ROBUST SPARSE RECOVERY FOR COMPRESSIVE SENSING IN IMPULSIVE NOISE USING ℓ_P-NORM MODEL FITTING

Fei Wen[1], *Peilin Liu*[1], *Yipeng Liu*[2], *Robert C. Qiu*[1], *Wenxian Yu*[1]

[1]Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China
[2]School of Electronic Engineering, University of Electronic Science and Technology of China, China
{wenfei, liupeilin, wxyu}@sjtu.edu.cn, yipengliu@uestc.edu.cn, rcqiu@sjtu.edu.cn

ABSTRACT

This work considers the robust sparse recovery problem in compressive sensing (CS) in the presence of impulsive measurement noise. We propose a robust formulation for sparse recovery using the generalized ℓ_p-norm with $0 < p < 2$ as the metric for the residual error under ℓ_1-norm regularization. An alternative direction method (ADM) has been proposed to solve this formulation efficiently. Moreover, a smoothing strategy has been used to derive a convergent method for the nonconvex case of $p < 1$. The convergence conditions of the proposed algorithm for both the convex and nonconvex cases have been provided . Numerical simulations demonstrated that the new algorithm can achieve state-of-the-art robust performance in highly impulsive noise.

***Index Terms*—** Compressive sensing, robust sparse recovery, alternating direction method, ℓ_p-norm data-fitting

1. INTRODUCTION

Compressive sensing (CS) allows us to acquire sparse signals at a significantly lower rate than the classical Nyquist sampling [1]. The CS theory states that if a signal $\mathbf{x} \in \mathbb{R}^n$ is sparse, only a small number of linear measurements $\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{R}^m$ $(m < n)$ of the signal suffice to accurately reconstruct it, $\mathbf{A} \in \mathbb{R}^{m\times n}$ is the sensing matrix. Taking the inevitable measurement noise into consideration, the compressed measurements can be modeled as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$$

where $\mathbf{n} \in \mathbb{R}^m$ is additive measurement noise.

In the CS setting, the recovery of $\mathbf{x}$ from $\mathbf{y}$ is generally ill-posed since $m < n$. However, provided that $\mathbf{x}$ is sparse and $\mathbf{A}$ satisfies some stable embedding conditions [2], $\mathbf{x}$ can be reliably recovered with an error upper bounded by the noise strength. To reconstruct $\mathbf{x}$, the BPDN and LASSO formulations [3], [4] are of the most popular, e.g., LASSO

$$\min_{\mathbf{x}} \left\{ \frac{1}{\mu} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \|\mathbf{x}\|_1 \right\} \tag{1}$$

where $\mu > 0$ is a regularization parameter.

As in BPDN, LASSO and many other variants, the ℓ_2-norm data-fitting model, which is optimal for Gaussian noise in the maximum likelihood sense, is the most widely used one. However, in practical applications, the measurement noise may be of different kinds or combinations. Impulsive noise in measurements may come from missing data in the measurement process, transmission problems [5]–[7], faulty memory locations [8], buffer overflow [9], and has been raised in many image and video processing works [10]–[13]. In these cases, the ℓ_2-norm model is inefficient as it is highly sensitive to outliers in the observations.

Recently, various robust formulations have been proposed for CS to suppress the outliers in measurements. In [14], the Lorentzian-norm has been employed as the metric for the residual error. In [15], the ℓ_1-norm has been used as the data-fitting model to obtain a robust formulation as

$$\min_{\mathbf{x}} \left\{ \frac{1}{\mu} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_1 + \|\mathbf{x}\|_1 \right\}. \tag{2}$$

It has been shown in [15] that, when the measurements contain impulsive noise, the ℓ_1-loss can result in dramatically better performance compared with the ℓ_2-one. Subsequently, the Huber penalty function has been used to design robust formulation for sparse recovery in [16].

In this work, we use the generalized ℓ_p-norm, $0 < p < 2$, as the loss function for the residual error to propose the following robust formulation

$$\min_{\mathbf{x}} \left\{ \frac{1}{\mu} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_p^p + \|\mathbf{x}\|_1 \right\}. \tag{3}$$

When $0 < p < 1$, $\|\cdot\|_p^p$ is the ℓ_p quasi-norm defined in a similar manner as the case of $p \geq 1$, i.e., $\|\mathbf{v}\|_p^p = \sum_{i=1}^m |v_i|^p$. The intuition behind utilizing ℓ_p-norm loss function is that, compared with the quadratic function, it is a less rapidly increasing function when $p < 2$, and, accordingly, is less sensitive to large outliers, especially when p is small.

Except for the special case of $p = 1$, the problem (3) has still not been well addressed. When $1 < p < 2$, it can be solved by traditional convex optimization methods such as

This work was supported in part by the National Natural Science Foundation of China (NSFC) under grants 61401501, 61472442 and 61171171.

interior-point methods. More specifically, this problem can be converted into

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_p \leq \epsilon \tag{4}$$

where $\epsilon > 0$ bounds the ℓ_p-norm of the residual error. A semi-definite program conversion method has been proposed in [17] to handle the problem (4). However, this approach is generally inefficient and impractical for large-scale problems. Moreover, when $1 < p < 2$, the ℓ_p-norm is smooth and convex but its gradient is not Lipschitz continuous, thus traditional proximal gradient methods cannot be directly applied.

When $0 < p < 1$, the problem (3) is more difficult to solve since in addition to the nonconvexity of the loss term, both the loss and regularization terms are nonsmooth. This case has still not been reported in the open literatures. The main contributions of this work are as follows.

First, we show that, with an appropriate choice of p, the proposed formulation can stably recover the desired signal with a finite recovery error even when the noise is highly impulsive with infinite variance. Second, we propose an efficient alternating direction method (ADM), termed Lp-ADM, for the optimization problem (3). Furthermore, for the nonconvex case of $p < 1$, a smoothing strategy has been employed to derive a convergent algorithm. Third, the convergence conditions of the new algorithm have been analyzed for both the convex and nonconvex cases. Finally, experimental results demonstrated that, with an appropriate choice of p, e.g., $p < 1$, the new algorithm has the capability to achieve the state-of-the-art robust performance in highly impulsive noise.

2. PRELIMINARIES

2.1. Analysis on ℓ_p-norm Data-Fitting Model

For each integer $s = 1, 2, \cdots$, let δ_s denote the s-restricted isometry constant of $\mathbf{A}$. It has been shown in [18] that if $\|\mathbf{n}\| \leq \epsilon_2$ and $\delta_{2s} < \sqrt{2} - 1$, the solution to the BPDN problem, denoted by $\hat{\mathbf{x}}$, obeys

$$\|\hat{\mathbf{x}} - \mathbf{x}\| \leq C\epsilon_2 \tag{5}$$

where C is a constant depends on δ_{2s}. For the proposed formulation (3), we have the following result (the proof will be presented in a later work).

Theorem 1. Suppose that $\mathbf{A}$ satisfies the restricted isometry property (RIP) of order $2s$ with $\delta_{2s} < \sqrt{2} - 1$. Then for any signal $\mathbf{x}$ supported on T_0 with $|T_0| \leq s$, and any measurement noise $\mathbf{n}$ with $\|\mathbf{n}\|_p \leq \epsilon_p$, $0 < p < 2$, the solution to (4), denoted by $\hat{\mathbf{x}}$, obeys

$$\|\hat{\mathbf{x}} - \mathbf{x}\| \leq C_s\epsilon_p \tag{6}$$

where C_s is a constant depends on δ_{2s}.

In Theorem 1, the condition of the noise is relaxed to $\|\mathbf{n}\|_p \leq \epsilon_p$, $0 < p < 2$, while that of BPDN is $\|\mathbf{n}\|_2 \leq \epsilon_2$. This result implies that, when the noise is highly impulsive with infinite variance, the proposed formulation and (4) has the capability to stably (in statistics) recover $\mathbf{x}$ while BPDN (also LASSO) cannot.

2.2. Proximity Operator for ℓ_p-Norm Functions

This subsection introduces the proximity operator which will be used in the proposed algorithm. Recall the proximity operator of a function $g(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^m$ with penalty η

$$\text{prox}_{g,\eta}(\mathbf{t}) = \arg\min_{\mathbf{x}} \left\{ g(\mathbf{x}) + \frac{\eta}{2}\|\mathbf{x} - \mathbf{t}\|^2 \right\}. \tag{7}$$

For $g(\mathbf{x}) = a\|\mathbf{x}\|_p^p$ with $p > 0$ and $a > 0$, the computation of $\text{prox}_{g,\eta}$ reduces to solving m univariate problems.

Case 1: $0 < p < 1$. The proximity operator in this case can be computed as [19]

$$\text{prox}_{g,\eta}(\mathbf{t})_i = \begin{cases} 0, & |t_i| < \tau \\ \{0, \text{sign}(t_i)\beta\}, & |t_i| = \tau \\ \text{sign}(t_i)z_i, & |t_i| > \tau \end{cases} \tag{8}$$

for $i = 1, \cdots, m$, where $\beta = [2a(1-p)/\eta]^{\frac{1}{2-p}}$, $\tau = \beta + ap\beta^{p-1}/\eta$, z_i is the solution of $h_1(z) = apz^{p-1} + \eta z - \eta|t_i| = 0$ over the region $(\beta, |t_i|)$. Since $h_1(z)$ is convex, when $|t_i| > \tau$, z_i can be efficiently solved, e.g., by a Newton's method.

Case 2: $p = 1$. In this case, the proximity operator reduces to the well-known soft-thresholding operator

$$\text{prox}_{g,\eta}(\mathbf{t})_i = S_{a/\eta}(\mathbf{t})_i = \text{sign}(t_i)\max\{|t_i| - a/\eta, 0\}.$$

Case 3: $1 < p < 2$. In this case, $g(\mathbf{x})$ is convex and smooth, and the proximity operator satisfies

$$\text{prox}_{g,\eta}(\mathbf{t})_i = \text{sign}(t_i)z_i \tag{9}$$

where z_i is the solution of the equality

$$h_2(z) = paz^{p-1} + \eta z - \eta|t_i| = 0, \quad z \geq 0. \tag{10}$$

Note that, $h_2(z)$ is an increasing and concave function for $z \geq 0$, with $h_2(0) < 0$ and $h_2(|t_i|) > 0$ when $t_i \neq 0$. Thus, when $t_i \neq 0$, the solution of (10) satisfies $0 < z_i < |t_i|$ and can be computed efficiently, e.g., by a Newton's method.

3. PROPOSED ALGORITHM

3.1. Lp-ADM Algorithm without Smoothing

In the ADM framework, the ℓ_p-norm loss term and the nonsmooth ℓ_1-regularization term are naturally separated. Specifically, using an auxiliary variable $\mathbf{v} \in \mathbb{R}^m$, the problem (3) can be equivalently reformulated as

$$\min_{\mathbf{x},\mathbf{v}} \left\{ \frac{1}{\mu}\|\mathbf{v}\|_p^p + \|\mathbf{x}\|_1 \right\} \quad \text{subject to} \quad \mathbf{A}\mathbf{x} - \mathbf{y} = \mathbf{v}. \tag{11}$$

The corresponding augmented Lagrangian function is

$$\mathcal{L}(\mathbf{v},\mathbf{x},\mathbf{w}) = \frac{1}{\mu}\|\mathbf{v}\|_p^p + \|\mathbf{x}\|_1 - \mathbf{w}^T(\mathbf{A}\mathbf{x}-\mathbf{y}-\mathbf{v}) + \frac{\rho}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}-\mathbf{v}\|^2$$

where $\mathbf{w} \in \mathbb{R}^m$ is the dual variable, $\rho > 0$ is a penalty parameter. Then, ADM applied to (11) consists of the following iterations

$$\mathbf{v}^{k+1} = \arg\min_{\mathbf{v}} \left(\frac{1}{\mu}\|\mathbf{v}\|_p^p + \frac{\rho}{2}\|\mathbf{A}\mathbf{x}^k - \mathbf{y} - \mathbf{v} - \frac{\mathbf{w}^k}{\rho}\|^2\right) \quad (12)$$

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \left(\|\mathbf{x}\|_1 + \frac{\rho}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y} - \mathbf{v}^{k+1} - \frac{\mathbf{w}^k}{\rho}\|^2\right) \quad (13)$$

$$\mathbf{w}^{k+1} = \mathbf{w}^k - \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{y} - \mathbf{v}^{k+1}). \quad (14)$$

The $\mathbf{x}$-subproblem (13) itself is an ℓ_2-ℓ_1 minimization problem as (1). We can approximately solve this subproblem by linearizing the quadratic term of its objective. Specifically, at a given point $\mathbf{x}^k$ we have

$$\frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{u}^k\|^2 \approx \frac{1}{2}\|\mathbf{A}\mathbf{x}^k - \mathbf{u}^k\|^2 + (\mathbf{x}-\mathbf{x}^k)^T d_1(\mathbf{x}^k) + \frac{L_1}{2}\|\mathbf{x}-\mathbf{x}^k\|^2$$

where $d_1(\mathbf{x}^k) = \mathbf{A}^T(\mathbf{A}\mathbf{x}^k - \mathbf{u}^k)$, $\mathbf{u}^k = \mathbf{y} + \mathbf{v}^{k+1} + \mathbf{w}^k/\rho$, $L_1 > 0$ is a proximal parameter. With this linearization, the $\mathbf{x}$-subproblem degenerates to the soft-thresholding operator

$$\mathbf{x}^{k+1} = S_{1/(\rho L_1)}\left(\mathbf{x}^k - \frac{1}{L_1}\mathbf{A}^T(\mathbf{A}\mathbf{x}^k - \mathbf{u}^k)\right). \quad (15)$$

The $\mathbf{v}$-subproblem (12) is a form of the proximity operator (7), which can be efficiently solved as

$$\mathbf{v}^{k+1} = \mathrm{prox}_{\frac{1}{\mu}\|\mathbf{v}\|_p^p,\rho}(\mathbf{b}^k) = \begin{cases} \text{solved as (8)}, & 0<p<1 \\ S_{1/(\mu\rho)}(\mathbf{b}^k), & p=1 \\ \text{solved as (9)}, & 1<p<2 \\ \rho\mathbf{b}^k/(\rho+2/\mu), & p=2 \end{cases}$$

where $\mathbf{b}^k = \mathbf{A}\mathbf{x}^k - \mathbf{y} - \mathbf{w}^k/\rho$.

3.2. Lp-ADM Algorithm Using Smoothed ℓ_1-Regularization For the Nonconvex Case

In the nonconvex case with $p < 1$, the above algorithm is not guaranteed to converge. To address this problem, we propose to solve a smoothed version of the problem (3) when $p < 1$. Specifically, the ℓ_1-norm regularization in (3) is smoothed as

$$\|\mathbf{x}\|_{1,\varepsilon} = \sum\nolimits_i (x_i^2 + \varepsilon^2)^{\frac{1}{2}}.$$

$\varepsilon > 0$ is an approximation parameter and we have

$$\lim_{\varepsilon\to 0}\|\mathbf{x}\|_{1,\varepsilon} = \|\mathbf{x}\|_1$$

which means that with a small ε, $\|\mathbf{x}\|_{1,\varepsilon}$ accurately approximates the ℓ_1-norm of $\mathbf{x}$. More importantly, with $\varepsilon > 0$, $\|\mathbf{x}\|_{1,\varepsilon}$ is strictly convex and its gradient is Lipschitz continuous. In this case, the derived algorithm is guaranteed to converge if the penalty parameter is chosen sufficiently large (see Theorem 3).

Using $\|\mathbf{x}\|_{1,\varepsilon}$ as the regularization, the problem becomes

$$\min_{\mathbf{x},\mathbf{v}} \left\{\frac{1}{\mu}\|\mathbf{v}\|_p^p + \|\mathbf{x}\|_{1,\varepsilon}\right\} \quad \text{subject to} \quad \mathbf{A}\mathbf{x} - \mathbf{y} = \mathbf{v}. \quad (16)$$

Accordingly, the $\mathbf{x}$-subproblem becomes

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \left(\|\mathbf{x}\|_{1,\varepsilon} + \frac{\rho}{2}\|\mathbf{A}\mathbf{x} - \mathbf{u}^k\|^2\right). \quad (17)$$

We linearize the term $\|\mathbf{x}\|_{1,\varepsilon}$ in the objective of (17) at $\mathbf{x}^k$ as

$$\|\mathbf{x}\|_{1,\varepsilon} \approx \|\mathbf{x}^k\|_{1,\varepsilon} + (\mathbf{x}-\mathbf{x}^k)^T d_2(\mathbf{x}^k) + \frac{L_2}{2}\|\mathbf{x}-\mathbf{x}^k\|^2$$

which results in the following closed-form solution of (16)

$$\mathbf{x}^{k+1} = (L_2\mathbf{I}_n + \rho\mathbf{A}^T\mathbf{A})^{-1}[L_2\mathbf{x}^k - d_2(\mathbf{x}^k) + \rho\mathbf{A}^T\mathbf{u}^k] \quad (18)$$

where $d_2(\mathbf{x}^k) = \nabla\|\mathbf{x}^k\|_{1,\varepsilon}$ with $d_2(\mathbf{x}^k)_i = x_i(x_i^2+\varepsilon^2)^{-\frac{1}{2}}$, $L_2 > 0$ is a proximal parameter. Note that, we do not linearize the quadratic term in the objective as the previous case since it does not yield a closed-form solution when $\varepsilon > 0$.

4. CONVERGENCE ANALYSIS

The proof of the following convergence properties of Lp-ADM will be presented in a later work.

First, we give the convergence condition of Lp-ADM for $p \geq 1$ when the $\mathbf{x}$-subproblem is updated via (15).

Theorem 2. For any $\rho > 0$, $p \geq 1$, and arbitrary starting point $(\mathbf{x}^0, \mathbf{w}^0)$, the sequence $\{(\mathbf{v}^k, \mathbf{x}^k, \mathbf{w}^k)\}$ generated by Lp-ADM via (12), (15) and (14) with $L_1 > \lambda_{\max}(\mathbf{A}^T\mathbf{A})$ converges to a solution of (11).

Then, we give a sufficient condition for the convergence of Lp-ADM for the generalized case of $p > 0$ when the $\mathbf{x}$-subproblem is updated via (18).

Theorem 3. Suppose that $\varepsilon > 0$ and $\mathbf{A}\mathbf{A}^T \succeq \mu_A\mathbf{I}_m$ with some $\mu_A > 0$, then, for any $p > 0$ if

$$\rho > \frac{C}{\varepsilon} \quad \text{with} \quad C = \frac{4(2\alpha^2 + 2\alpha + 1)}{\mu_A(2\alpha - 1)}, \quad (19)$$

the sequence $\{(\mathbf{v}^k, \mathbf{x}^k, \mathbf{w}^k)\}$ generated by the ADM algorithm via (12), (18) and (14) with $L_2 = \frac{\alpha}{\varepsilon} > \frac{1}{2\varepsilon}$ (i.e., $\alpha > \frac{1}{2}$) converges to a stationary point of the problem (16).

When $\varepsilon \to 0$, the problem (16) reduces to the problem (11) and the approximate error vanishes. However, in this case the sufficient condition (19) requires $\rho \to \infty$. In general, an ADM tends to be very slow when the penalty parameter ρ gets very large. Thus, a tradeoff should be made between the approximating accuracy and the algorithm convergent rate.

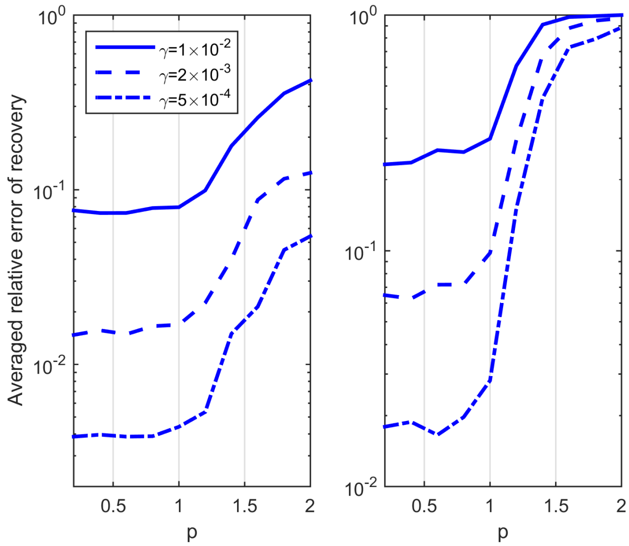

Fig. 1. Recovery performance of Lp-ADM versus p in $S\alpha S$ noise. *Left:* $\alpha = 1$ (Cauchy noise). *Right:* $\alpha = 0.5$.

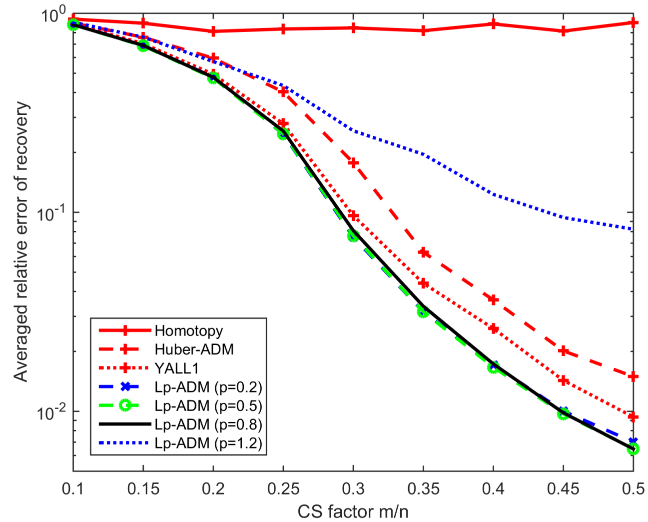

Fig. 2. Recovery performance versus CS factor m/n for the compared algorithms in $S\alpha S$ noise with $\alpha = 0.5$ and $\gamma = 5 \times 10^{-4}$.

5. NUMERICAL EXPERIMENTS

In this section, we evaluate the performance of the new method via numerical simulations. It is compared with a standard reconstruction algorithm, Homotopy [20], and two robust algorithms, Huber-FISTA [16], and YALL1 [15]. We use a simulated K-sparse (with $K = 30$) signal of length $n = 512$. The positions of the K nonzeros are uniformly randomly chosen while the amplitude of each nonzero entry is generated according to the Gaussian distribution. An $m \times n$ orthonormal Gaussian random matrix is used as the sensing matrix. Each provided experimental result is an average over 200 independent runs. When $p \geq 1$, Lp-ADM is run with $\rho = 10^2$ and the $\mathbf{x}$-subproblem is updated via (15) with $L_1 = 2$. When $p < 1$, Lp-ADM is run with $\rho = 2 \times 10^4$ and the $\mathbf{x}$-subproblem is updated via (18) with $\varepsilon = 10^{-3}$ and $L_2 = \frac{1}{\varepsilon}$. In implementing Lp-ADM in the nonconvex case, we firstly run it with $p = 1$ to obtain a starting point.

Fig. 1 shows the recovery performance of Lp-ADM versus p in symmetric α-stable ($S\alpha S$) noise. We set $m = 200$. Two impulsive conditions, with characteristic exponents $\alpha = 1$ (Cauchy noise) and $\alpha = 0.5$, and three noise levels, with scale parameters of $\gamma \in \{10^{-2}, 2 \times 10^{-3}, 5 \times 10^{-4}\}$, are considered. Since the variance of such $S\alpha S$ noise is infinite, the ℓ_2-norm loss formulation is unstable in statistics in this case. Accordingly, as shown in Fig. 1, the performance corresponds to $p = 2$ deteriorates drastically when the noise gets more impulsive. Meanwhile, using a smaller value of p can yield significantly better performance.

In the second experiment, we compare the proposed algorithm with the Homotopy, Huber-FISTA, and YALL1 algorithms. Fig. 2 shows the performance of the compared algorithms versus CS factor m/n in $S\alpha S$ noise with $\alpha = 0.5$ and $\gamma = 5 \times 10^{-4}$. We set $K = 30$ and $n = 512$. Four typical values of p, $p \in \{0.2, 0.5, 0.8, 1.2\}$, are examined for the new algorithm.

It can be seen from Fig. 2 that, the robust algorithms distinctly outperform Homotopy when $m/n > 0.1$. As the CS factor increases, the recovery accuracy of each robust algorithm improves significantly, but that of Homotopy does not improve distinctly. This is due to the fact that the considered $S\alpha S$ noise is highly impulsive, and the ℓ_2-loss function is very sensitive to extremely large outliers. Huber-FISTA is more robust than Homotopy but less robust than YALL1. That is due the nature that Huber function fitting lies in between the least-squares and least-absolute-deviations. When $m/n > 0.25$, Lp-ADM with $p < 1$ achieves better performance than Huber-FISTA and YALL1.

6. CONCLUSION

This work introduced a robust formulation for sparse recovery, which employs the ℓ_p-norm with $0 < p < 2$ as the loss function. An ADM algorithm has been proposed to solve this formulation efficiently. Moreover, we have provided the convergence conditions of the new algorithm for both the convex and nonconvex cases. Simulation results demonstrated that, in highly impulsive noise, the new algorithm with an appropriate choice of p ($p < 1$) has the capability to achieve distinctly better recovery accuracy compared with existing robust algorithms.

7. REFERENCES

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

[2] E.J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[3] 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.

[4] R. Tibshirani, “Regression shrinkage and selection via the LASSO,” *J. R. Statist. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[5] E.J. Candès and T. Tao, “Decoding by linear programming,” *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[6] E.J. Candès and P.A. Randall, “Highly robust error correction by convex programming,” *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 2829–2840, 2008.

[7] B. Popilka, S. Setzer, and G. Steidl, “Signal recovery from incomplete measurements in the presence of outliers,” *Inverse Problems Imag.*, vol. 1, no. 4, pp. 661–672, Nov. 2007.

[8] R. Chan, C.W. Ho, and M. Nikolova, “Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization,” *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1479–1485, Oct. 2005.

[9] T. Hashimoto, “Bounds on a probability for the heavy tailed distribution and the probability of deficient decoding in sequential decoding,” *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 990–1002, Mar. 2005.

[10] L. Bar, A. Brook, N. Sochen, and N. Kiryati, “Deblurring of color images corrupted by impulsive noise,” *IEEE Trans. Image Process.*, vol. 16, no. 4, pp. 1101–1111, Apr. 2007.

[11] P. Civicioglu, “Using uncorrupted neighborhoods of the pixels for impulsive noise suppression with ANFIS,” *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 759–773, Mar. 2007.

[12] P. Windyga, “Fast impulsive noise removal,” *IEEE Trans. Image Process.*, vol. 10, no. 1, pp. 173–179, Jan. 2001.

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

[14] R.E. Carrillo, K.E. 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.*, vol. 4, no. 2, pp. 392–408, Apr. 2010.

[15] J.F. Yang and Y. Zhang, “Alternating direction algorithms for ℓ_1-problems in compressive sensing,” *SIAM J. Sci. Comput.*, vol. 33, pp. 250–278, 2011.

[16] D.S. Pham, and S. Venkatesh, “Efficient algorithms for robust recovery of images from compressed data,” *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4724–4737, 2013.

[17] F. Alizadeh and D. Goldfarb, “Second-order cone programming,” *Mathematical programming*, vol. 95, no. 1, pp. 3–51, 2003.

[18] E.J. Candès, “The restricted isometry property and its implications for compressed sensing,” *Comptes. Rendus. Mathematique*, vol. 346, no. 9, pp. 589–592, 2008.

[19] G. Marjanovic and V. Solo, “On ℓ_q optimization and matrix completion,” *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5714-5724, 2012.

[20] Y. Tsaig and D.L. Donoho, “Fast solution of ℓ_1-norm minimization problems when the solution may be sparse,” *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4789–4812, Nov. 2008.