\hat{\mathbf{s}}$ to (3) is needed. *Huber's weight function* is defined as

$$w_{\mathrm{H}}(e) \triangleq \min(1, k/|e|) \tag{5}$$

and $\psi_{\mathrm{H}}(e) = \mathrm{clip}_k(e) \triangleq e \cdot w_{\mathrm{H}}(e) = \max[-k, \min(k, e)]$. The popular *Tukey's biweight* (or bisquare) weight function is

$$w_{\mathrm{B}}(e) \triangleq \left\{ \left(1 - \frac{e^2}{c^2}\right)_+ \right\}^2, \tag{6}$$

with $\psi_{\mathrm{B}}(e) = e \cdot w_{\mathrm{B}}(e)$, where $a_+ = \max(0, a)$. Note that k and c above are user-defined tuning (threshold) constants that affect robustness (and efficiency) of the methods; see [10] for details. In our simulations we use the standard values $k = 1.345$ and $c = 4.685$ which yield 95% efficiency in Gaussian noise [10]. Huber's ψ function is also an MLE for the "least favorable distribution" but Tukey's ψ-function can not be linked with any error density and hence does not have an ML-interpretation. Functions (4)-(6) are depicted in Figure 1. Note that Huber's ψ-function can be interpreted as winzorizing ("clipping") function. Tukey's biweight penalizes the most large residuals as it strongly redescends to zero.

The M-estimate is commonly calculated by an Iteratively Re-weighted LS (IRLS) algorithm, denoted by $\mathrm{MFIT}(\mathbf{y}, \mathbf{\Phi}, w)$,

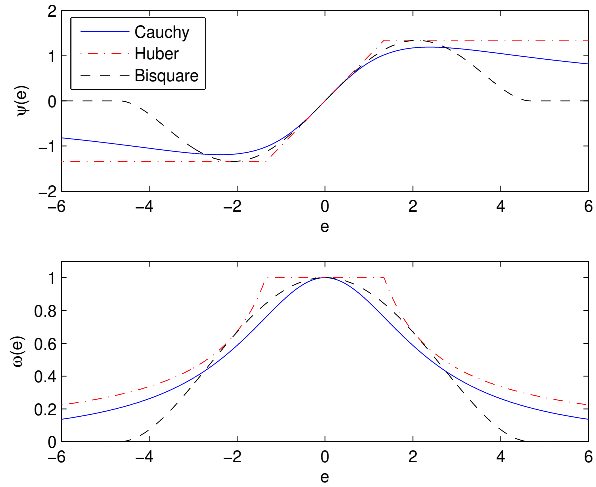

Fig. 1: Functions ψ and w where $\psi(e) = e \cdot w(e)$ of (4)-(6).

where w is the weight function. Since σ is unknown in practice, it is commonly replaced at each iteration by a robust estimate $\hat{\sigma}$ calculated from the current residuals. We use the median absolute deviation (MAD) $\hat{\sigma} = \mathrm{MAD}(\mathbf{e}) = 1.4286 \cdot \mathrm{med}_i(|e_i - \mathrm{med}_i(e_i)|)$ which is the default choise in `robustfit` routine of Matlab, where $\mathrm{med}(\cdot)$ denotes the median function. Another approach is to use joint M-estimation of regression and scale such as the method introduced in [11] called robust Cauchy-based M-estimation.

Robust ridge regression: In the $M < N$ case (more variables than responses), the *ridge regression estimate* [8] $\hat{\mathbf{s}}$ minimizes a penalized residual sum of squares $\sum_{i=1}^{M} e_i^2 + \lambda\|\mathbf{s}\|_2^2$. Then the bigger the ridge (shrinkage) parameter λ, the greater the amount of shrinkage of coefficients. See [9] for details on regularization parameter selection. The method is called *Tikhonov regularization* in many other fields. This approach can be generalized by considering a penalized loss function $\sum_{i=1}^{M} \rho(e_i/\sigma) + \lambda\|\mathbf{s}\|_2^2$, where $\psi = \rho'$. We thus define the ridge regression M-estimate (RRM) as the stationary point of the loss function, i.e. $\hat{\mathbf{s}}$ is the solution to the *ridge M-estimating equation* $-\sum_{i=1}^{n} \psi(e_i/\sigma)(\boldsymbol{\phi}_i/\sigma) + \lambda\mathbf{s} = \mathbf{0}$, i.e.:

$$\hat{\mathbf{s}} = (\mathbf{\Phi}^\top \mathbf{W}\mathbf{\Phi} + \lambda\sigma^2\mathbf{I})^{-1}\mathbf{\Phi}^\top \mathbf{W}\mathbf{y} \tag{7}$$

where $\mathbf{W}$ is as earlier; see also [14]. If $\psi(e) = e$ and $\sigma = 1$, then the regular ridge estimator is obtained. When robust weight functions (4)-(6) are used, then the estimating equation is implicit and one can employ IRLS approach but now replacing (3) by (7) at each iteration of the algorithm. Furthermore, we stop after $t_{\max}$ iterations to reduce the computation load and call the estimate as $t_{\max}$-step RRM-estimator. In our simulations, we use $t_{\max} = 5$ and $\lambda = 5.5$ which usually yield a sufficiently robust estimate. Again we utilize the MAD estimate $\hat{\sigma}$ to (re)scale the residuals at each iteration. Note that ridge regression cannot be used for sparse recovery alone as it does not provide sufficiently sparse solution. However, it could be used to find an estimate of a sparse support,

Table 1: Ridge Regression M-estimate (RRM) algorithm

Input: $\mathbf{y}$, $\mathbf{\Phi}$, and an initial estimate $\mathbf{s}^0$ of $\mathbf{s}$
Output: $\mathbf{s}^{t_{\max}}$ an estimate of $\mathbf{s}$
Initialization: $t \leftarrow 0$
repeat
$t \leftarrow t+1$
$\mathbf{e} \leftarrow \mathbf{y} - \mathbf{\Phi}\mathbf{s}^{t-1}$ {Update the residuals}
$\hat{\sigma} \leftarrow \mathrm{MAD}(\mathbf{e})$ {Calculate the robust estimate of scale}
$[\mathbf{w}]_i \leftarrow w(e_i/\hat{\sigma})$, $\mathbf{W} \leftarrow \mathrm{diag}(\mathbf{w})$ {Update the weights}
$\mathbf{s}^t \leftarrow (\mathbf{\Phi}^\top \mathbf{W}\mathbf{\Phi} + \hat{\sigma}^2\lambda\mathbf{I})^{-1}\mathbf{\Phi}^\top\mathbf{W}\mathbf{y}$ {weighted LS estimate}
until $t \geq t_{\max}$

e.g. as $\Omega = \mathrm{supp}(H_K(\hat{\mathbf{s}}))$. See also Table 4. The summary of the algorithm is given in Table 1.

3. ROBUST GREEDY CS RECONSTRUCTION

3.1. The robust CoSaMP algorithm

CoSaMP [12] is comprised of the following four steps:
Step 1: *Update the residuals* using the current estimate of $\mathbf{s}$.
Step 2: *Identify* the $2K$ columns of the measurement matrix correlated with the residual vector the most and *merge* it with the support of the current estimate of $\mathbf{s}$.
Step 3: *Estimate* the signal vector by LS-regression using the merged support set as the selected variables.
Step 4: *Prune the estimate* by retaining only the K largest (in magnitude) coefficients.

To robustify CoSaMP, we make three modifications to the method. First, we use a robust initial estimator $\mathbf{s}^0$ of $\mathbf{s}$. Note that the initial estimate is $\mathbf{s}^0 = \mathbf{0}$ in the original CoSaMP algorithm. Using an initial robust estimator both decreases the number of iterations of the algorithm and improves on the robustness and accuracy of the algorithm. Second, we identify largest correlations between the columns of $\mathbf{\Phi}$ and the *residual pseudo-values* $\mathbf{e}_\psi = \psi(\mathbf{e}/\hat{\sigma}) \triangleq (\psi(e_1/\hat{\sigma}), \ldots, \psi(e_M/\hat{\sigma}))^T$. Due to relation $\psi(e) = ew(e)$ this can be interpreted as downweighting the large residuals. For example, if we use Huber's psi-function $\psi(e) = \mathrm{clip}_k(e)$ we are winzorizing (or "clipping") after a threshold k. Also note that $\psi(e) = e$ (so $w(e) = 1$) in the original CoSaMP. Third, we replace the LS-estimator in *Step 3* by a robust M-estimator of regression that uses a robust weight function $w(\cdot)$ such as (4)-(6). The summary of the Robust CoSaMP algorithm is given in Table 2.

3.2. Robust Iterative Hard Thresholding Algorithm

Iterative Hard Thresholding (IHT) [1] is an efficient greedy algorithm which optimizes the cost function $J(\mathbf{s}) = \|\mathbf{y} - \mathbf{\Phi}\mathbf{s}\|_2$ under the constraint $\|\mathbf{s}\|_0 \leq K$. The K-sparse vector $\mathbf{s}$ is found by iterating $\mathbf{s}^{t+1} = H_K(\mathbf{s}^t - \mu^t \nabla J(\mathbf{s}^t))$ until convergence. Above μ^t is the step size and $\mathbf{s}^0 = \mathbf{0}$. Substituting $\nabla J(\mathbf{s}) = -\mathbf{\Phi}^\top \mathbf{e}$, where $\mathbf{e} = \mathbf{y} - \mathbf{\Phi}\mathbf{s}$, we obtain

$$\mathbf{s}^{t+1} = H_K(\mathbf{s}^t + \mu^t \mathbf{\Phi}^\top \mathbf{e}^t),$$

Table 2: Robust CoSaMP algorithm

Input: $\mathbf{y}$, $\mathbf{\Phi}$, sparsity level K and an initial estimate $\mathbf{s}^0$ of $\mathbf{s}$
Output: $\mathbf{s}^t$ an estimate of $\mathbf{s}$
Initialization: $t \leftarrow 0$
repeat
$t \leftarrow t+1$
Step 1:
$\mathbf{e} \leftarrow \mathbf{y} - \mathbf{\Phi}\mathbf{s}^{t-1}$ {Update the residuals}
$\hat{\sigma} \leftarrow \mathrm{MAD}(\mathbf{e})$ {Calculate the robust estimate of scale}
$\mathbf{e}_\psi \leftarrow \psi(\mathbf{e}/\hat{\sigma})$ {residual pseudo-values (downweighting) }
Step 2:
$\Omega \leftarrow \mathrm{supp}(H_{2K}(\mathbf{\Phi}^\top \mathbf{e}_\psi))$ {Identify the $2K$ largest correlations}
$T \leftarrow \Omega \cup \mathrm{supp}(\mathbf{s}^{t-1})$ {Merge the supports}
Step 3:
$\boldsymbol{\beta} \leftarrow \mathrm{MFIT}(\mathbf{y}, \mathbf{\Phi}_T, w)$ {Compute the M-estimates of regression }
$\mathbf{b} \leftarrow \mathbf{0}_{M\times 1}$, $\mathbf{b}_T \leftarrow \boldsymbol{\beta}$
Step 4:
$\mathbf{s}^t \leftarrow H_K(\mathbf{b})$ {Prune the signal estimate}
until $\frac{\|\mathbf{s}^t - \mathbf{s}^{t-1}\|}{\|\mathbf{s}^t\|} \leq \delta$ or $t > t_{max}$.

Table 3: Robust OMP algorithm.

Input: $\mathbf{y}$, $\mathbf{\Phi}$, sparsity level K **Output:** $\mathbf{s}^t$ an estimate of $\mathbf{s}$
Initialization: $t \leftarrow 0$, $\Omega^0 \leftarrow \{\}$, $\mathbf{e} \leftarrow \mathbf{y}$, $\mathbf{s} \leftarrow \mathbf{0}$
repeat
$t \leftarrow t+1$
$\hat{\sigma} \leftarrow \mathrm{MAD}(\mathbf{e})$ {Calculate the robust estimate of scale}
$\mathbf{e}_\psi \leftarrow \psi(\mathbf{e}/\hat{\sigma})$ { Calculate the residual pseudo-values }
$\Omega^t \leftarrow \Omega^{t-1} \cup \mathrm{supp}(H_1(\mathbf{\Phi}^\top \mathbf{e}_\psi))$ {Add the index of the atom correlated with the residual pseudo-values the most to the index set.}
$\boldsymbol{\beta} \leftarrow \mathrm{MFIT}(\mathbf{y}, \mathbf{\Phi}_{\Omega^t}, w)$ {Compute the M-estimates of regression }
$\mathbf{e} \leftarrow \mathbf{y} - \mathbf{\Phi}_{\Omega^t}\boldsymbol{\beta}$ {update the residuals}
until $\lvert\Omega^t\rvert \geq K$
$\mathbf{s}_{\Omega^t} \leftarrow \boldsymbol{\beta}$

As in M-regression, replacing the non-robust ℓ_2-objective function by a more general objective function $J(\mathbf{s}) = \sum_{i=1}^M \rho\big((y_i - \boldsymbol{\phi}_i^\top \mathbf{s})/\sigma\big)$, yields $\nabla J(\mathbf{s}) = -(1/\sigma^2)\mathbf{\Phi}^\top \mathbf{W}\mathbf{e}$, where $\mathbf{W}$ is the diagonal weight matrix defined as in Section 2 and σ is the scale. This ends up with

$$\mathbf{s}^{t+1} = H_K(\mathbf{s}^t + (\mu^t/(\hat{\sigma}^t)^2)\mathbf{\Phi}^\top \mathbf{W}^t \mathbf{e}^t), \tag{8}$$

as the update equation, where $\mathbf{W}^t$ is the diagonal matrix with weights $w_i^t = w(e_i^t/\hat{\sigma}^t)$, and as in M-regression, the practically unknown scale parameter σ is replaced by a robust estimate of it at each iteration t, e.g., $\hat{\sigma}^t = \mathrm{MAD}(\mathbf{e}^t)$.

Let $\mathbf{g}$ denote the negative gradient at the current estimate $\mathbf{s}^t$, i.e. $\mathbf{g} = (\hat{\sigma}^t)^{-2}\mathbf{\Phi}^\top \mathbf{W}^t \mathbf{e}^t$, and $\Gamma^t = \mathrm{supp}(\mathbf{s}^t)$. Assuming that we have identified the correct support at iteration t, one approach for finding a reasonable (namely, easily computable) step size μ^t for the gradient ascent direction $\mathbf{s}_{\Gamma^t} + \mu^t \mathbf{g}_{\Gamma^t}$ is minimizing the weighted sum of squared errors (WSSE) $\min_\mu \sum_{i=1}^M w_i^t (y_i - \boldsymbol{\phi}_i^\top(\mathbf{s}_{\Gamma^t}^t + \mu \mathbf{g}_{\Gamma^t}))^2$. After some elementary calculations

$$\mu^t = \frac{(\hat{\sigma}^t)^2 \mathbf{g}_{\Gamma^t}^\top \mathbf{g}_{\Gamma^t}}{\mathbf{g}_{\Gamma^t}^\top \mathbf{\Phi}_{\Gamma^t}^\top \mathbf{W}^t \mathbf{\Phi}_{\Gamma^t} \mathbf{g}_{\Gamma^t}}. \tag{9}$$

Table 4: Comparing support recovery $H_{2K}(\mathbf{\Phi}^T \mathbf{e}_\psi)$ of a sparse signal when $(N, M, K) = (512, 128, 5)$. Cauchy, biweight and RRM exhibit relatively high robustness in Cauchy noise while maintaining comparable performance to the ordinary methods in Gaussian noise.

	Gaussian Noise		Cauchy Noise	
ψ-function	PFI	NIC	PFI	NIC
identity (ordinary LS)	94.38	4.943	0.09	0.513
Cauchy	91.92	4.918	31.28	4.025
Biweight	91.6	4.915	29.78	3.986
Ordinary Ridge estimator	96.88	4.968	0.01	0.425
RRM (Cauchy)	92.46	4.924	37.2	4.141
RRM (Biweight)	89.09	4.886	33.84	4.06

When we use μ^t as above and the found support Γ^{t+1} of $\mathbf{s}^{t+1}$ is the same as Γ^t, then we are quaranteed to have a maximal reduction in WSSE. If $\Gamma^{t+1} \neq \Gamma^t$, then the optimality of this stepsize is no longer guaranteed. In this case, if the new objective function $\sum_{i=1}^{M} \rho((y_i - \mathbf{\Phi}_i^\top \mathbf{s}^{t+1})/\hat{\sigma}^{t+1})$ is smaller than the old objective function $\sum_{i=1}^{M} \rho((y_i - \mathbf{\Phi}_i^\top \mathbf{s}^{t})/\hat{\sigma}^{t})$, we still accept $\mathbf{s}^{t+1}$ as the new estimate, otherwise we set $\mu^t \leftarrow \mu^t/2$ and recalculate a new proposal from (8). This is continued until a new proposal with smaller objective function is obtained, which is then accepted as the new estimate $\mathbf{s}^{t+1}$.

3.3. The robust OMP algorithm

OMP [15] is an iterative greedy algorithm that selects at each step the column which is most correlated with the current residuals, and iterates it until K distinct atoms have been selected. Due to the lack of space we do not discuss the OMP algorithm in detail here and only introduce its robust version. Similar to robust CoSaMP, we use correlation between columns of $\mathbf{\Phi}$ and residual pseudo-values $\psi(\mathbf{e})$ and exploit robust M-regression for estimating the parameter vector in each step. The summary of robust OMP is given in Table 3.

4. SIMULATION RESULTS

The simulation part of this paper contains three experiments. In the first simulation example we study how the residual weight function w used in Step 1 affects the support recovery of a sparse signal in the identification step (*Step 2*) of the CoSaMP algorithm. In this experiment the length of signal is $N = 512$, the number of measurements is $M = 128$, the sparsity level is $K = 5$, the number of random trials is $Q = 10000$. The elements of the measurement matrix $\mathbf{\Phi}$ are drawn from $\mathcal{N}(0, 1)$ and its columns are normalized to have unit norm. Table 4 reports the average Percentage of Full Identification (PFI), PFI $\triangleq \frac{100}{Q} \sum_{q=1}^{Q} \mathrm{I}(\mathrm{supp}(\mathbf{s}^{[q]}) \subset \Omega^{[q]})$ and the average Number of Identified Components (NIC), NIC $\triangleq \frac{1}{Q} \sum_{q=1}^{Q} |\mathrm{supp}(\mathbf{s}^{[q]}) \cap \Omega^{[q]}|$ at the first iteration of the algorithm and with $\mathbf{s}^0 = \mathbf{0}$ given as the initial estimate. Above $\mathbf{s}^{[q]}$ and $\Omega^{[q]}$ are , respectively, the true signal and the the $2K$ support set calculated in Step 2 of the CoSaMP algorithm in the q-th trial. The amplitude of all K nonzero elements of $\mathbf{s}$ is set to 10, and the two noise models considered here are

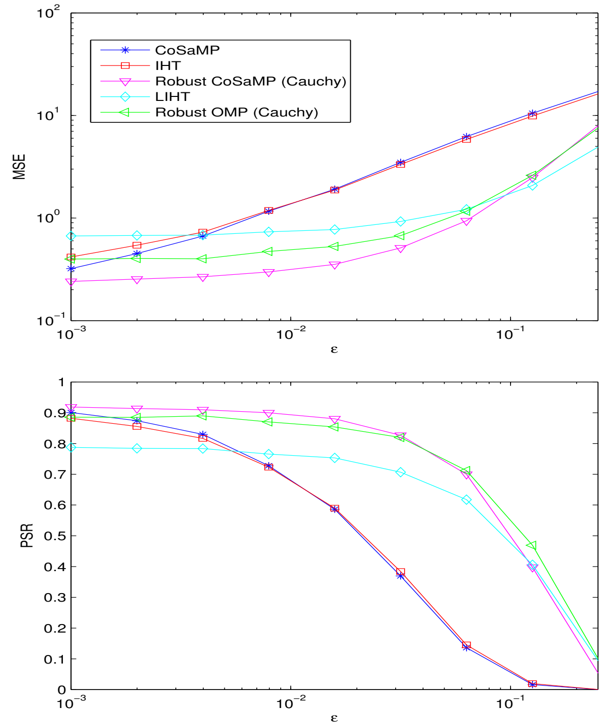

Fig. 2: MSE (up) and PSR (down) in reconstructing a sparse signal by five different sparse reconstruction methods in the presence of ε-contaminated noise. Robust CoSaMP and Robust OMP are both superior to LIHT unless for large ε, which less happens in pactice.

Cauchy distribution with scale $\sigma = 1$ and zero-mean Gaussian with variance one. Table 4 provides figures for the following functions: the original CoSaMP weight $\psi(e) = e$, the robust functions (4) and (6), and the ridge regression estimators discussed in Section 2. As can be seen, while in the Gaussian case the identity function ($\psi(e) = e$, so weight $w(e) = 1$) used in the original CoSaMP works very well (as expected since it corresponds to optimal ML-score function $\psi(\cdot)$ for Gaussian errors), its performance severely degrades in the presence of heavy-tailed Cauchy noise. On the other hand, RRM has the best performance in Cauchy noise. The performances of Cauchy (with weight function (4)) and biweight (with weight function (6)) are very close to RRM.

The goal of the next two examples is to evaluate the performance of robust CoSaMP, robust OMP, and robust IHT for two different non-Gaussian noise models. The measures of performance are the Mean Squared Error defined as MSE $\triangleq \frac{1}{Q} \sum_{q=1}^{Q} \|\mathbf{s}^{[q]} - \hat{\mathbf{s}}^{[q]}\|_2^2$, and the Proportion of full Support Recovery (PSR) defined as PSR $\triangleq \frac{1}{Q} \sum_{q=1}^{Q} \mathrm{I}(\mathrm{supp}(\mathbf{s}^{[q]}) = \mathrm{supp}(\hat{\mathbf{s}}^{[q]}))$ where $\hat{\mathbf{s}}^{[q]}$ is the reconstructed signal in the q-th trial. In both experiments we have $(N, M, K) = (1024, 64, 4)$ and the number of trials is $Q = 10000$. In the second experiment, we use the ε-contaminated noise model $(1 - \varepsilon)\mathcal{N}(0, \sigma_1^2) + \varepsilon\mathcal{N}(0, \sigma_2^2)$ for each component of the noise vector $\boldsymbol{\epsilon}$. We set $\sigma_1 = 1$, $\sigma_2 = 10$, and ε is varied from 10^{-3} to 0.25. Note that ε is

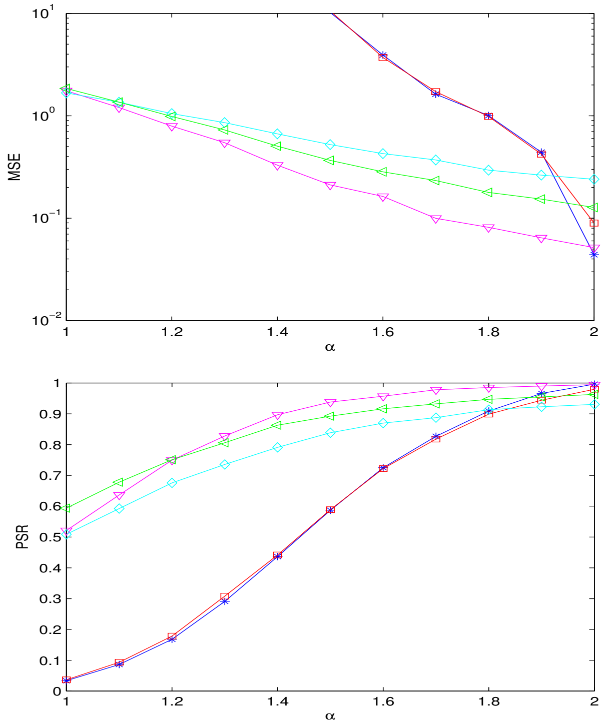

Fig. 3: MSE (up) and PSR (down) in reconstructing a sparse signal by five different sparse reconstruction methods in the presence of α-stable noise. Graphical conventions are the same as in figure 2. Robust CoSaMP and Robust OMP both surpass LIHT.

the probability of obtaining noise from an impulsive (higher variance) noise model and it describes the proportion of expected outliers in the sample. We compare five reconstruction algorithms: CoSaMP, IHT, Robust IHT with Cauchy weights (which coincides with LIHT [5]), robust OMP and robust CoSaMP both with Cauchy weights. The nonzero elements of $\mathbf{s}$ have fixed amplitude equal to 1. As can be seen in Fig. 2, the robust CoSaMP and robust OMP both outperform LIHT unless for large values of ε, which rarely happen in practice.

In the third experiment, we examine the performance in centered symmetric α-stable noise with scale parameter $\sigma = 0.2$ when α is varied from $\alpha = 1$ (Cauchy distribution) to $\alpha = 2$ (Gaussian distribution). As can be seen in Fig. 3, both robust CoSaMP and robust OMP are superior to LIHT.

5. CONCLUDING REMARKS

We proposed robust versions of three greedy CS reconstruction algorithms, the CoSaMP, OMP and IHT. The update of residuals were robustified by a robust weighting scheme and the non-robust LS estimation was replaced by robust M-estimation of regression. Simulations showed that the proposed modifications significantly improved on the accuracy of sparse reconstruction in the heavy-tailed noise comparing to the state-of-the-art method, LIHT. Moreover, the robust CS algorithms were formulated for general weight functions.

6. REFERENCES

[1] T. Blumensath and M. E. Davies. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265-274, Nov. 2009.

[2] E. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. on Inf. Theory*, 52(2):489-509, Feb. 2006.

[3] E.J. Candes and M.B. Wakin. An introduction to compressive sampling. *IEEE Signal Proc. Mag.*, 25(2):21-30, March 2008.

[4] R. E. Carrillo, K. E. Barner, and Aysal T. C. Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise. *IEEE Journal of Selected Topics in Signal Proc.*, 4(2):392–408, Apr. 2010.

[5] R. E. Carrillo and Barner K. E. Lorentzian based iterative hard thresholding for compressed sensing. In *ICASSP 2011*, pages 3664–3667, Prague, Czech Rep.

[6] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *Information Theory, IEEE Transactions on*, 55(5):2230 –2249, May 2009.

[7] M. A. T. Figueiredo et.al. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Proc.*, 1(4):586 –597, Dec. 2007.

[8] A. E. Hoerl and R. W. Kennard. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, February 1970.

[9] M. E. Kilmer and D. P. O'Leary. Choosing regularization parameters in iterative methods for ill-posed problems. *SIAM. J. Matrix Anal. and Appl.*, 2001.

[10] R. A. Maronna, R. D. Martin, and V. J. Yohai. *Robust Statistics: Theory and Methods*. John Wiley, 2006.

[11] I. Mizera and C. H. Muller. Breakdown points of Cauchy regression-scale estimators. *Statistics and Probability Letters*, 57(1):79-89, Mar. 2002.

[12] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, May 2009.

[13] J. L. Paredes and G. R. Arce. Compressive sensing signal reconstruction by weighted median regression estimates. *IEEE Trans. on Signal Processing*, 59(6):2585–2601, June 2011.

[14] M. J. Silvapulle. Robust ridge regression based on an M-estimator. *Australian Journal of Statistics*, 33(3):319333, Sep. 1991.

[15] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. on Information Theory*, 53(12):4655–4666, Dec. 2007.

[16] J. A. Tropp and S.J. Wright. Computational methods for sparse solution of linear inverse problems. *Proceedings of the IEEE*, 98(6):948 –958, June 2010.