Negentropy-Based Sparsity-Promoting Reconstruction with Fast Iterative Solution from Noisy Measurements

Yingxin Zhao [1,2] ![ID], **Yingjie Huang** [1,2] ![ID], **Hong Wu** [1,2,*], **Ming Zhang** [1,2], **Zhiyang Liu** [1,2] ![ID] **and Shuxue Ding** [1,2]

[1] College of Electronic Information and Optical Engineering, Nankai University, Tianjin 300350, China; zhaoyx@nankai.edu.cn (Y.Z.); yjhuang97@mail.nankai.edu.cn (Y.H.); 2120170366@mail.nankai.edu.cn (M.Z.); liuzhiyang@nankai.edu.cn (Z.L.); sding@guet.edu.cn (S.D.)

[2] Tianjin Key Laboratory of Optoelectronic Sensor and Sensing Network Technology, Tianjin 300350, China

* Correspondence: wuhong@nankai.edu.cn

check for updates

Abstract: Compressed sensing provides an elegant framework for recovering sparse signals from compressed measurements. This paper addresses the problem of sparse signal reconstruction from compressed measurements that is more robust to complex, especially non-Gaussian noise, which arises in many applications. For this purpose, we present a method that exploits the maximum negentropy theory to promote the adaptability to noise. This problem is formalized as a constrained minimization problem, where the objective function is the negentropy of measurement error with sparse constraint $\ell_p(0 < p < 1)$-norm. On the minimization issue of the problem, although several promising algorithms have been proposed in the literature, they are very computationally demanding and thus cannot be used in many practical situations. To improve on this, we propose an efficient algorithm based on a fast iterative shrinkage-thresholding algorithm that can converge fast. Both the theoretical analysis and numerical experiments show the better accuracy and convergent rate of the proposed method.

Keywords: compressed sensing; sparse constraint; first order Taylor expansion; fast iterative; non-Gaussian noise

1. Introduction

In recent years, compressed sensing (CS) has attracted considerable attention in areas of computer science, signal processing and wireless communication [1–5]. It can be considered as a signal sampling or coding approach that leads full use of signal sparsity, which may make it possible to surpass the traditional limits of sampling theory. CS suggests that if the signal is sparse or sparse in a transform domain such as the Fourier transform domain or wavelet transform domain, it can be projected onto a low-dimensional space via a measurement matrix that is uncorrelated with the transform basis. The original signal can be represented using only a few non-zero coefficients in a suitable basis or dictionary. Nonlinear optimization can then enable reconstruction of such signals from very few measurements with high probability. This reconstruction process can also be called sparse coding, which is described mathematically as finding an unknown signal α to satisfy the equation

$$\mathbf{x} = \mathbf{D}\alpha + \mathbf{n} \tag{1}$$

where $\mathbf{D} \in \mathbb{R}^{M \times N}(M < N)$ is a measurement matrix that satisfies the restricted isometry property (RIP) criteria [6], $\mathbf{x} \in \mathbb{R}^M$ is a low-dimensional measurement signal, $\mathbf{n} \in \mathbb{R}^M$ is an unknown noise vector and $\alpha \in \mathbb{R}^N$ is a sparse signal or a sparse representation of the signal in a transform basis.

Obviously, this is solving an underdetermined equation, and there are infinite solutions. Given the measurements $\mathbf{x}$ and the knowledge that the original signal $\boldsymbol{\alpha}$ is sparse or compressible, it is natural to attempt to recover $\boldsymbol{\alpha}$ by treating this feature as a constraint [7]. Then, we can search for an optimal one from infinite solutions by solving an optimization problem of the form

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \text{ subject to } \mathbf{x} = \mathbf{D}\boldsymbol{\alpha} + \mathbf{n} \tag{2}$$

where $\|\boldsymbol{\alpha}\|_0$, formally called ℓ_0 regularization, counts the number of non-zero complements of $\boldsymbol{\alpha}$ and is used to promote the sparsity of the solution [8]. However, if we want to solve the above problem directly, we need to list the combinations of all non-zero element positions separately. This is an NP-hard problem, which is computation prohibited [9]. In order to overcome the difficulty, many researchers are engaging in developing simple and effective algorithms to solve the ℓ_0 regularization problem.

A typical method is using a greedy algorithm such as orthogonal matching pursuit (OMP) [10], compressive sampling matching pursuit (CoSaMP) [11], regularized orthogonal matching pursuit (ROMP) [12] and so on. The greedy strategies rely on iterative approximation of the signal coefficients and support, which are very fast for low-dimensional problems. Another approach is to relax ℓ_0 regularization and, instead, to consider the ℓ_1 regularization:

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \text{ subject to } \mathbf{x} = \mathbf{D}\boldsymbol{\alpha} + \mathbf{n} \tag{3}$$

where $\|\boldsymbol{\alpha}\|_1$ is ℓ_1 norm counting the sum of the absolute values of non-zero elements in vector $\boldsymbol{\alpha}$. Thus, the ℓ_0 regularization problem is replaced by a more tractable formulation. That is, $\|\cdot\|_0$ is replaced with its convex approximation $\|\cdot\|_1$, whose minimization can be performed efficiently using conventional convex optimization methods. In fact, the resulting problem can be posed as the linear program. For problems (2) and (3), we can search unknown variable $\boldsymbol{\alpha}$ using the method of gradient descent, which is called the iterative shrinkage-thresholding algorithm (ISTA) [13]. In addition, the basis pursuit (BP) algorithm [14] is also used for solving such problems.

There are a variety of algorithms that have been used for solving optimization problems in applications such as digital signal process and wireless communication. In these fields, noise inevitably affects the performance of the algorithm. The ability to perfectly reconstruct a sparse signal from noise-free measurements represents a very promising result. However, in most real-world systems the measurements are likely to be contaminated by some form of noise. Moreover, systems which are implemented in physical hard-ware will be subject to a variety of different types of noise depending on the setting [15].

In most previous studies, the constraint in (2) or (3) is formulated as a penalty function with the mean square error (MSE),

$$\min_{\boldsymbol{\alpha}} \left\{ f(\boldsymbol{\alpha}) = \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 \right\} \tag{4}$$

so that the problem becomes an unconstrained optimization problem and can be solved using a convex optimization algorithm. Here, MSE is adopted since it is optimal if the noise, $\mathbf{n}$, is Gaussian distributed, which is indeed a suitable assumption for many applications.

However, there are many other applications in which the noise may not be Gaussian distributed. In these cases, if we still use the MSE measurement, it results in performance degrading and thus becomes not applicable in the targeted applications. In fact, non-Gaussian noise is quite common, such as the salt and pepper noise in image [16] and impulse noise in underwater acoustic channels [17]. Therefore, it is necessary to develop new methods and algorithms that are applicable to tasks with non-Gaussian noise.

1.1. Related Works

In order to achieve robust sparse recovery in the presence of non-Gaussian noise, various robust formulations have been proposed in recent years. The Lorentzian norm was employed as metrics

for the residual error in [18]. In [19,20], optimization-based approaches using the Huber function as the objective function were proposed. Moreover, to tackle the ℓ_1-regularized least-absolute (ℓ_1-LA) minimization problem, many effective algorithms have been proposed such as YALL1 [21], which solves the ℓ_1-LA formulation based on the alternating direction method of multipliers (ADMM) framework. Though ℓ_1-norm regularization has been widely used, $\ell_p(0 < p < 1)$-norm regularization has shown significantly better performance in sparse recovery. To achieve more accurate recovery, research focusing on the ℓ_p-LA minimization problem was reported in [22]. Meanwhile, it was demonstrated in [23] that the negentropy of measurement error can achieve good recovery performance with non-Gaussian noise. Whilst the focus of [23] is to present a novel method of error measurement, it also suggests an optimized algorithm that combines forward-backward splitting to solve the robust CS formulation. However, the efficiency and capacity of the method proposed in [23] have not reached the requirement of real-world applications that require fast processing, such as magnetic resonance imaging (MRI), thus, we propose in this paper more computationally efficient algorithms to achieve the goal.

1.2. Contributions

In this paper, we mainly study sparse signal reconstruction algorithms against non-Gaussian noise. A novel sparse representation model is proposed, which takes the negentropy [23] of measurement error as the objective function and $\ell_p(0 < p < 1)$ norm as the sparsity measurement. The negentropy can be exploited to measure the non-Gaussian properties of random variables, here the fitting error, and the stronger the non-Gaussian noise is, the larger the negentropy value is. Thus, this model is appropriate for reconstructing sparse signals from non-Gaussian noisy measurements, and α, which maximizes the negentropy of the measurement error, is the optimal solution if the constraint is satisfied. In summary, the contribution of this paper is as follows:

(a)   We propose a new sparse representation model for accurately recovering the original signal from noisy measurements, especially if it is non-Gaussian. The negentropy of measurement error emerges as the loss function with optimum $\ell_p(0 < p < 1)$ sparse constraint.

(b)   We develop a solution for the proposed model. The first order Taylor formula is used to approximate the value in the neighborhood at any point on the objective function, and then, we find the minimum point of the function in the neighborhood.

(c)   We develop an optimization method of the fast iterative shrinkage-thresholding algorithm (FISTA) to speed up the convergence of the algorithm.

(d)   We evaluated the new algorithm using sparse signal recovery, image denoising and recovery of MRI images. The results showed that the new algorithm is more accurate and faster than several existing methods and has better adaptability to different types of noise.

1.3. Organization

The rest of this paper is organized as follows. In Section 2, we survey the iterative shrinkage-thresholding algorithm (ISTA), since this technique will be used for the new model proposed in this paper. Then, we propose the algorithm model based on negentropy maximization and $\ell_p(0 < p < 1)$ norm and demonstrate the solution process of the model and the principle of accelerating convergence in detail. In Section 3, we show the experiments with a series of recovering sparse signals in different conditions to analyze the accuracy and stability of the developed algorithm. Furthermore, we show the performance of the algorithm in image denoising and MRI images recovery to indicate the performance in applications. Finally, Section 4 concludes the paper.

2. Methods

The problem that needs to be solved in this paper is to recover the sparse signal with high precision from the linearly compressed measurements with different types of noise, which is formulated by (1).

Here, $\boldsymbol{\alpha} \in \mathbb{R}^N$ is an unknown sparse signal in which the number k of non-zero element is less than N. $\mathbf{D} \in \mathbb{R}^{M \times N}$ is a known measurement matrix with $M < N$, where the ratio of M to N is called the compression ratio. The measurement signal $\mathbf{x} \in \mathbb{R}^N$, as a linear combination of column vectors in the $\mathbf{D}$, is also known. The vector $\mathbf{n}$ may be Gaussian or non-Gaussian noise. In order to find the optimal solution, it is very necessary to establish a suitable mathematical model, which will play an important role in guaranteeing the accuracy of the solution.

2.1. ISTA Algorithm

ISTA is a typical algorithm used for solving problem (3); the detailed steps are shown in Algorithm 1, which can be frequently transformed into the following problem:

$$\min_{\boldsymbol{\alpha}}\left\{f(\boldsymbol{\alpha}) = \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1\right\} \tag{5}$$

where $\|\boldsymbol{\alpha}\|_1 = \sum_i^n |\alpha_i|$ is a sparse constrain term and N is the count of elements of signal $\boldsymbol{\alpha}$. $\lambda > 0$ is a regularization parameter to adjust the effect of the sparse constraint. ISTA gradually approaches the global minimum after multiple iterations by finding the local minimum, which approaches the optimal solution according to following recursive equation [24]:

$$\boldsymbol{\alpha}_k = \pi_{\lambda\eta}(\boldsymbol{\alpha}_{k-1} - \eta\nabla f(\boldsymbol{\alpha}_{k-1})) \tag{6}$$

where $\eta \in \left(0, 1/\|\mathbf{D}^T\mathbf{D}\|\right)$ is a suitable step and $\pi_{\lambda\eta}$ is a nonlinear soft-thresholding operator, which is defined in Equation (7).

$$\pi_{\lambda\eta}(\boldsymbol{\alpha})_i = \text{sign}(\alpha_i)\max\left\{\left(|\alpha_i| - \lambda\eta\right), 0\right\} \tag{7}$$

ISTA has been developed for a long time and is used in applications, but its application range is limited due to its poor performance against non-Gaussian noise. At the same time, the effect of ℓ_1 norm as sparse constraint needs to be improved. Therefore, we aim, in this paper, to improve the robustness and accuracy of the algorithm by further changing the solution model.

Algorithm 1 ISTA

1: Input: measurement matrix $\mathbf{D}$, measurement signal $\mathbf{x}$.
2: Initialization: $\eta \in \left(0, 1/\|\mathbf{D}^T\mathbf{D}\|\right)$, $\lambda > 0$, $\boldsymbol{\alpha}_0 = 0$, $k = 0$
3: Main iteration: increment k by 1
4: $\boldsymbol{\alpha}_k = \pi_{\lambda\eta}(\boldsymbol{\alpha}_{k-1} - \eta\nabla f(\boldsymbol{\alpha}_{k-1}))$
5: Stopping rule: stop if $\boldsymbol{\alpha}_k$ has converged
6: Output: sparse signal $\boldsymbol{\alpha}$

2.2. Negentropy-Based and Sparsity-Promoting Algorithm

We propose to exploit the negentropy of measurement error, $e = \mathbf{x} - \mathbf{D}\boldsymbol{\alpha}$, as the objective function as a new model. According to information theory, among all random variables with equal variance, the entropy of Gaussian variables is the largest [23]. Obviously, entropy can measure the Gaussian property of a variable, so negentropy can be used to measure its non-Gaussian property; the larger the negentropy is, the stronger the non-Gaussian variable is. Therefore, it is possible to find an optimal $\boldsymbol{\alpha}$ which maximizes the negentropy of the measurement error so that an accurate estimation is obtained from the signal affected by the non-Gaussian noise. The negentropy is defined as follows:

$$N(e_i) = H(e_{i_{gauss}}) - H(e_i), \ for \ i = 1, 2, \cdots, M \tag{8}$$

where $e_{i_{gauss}}$ is a Gaussian distributed variable which has the same variance as e_i and $H(\cdot)$ denotes the differential entropy of the random variable. It is worthy of notice that $N(e_i)$ is positive-definite. It is known from the negentropy definition that when e_i follows the Gaussian distribution, $N(e_i)$ is zero.

As e_i becomes more non-Gaussian, $N(e_i)$ becomes larger. Because it is difficult to solve the differential entropy of random variables, the approximate expression (9) of negentropy is commonly used in practical applications.

$$N(e) = \left\{ E[G(e_i)] - E[G(e_{i_{gauss}})] \right\}^2 \tag{9}$$

where $G(\cdot)$ is a non-linear function, such as $\tanh(c \cdot e)$, $e \cdot \exp(-e^2/2)$ and e^3. $E(\cdot)$ is the expected value operation. Since $E[G(e_i)]$ is less than or equal to $E[G(e_{i_{gauss}})]$, maximizing $N(e)$ can be converted to minimize $E[G(e)]$ and written as

$$\min_{\boldsymbol{\alpha}}\{f(\boldsymbol{\alpha})\} = \sum_{i=1}^{M}\{E[G(e_i)]\} = \sum_{i=1}^{M}\left\{E[G(\mathbf{x} - \mathbf{D}\boldsymbol{\alpha})_i]\right\}, \boldsymbol{\alpha} \in \mathbb{R}^N \tag{10}$$

To solve the above formulation, we find the local minimum to converge the objective function to the global optimal solution through multiple iterations. First, we approximate the function $f(\boldsymbol{\alpha})$ with the first-order Taylor formula near the initial point $\boldsymbol{\alpha}_0$.

$$f(\boldsymbol{\alpha}) = f(\boldsymbol{\alpha}_0) + \langle \boldsymbol{\alpha} - \boldsymbol{\alpha}_0, \nabla f(\boldsymbol{\alpha}_0) \rangle \tag{11}$$

where $f(\boldsymbol{\alpha})$ represents the function value in the neighborhood of point $\boldsymbol{\alpha}_0$ and $\nabla f(\boldsymbol{\alpha}_0)$ is the gradient of function $f(\boldsymbol{\alpha})$ at this point. Then, we solve the minimum value point $\boldsymbol{\alpha}_k$ of the function in this neighborhood and use this point as the starting point of the next iteration. After multiple iterations, $\boldsymbol{\alpha}_k$ will converge to the global optimal solution $\boldsymbol{\alpha}^*$, which approaches the optimal solution according to the following recursive equation:

$$\boldsymbol{\alpha}_k = \operatorname*{argmin}_{\boldsymbol{\alpha}}\{f(\boldsymbol{\alpha}_{k-1}) + \langle \boldsymbol{\alpha} - \boldsymbol{\alpha}_{k-1}, \nabla f(\boldsymbol{\alpha}_{k-1}) \rangle\} \tag{12}$$

A surrogate function based on the proximal point algorithm mentioned in [25,26] can be defined. Then, the objective function $f(\boldsymbol{\alpha})$ is replaced by

$$J(\boldsymbol{\alpha}, \boldsymbol{\alpha}_{k-1}) = f(\boldsymbol{\alpha}_{k-1}) + \langle \boldsymbol{\alpha} - \boldsymbol{\alpha}_{k-1}, \nabla f(\boldsymbol{\alpha}_{k-1}) \rangle + \frac{1}{2\eta}\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_{k-1}\|_2^2 \tag{13}$$

where $\eta \in \left(0, 1/\|\mathbf{D}^T\mathbf{D}\|\right)$ is a restriction parameter. For simplification, the above formula can be written as

$$J(\boldsymbol{\alpha}, \boldsymbol{\alpha}_{k-1}) = \frac{1}{2\eta}\|\boldsymbol{\alpha} - (\boldsymbol{\alpha}_{k-1} - \eta\nabla f(\boldsymbol{\alpha}_{k-1}))\|^2 + K \tag{14}$$

where $K = f(\boldsymbol{\alpha}_{k-1} + \eta/2\nabla f(\boldsymbol{\alpha}_{k-1}))$, a constant term, can be ignored, so Equation (12) can be rewritten as

$$\boldsymbol{\alpha}_k = \operatorname*{argmin}_{\boldsymbol{\alpha}}\{\frac{1}{2\eta}\|\boldsymbol{\alpha} - (\boldsymbol{\alpha}_{k-1} - \eta\nabla f(\boldsymbol{\alpha}_{k-1}))\|^2\} \tag{15}$$

In this paper, $\log[\cosh(c \cdot \mathbf{e})]$ is used as the nonlinear function in the negentropy formula, so the gradient of the function $f(\boldsymbol{\alpha})$ is expressed as $\nabla f(\boldsymbol{\alpha}) = \mathbf{D}^T\tanh(c(\mathbf{D}\boldsymbol{\alpha} - \mathbf{x}))$. Adopting the ℓ_1 regularization as sparse constraint to ensure the sparsity of the solution, $\boldsymbol{\alpha}_k$ can be gained iteratively as follows:

$$\boldsymbol{\alpha}_k = \operatorname*{argmin}_{\boldsymbol{\alpha}}\{\frac{1}{2\eta}\|\boldsymbol{\alpha} - (\boldsymbol{\alpha}_{k-1} - \eta\mathbf{D}^T\tanh(c(\mathbf{D}\boldsymbol{\alpha}_{k-1} - \mathbf{x})))\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1\} \tag{16}$$

Since the ℓ_1 norm is separable, the computation of $\boldsymbol{\alpha}_k$ reduces to solving a one-dimensional minimization problem for each of its components, which can be produced by calculating the following formula:

$$\alpha_i^k = \operatorname{sign}(\beta_i^{k-1})\max(0, |\beta_i^{k-1}| - \lambda\eta) \tag{17}$$

where we use the fact that $\beta_i^{k-1} = \alpha_i^{k-1} - \eta \nabla f(\alpha_i^{k-1})$; the detailed steps based on ℓ_1 norm are shown in Algorithm 2, which has the same sparse constraint and different objective function as ISTA, so that we can further illustrate the advantages of the proposed objective function by comparing the two methods.

Algorithm 2 Negentropy + ℓ_1 Norm

1: Input: measurement matrix $\mathbf{D}$, measurement signal $\mathbf{x}$.
2: Initialization: $\eta \in \left(0, 1/\left\|\mathbf{D}^T\mathbf{D}\right\|\right)$, $\lambda > 0$, $\boldsymbol{\alpha}_0 = 0$, $k = 0$
3: Main iteration: increment k by 1
4: $\beta_i^{k-1} = \alpha_i^{k-1} - \eta \nabla f(\alpha_i^{k-1})$
5: $\alpha_i^k = \text{sign}(\beta_i^{k-1})\max(0, \left|\beta_i^{k-1}\right| - \lambda\eta)$
6: Stopping rule: stop if $\boldsymbol{\alpha}_k$ has converged
7: Output: sparse signal $\boldsymbol{\alpha}$

$\ell_p(0 < p < 1)$ regularization, a nonconvex and non-smooth optimization problem, has been proven to solve more accurate sparse solutions than ℓ_1 regularization as a relaxation approach in [27–29]. Weighted ℓ_1 regularization is an approximate representation of ℓ_p regularization [30,31]. Adopting the weighted ℓ_1 regularization as sparse constraint, the iterative scheme can be written as follows:

$$
\begin{aligned}
\boldsymbol{\alpha}_k &= \underset{\boldsymbol{\alpha}}{\arg\min}\{\frac{1}{2\eta}\|\boldsymbol{\alpha} - (\boldsymbol{\alpha}_{k-1} - \eta\mathbf{D}^T\tanh(c(\mathbf{D}\boldsymbol{\alpha}_{k-1} - \mathbf{x})))\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_p^p\} \\
&= \underset{\boldsymbol{\alpha}}{\arg\min}\underset{i}{\Sigma}\{\frac{1}{2\eta}\|\alpha_i - (\alpha_i^{k-1} - \eta\mathbf{D}^T\tanh(c(\mathbf{D}\boldsymbol{\alpha}_{k-1} - \mathbf{x})))\|_2^2 + \lambda w_i^{k-1}|\alpha_i|\}
\end{aligned}
\tag{18}
$$

where $\|\boldsymbol{\alpha}\|_p^p = \sum_{i=1}^N |\alpha_i|^p \approx \sum_{i=1}^N (|\alpha_i| + \delta)^{p-1}|\alpha_i|$ is an approximate representation of the ℓ_p norm, δ is close to 0 and $w_i^k = \left(\left|\alpha_i^{k-1} + \delta\right|\right)^{p-1}$ is a constant. It is clear from (18) that α_i respectively in each dimension of vector $\boldsymbol{\alpha}_k$ can be solved and updated by calculating the following formula:

$$
\alpha_i^k = \text{sign}(\beta_i^{k-1})\max(0, \left|\beta_i^{k-1}\right| - \lambda\eta w_i^{k-1})
\tag{19}
$$

The detailed steps based on a weighted ℓ_1 norm are shown in Algorithm 3, which improve accuracy of the solution.

Algorithm 3 Negentropy + Weighted ℓ_1 Norm

1: Input: measurement matrix $\mathbf{D}$, measurement signal $\mathbf{x}$.
2: Initialization: choose proper η, λ, c, p and δ, $\boldsymbol{\alpha}_0 = 0$, $k = 0$
3: Main iteration: increment k by 1
4: $\beta_i^{k-1} = \alpha_i^{k-1} - \eta \nabla f(\alpha_i^{k-1})$
5: $\alpha_i^k = \text{sign}(\beta_i^{k-1})\max(0, \left|\beta_i^{k-1}\right| - \lambda\eta w_i^{k-1})$
6: Stopping rule: stop if $\boldsymbol{\alpha}_k$ has converged
7: Output: sparse signal $\boldsymbol{\alpha}$

An exact solution can also be solved by analyzing the trend of the function. For ℓ_p regularization, there will be different sparse constraint effects with different p. When $p \in (0, 0.5]$, there is no significant difference, but when $p \in [0.5, 1)$, the smaller the value of p is, the better the sparse effect is [32,33]. Thus, ℓ_p norm, where $p = 0.5$, is selected as sparse constrain in the proposed algorithm. Adopting $\ell_{0.5}$ regularization as sparse constraint, the iterative scheme can be written as

$$
\boldsymbol{\alpha}_k = \underset{\boldsymbol{\alpha}}{\arg\min}\{\frac{1}{2\eta}\|\boldsymbol{\alpha} - (\boldsymbol{\alpha}_{k-1} - \eta\mathbf{D}^T\tanh(c(\mathbf{D}\boldsymbol{\alpha}_{k-1} - \mathbf{x})))\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_{1/2}^{1/2}\}
\tag{20}
$$

According to the iterative half thresholding algorithm (IHTA) [34], this can be rewritten as

$$\alpha_i^k = \frac{2}{3}x(1 + \cos(\frac{2}{3}\cos^{-1}(-\frac{3^{3/2}}{4}\lambda\eta|x|^{-3/2})))\text{sign}(\max(0, |\beta_i^{k-1}| - \frac{3}{2}\lambda\eta^{\frac{2}{3}})) \tag{21}$$

The detailed steps based on $\ell_{0.5}$ norm are shown in Algorithm 4.

Algorithm 4 Negentropy + $\ell_{0.5}$ Norm

1: Input: measurement matrix **D**, measurement signal **x**.
2: Initialization: choose proper η, λ, c, $\boldsymbol{\alpha}_0 = 0$, $k = 0$
3: Main iteration: increment k by 1
4: $\beta_i^{k-1} = \alpha_i^{k-1} - \eta\nabla f(\alpha_i^{k-1})$
5: $\alpha_i^k = \frac{2}{3}x(1 + \cos(\frac{2}{3}\cos^{-1}(-\frac{3^{3/2}}{4}\lambda\eta|x|^{-3/2})))\text{sign}(\max(0, |\beta_i^{k-1}| - \frac{3}{2}\lambda\eta^{\frac{2}{3}}))$
6: Stopping rule: stop if $\boldsymbol{\alpha}_k$ has converged
7: Output: sparse signal $\boldsymbol{\alpha}$

2.3. Optimizing the Rate of Convergence

The proposed algorithm based on maximizing negentropy and ℓ_p regularization can reconstruct sparse signals more accurately against non-Gaussian noise, which has already been shown in the following experiment. Equally importantly, further optimization of the convergence rate is required since wide application of a CS algorithm, especially in processing large scale data [35], is decided by the optimization process for reconstruction.

In this paper, a fast way to find the optimal solution by optimizing the starting point in each iteration is presented, which is called the fast iterative shrinkage-thresholding algorithm (FISTA). We optimize the convergence rate of the proposed algorithm according to the way of selecting the starting point in FISTA. The detailed steps can be written as follows [24]:

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2} \tag{22}$$

$$\theta_{k+1} = \boldsymbol{\alpha}_k + \left(\frac{t_k - 1}{t_{k+1}}\right)(\boldsymbol{\alpha}_k - \boldsymbol{\alpha}_{k-1}) \tag{23}$$

where $t_1 = 1$, θ_{k+1} depends on the linear combination of $\boldsymbol{\alpha}_k$ and $\boldsymbol{\alpha}_{k-1}$. After one iteration, we optimize the local minimum point as the input for the next iteration. ISTA has a worst-case complexity result of $O(1/k)$, but FISTA has an improved complexity result of $O(1/k^2)$. Our analysis also provides proof as follows.

Let $F(\boldsymbol{\alpha}) = f(\boldsymbol{\alpha}) + g(\boldsymbol{\alpha})$, $Q(\boldsymbol{\alpha}, \boldsymbol{\alpha}_n) = J(\boldsymbol{\alpha}, \boldsymbol{\alpha}_n) + g(\boldsymbol{\alpha})$, $g(\boldsymbol{\alpha})$ is a regularization function.

Lemma 1. *Let $\boldsymbol{\alpha}_n \in \mathbb{R}^N$ and $\eta > 0$ be such that*

$$F(\boldsymbol{\alpha}_{n+1}) \leq Q(\boldsymbol{\alpha}_{n+1}, \boldsymbol{\alpha}_n) \tag{24}$$

Then for any $\boldsymbol{\alpha} \in \mathbb{R}^N$,

$$F(\boldsymbol{\alpha}) - F(\boldsymbol{\alpha}_n) \geq \frac{1}{2\eta}\|\boldsymbol{\alpha}_{n+1} - \boldsymbol{\alpha}_n\|^2 + \frac{1}{\eta}\langle\boldsymbol{\alpha}_n - \boldsymbol{\alpha}, \boldsymbol{\alpha}_{n+1} - \boldsymbol{\alpha}_n\rangle \tag{25}$$

Theorem 1. *Let* $\{\alpha_k\}$ *be the sequence generated by (6). Then for any* $k \geq 1$

$$F(\alpha_k) - F(\alpha^*) \leq \frac{\|\alpha^* - \alpha_0\|^2}{2k\eta} \forall \alpha^* \in X_* \tag{26}$$

Proof. Invoking Lemma 1 with $\alpha = \alpha^*$, we obtain

$$
\begin{aligned}
2\eta(F(\alpha^*) - F(\alpha_{n+1})) \quad &\geq \|\alpha_{n+1} - \alpha_n\|_2 + 2\langle \alpha_n - \alpha^*, \alpha_{n+1} - \alpha_n \rangle \\
&= \|\alpha^* - \alpha_{n+1}\|_2 - \|\alpha^* - \alpha_n\|^2
\end{aligned}
\tag{27}
$$

Summing this inequality over $n = 0, \cdots, k - 1$ gives

$$2\eta\left(kF(\alpha^*) - \sum_{n=0}^{k-1} F(\alpha_{n+1})\right) \geq \|\alpha^* - \alpha_k\|^2 - \|\alpha^* - \alpha_0\|^2 \tag{28}$$

Invoking Lemma 1 one more time with $\alpha = \alpha_n$ yields

$$2\eta(F(\alpha_n) - F(\alpha_{n+1})) \geq \|\alpha_{n+1} - \alpha_n\|^2 \tag{29}$$

Multiplying the last inequality by n and summing over $n = 0, \cdots, k - 1$, we obtain

$$2\eta \sum_{n=0}^{k-1} (nF(\alpha_n) - (n+1)F(\alpha_{n+1}) + F(\alpha_{n+1})) \geq \sum_{n=0}^{k-1} n\|\alpha_{n+1} - \alpha_n\|^2 \tag{30}$$

which simplifies to

$$2\eta\left(-kF(\alpha_k) + \sum_{n=0}^{k-1} F(\alpha_{n+1})\right) \geq \sum_{n=0}^{k-1} n\|\alpha_{n+1} - \alpha_n\|^2 \tag{31}$$

Adding (28) and (31), we obtain

$$2k\eta(F(\alpha^*) - F(\alpha_k)) \geq \|\alpha^* - \alpha_k\|^2 + \sum_{n=0}^{k-1} n\|\alpha_{n+1} - \alpha_n\|^2 - \|\alpha^* - \alpha_0\|^2 \tag{32}$$

and hence it follows that

$$k \leq \frac{\|\alpha^* - \alpha_0\|^2}{2\eta(F(\alpha_k) - F(\alpha^*))} \tag{33}$$

The above result shows that when α_k converge to the optimal solution, the number of iterations is at most $\|\alpha^* - \alpha_0\|^2 / 2\eta(F(\alpha_k) - F(\alpha^*))$. □

Lemma 2. *The sequences* $\{\alpha_k, \theta_k\}$ *generated via FISTA with either a constant or backtracking stepsize the satisfy for every* $k \geq 1$

$$2\eta(t_k^2 v_k - t_{k+1}^2 v_{k+1}) \geq \|u_{k+1}\|_2 - \|u_k\|^2 \tag{34}$$

where $v_k = F(\alpha_k) - F(\alpha^*)$.

Lemma 3. *Let* $\{a_k, b_k\}$ *be the position sequence of real satisfying*

$$a_k - a_{k+1} > b_{k+1} - b_k \ \forall k \geq 1, \ with \ a_1 + b_1 \leq c, c > 0 \tag{35}$$

Then, $a_k < c$ for every $k \geq 1$.

Lemma 4. *The positive sequence $\{t_k\}$ generated in FISTA via (2.3) with $t_1 = 1$ satisfies $t_k \geq (k+1)/2$ for all $k \geq 1$.*

Theorem 2. *Let $\{a_k\}$, $\{b_k\}$ be generated by FISTA. Then, for any $k \geq 1$*

$$F(\boldsymbol{\alpha}_k) - F(\boldsymbol{\alpha}^*) \leq \frac{\|\boldsymbol{\alpha}_0 - \boldsymbol{\alpha}^*\|^2}{2\eta(k+1)^2} \, \forall \boldsymbol{\alpha}^* \in X_* \tag{36}$$

Proof. Let us define the quantities

$$a_k = 2\eta t_k^2 v_k, \quad b_k = \|u_k\|^2, \quad c = \|\boldsymbol{\theta}_1 - \boldsymbol{\alpha}^*\|_2 - \|\boldsymbol{\alpha}_0 - \boldsymbol{\alpha}^*\|^2$$

and recall that $v_k = F(\boldsymbol{\alpha}_k) - F(\boldsymbol{\alpha}^*)$. Then, using Lemma 3, we have for every $k \geq 1$

$$a_k - a_{k+1} > b_{k+1} - b_k$$

and hence assuming that $a_1 + b_1 \leq c$ holds true, invoking Lemma 3, we obtain that

$$2\eta t_k^2 v_k \leq \|\boldsymbol{\alpha}_0 - \boldsymbol{\alpha}^*\|^2 \tag{37}$$

which, combined with $t_k \geq (k+1)/2$, yields. $\square$

$$k \leq \sqrt{\frac{\|\boldsymbol{\alpha}_0 - \boldsymbol{\alpha}^*\|^2}{2\eta(F(\boldsymbol{\alpha}_k) - F(\boldsymbol{\alpha}^*))}} - 1 \tag{38}$$

It is important to note that, when $\boldsymbol{\alpha}_k$ converge to the optimal solution, the number of iterations is at most $\sqrt{\|\boldsymbol{\alpha}^* - \boldsymbol{\alpha}_0\|^2 / 2\eta(F(\boldsymbol{\alpha}_k) - F(\boldsymbol{\alpha}^*))} - 1$. Obviously, FISTA has fewer iterations than ISTA. The optimized algorithms are shown separately in Algorithms 5–7.

Algorithm 5 Fast Negentropy + ℓ_1 Norm

1: Input: measurement matrix $\mathbf{D}$, measurement signal $\mathbf{x}$.
2: Initialization: $\eta \in \left(0, 1/\|\mathbf{D}^T\mathbf{D}\|\right)$, $\lambda > 0$, $\theta_1 = \boldsymbol{\alpha}_0 = 0$, $k = 0$, $t_1 = 1$
3: Main iteration: increment k by 1
4: $\beta_i^{k-1} = \theta_i^{k-1} - \eta\nabla f(\theta_i^{k-1})$
5: $\alpha_i^k = \text{sign}(\beta_i^{k-1})\max(0, |\beta_i^{k-1}| - \lambda\eta)$
6: $t_{k+1} = \dfrac{1 + \sqrt{1 + 4t_k^2}}{2}$
7: $\boldsymbol{\theta}_{k+1} = \boldsymbol{\alpha}_k + \left(\dfrac{t_k - 1}{t_{k+1}}\right)(\boldsymbol{\alpha}_k - \boldsymbol{\alpha}_{k-1})$
8: Stopping rule: stop if $\boldsymbol{\alpha}_k$ has converged
9: Output: sparse signal $\boldsymbol{\alpha}$

Algorithm 6 Fast Negentropy + Weighted ℓ_1 Norm

1: Input: measurement matrix **D**, measurement signal **x**.
2: Initialization: choose proper η, λ, c, p and δ, $\theta_1 = \boldsymbol{\alpha}_0 = 0$, $k = 0$, $t_1 = 1$
3: Main iteration: increment k by 1
4: $\beta_i^{k-1} = \theta_i^{k-1} - \eta \nabla f(\theta_i^{k-1})$
5: $\alpha_i^k = \text{sign}(\beta_i^{k-1}) \max(0, |\beta_i^{k-1}| - \lambda \eta w_i^{k-1})$
6: $t_{k+1} = \dfrac{1 + \sqrt{1 + 4t_k^2}}{2}$
7: $\boldsymbol{\theta}_{k+1} = \boldsymbol{\alpha}_k + \left(\dfrac{t_k - 1}{t_{k+1}}\right)(\boldsymbol{\alpha}_k - \boldsymbol{\alpha}_{k-1})$
8: Stopping rule: stop if $\boldsymbol{\alpha}_k$ has converged
9: Output: sparse signal $\boldsymbol{\alpha}$

Algorithm 7 Fast Negentropy + $\ell_{0.5}$ Norm

1: Input: measurement matrix **D**, measurement signal **x**.
2: Initialization: choose proper η, λ, c,
3: $\theta_1 = \boldsymbol{\alpha}_0 = 0, k = 0, t_1 = 1$
4: Main iteration: increment k by 1
5: $\beta_i^{k-1} = \theta_i^{k-1} - \eta \nabla f(\theta_i^{k-1})$
6: $\alpha_i^k = \frac{2}{3} x (1 + \cos(\frac{2}{3} \cos^{-1}(-\frac{3^{3/2}}{4} \lambda \eta |x|^{-3/2}))) \text{sign}(\max(0, |\beta_i^{k-1}| - \frac{3}{2} \lambda \eta^{\frac{2}{3}}))$
7: $t_{k+1} = \dfrac{1 + \sqrt{1 + 4t_k^2}}{2}$
8: $\boldsymbol{\theta}_{k+1} = \boldsymbol{\alpha}_k + \left(\dfrac{t_k - 1}{t_{k+1}}\right)(\boldsymbol{\alpha}_k - \boldsymbol{\alpha}_{k-1})$
9: Stopping rule: stop if $\boldsymbol{\alpha}_k$ has converged
10: Output: sparse signal $\boldsymbol{\alpha}$

---

3. Numerical Experiments

In this section, we conduct two groups of experiments to evaluate the proposed algorithm. In the first group of experiments, we compare the proposed solution with an MSE-based method to recover simulated sparse signals and image denoising under Gaussian and non-Gaussian noise conditions so as to illustrate the superiority of using negentropy as the measurement error. Further, in the second group of experiments, the effectiveness and practicability of the proposed algorithm are verified by comparing with some robust sparse recovery algorithms to recover simulated sparse signals and real-world MRI images under non-Gaussian noise conditions.

3.1. Comparison with MSE-Based Method

(1) Sparse Signal Recovery

In this subsection, the performance of the proposed algorithm in recovering sparse signals from noisy measurements is presented. For the sake of comparison, the results of ISTA are also presented. The parameter settings are summarized in Table 1.

Table 1. The parameter setting of algorithms.

Algorithm	Parameters
ISTA	$\lambda = 3, \eta = 4 \times 10^{-3}$
Fast Negentropy + ℓ_1 norm	$\lambda = 3, \eta = 4 \times 10^{-3}, c = 1$
Fast Negentropy + weighted ℓ_1 norm	$\lambda = 3, \eta = 4 \times 10^{-3}, c = 1, \delta = 1 \times 10^{-7}, p = 0.9$
Fast Negentropy + $\ell_{0.5}$ norm	$\lambda = 2, \eta = 2.5 \times 10^{-3}, c = 1$

The experiments were performed on a 100-dimension 5-sparsity signal, where the non-zero values had random positions and the values were generated following a uniform distribution in the range $[-2, -1] \cup [1, 2]$. The measurement matrix **D** was generated by randomly drawing value from normal distribution $N(0, 1)$. In our experiments, we generate 100 samples of sparse signals and measurement matrices, and the corresponding measurements are obtained from Equation (1). The measurement signals were further corrupted by Gaussian and non-Gaussian noise. In the Gaussian case, the noise is generated from $N(0, 1)$ distribution. In the non-Gaussian case, impulse noise is adopted, which is generated by a mixed Gaussian distribution model as suggested in [36]. The reconstruction performance is evaluated by the relative norm error between the reconstructed signal and the original signal, which is defined as

$$\text{Relative norm error}(\boldsymbol{\alpha}_i) = \frac{\left\|\boldsymbol{\alpha}_{origi} - \boldsymbol{\alpha}_i\right\|_2}{\left\|\boldsymbol{\alpha}_{origi}\right\|_2} \tag{39}$$

Figure 1 presents the performance of ISTA and our proposed algorithms with 16 dB of Gaussian noise and non-Gaussian noise. As shown in Figure 1a, the relative error of ISTA is slightly smaller than that of the proposed negentropy algorithm with ℓ_1 norm constraint, which indicates that the proposed negentropy objective function has no obvious advantage in recovering signals with Gaussian noise. With non-Gaussian noise, however, it is shown in Figure 1b that the negentropy-based algorithms have higher recovery accuracy, which indicates that the proposed negentropy objective function has a stronger adaptability to non-Gaussian noise. Moreover, when weighted ℓ_1 norm and $\ell_{0.5}$ norm were adopted as sparse constraints, the error further decreases, which implies that ℓ_p norm as sparse constraint can improve the accuracy of algorithms.
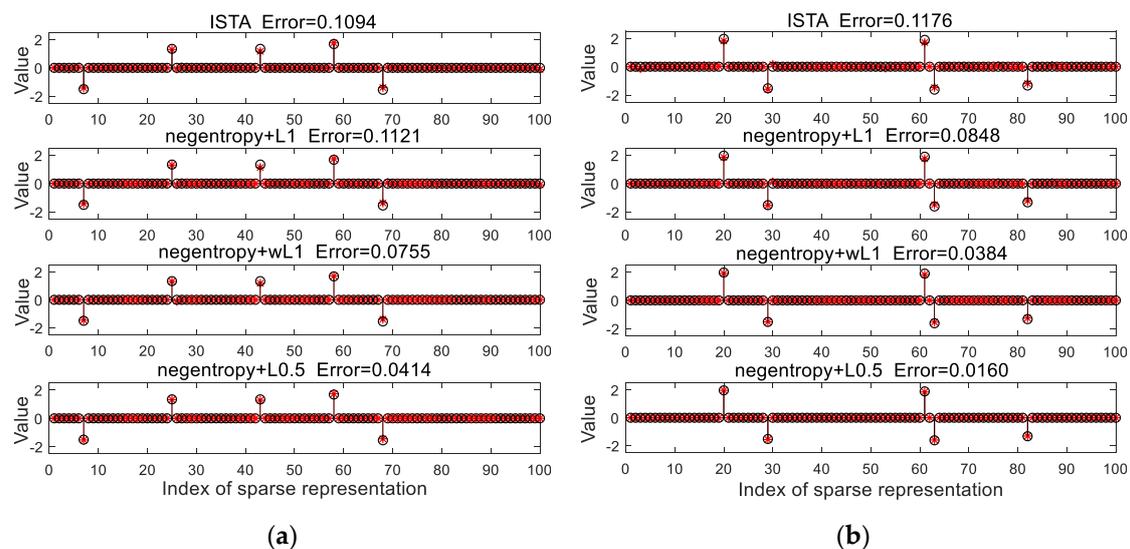
Figure 1. Comparison between original signal (black) and recovered signal (red): (**a**) with Gaussian noise; (**b**) with non-Gaussian noise.

Figure 2 presents how the average reconstruction error varies with the signal-to-noise ratio (SNR). The average relative norm error is obtained by averaging over 100 samples of signals and measurement matrices. As Figure 2a shows, the proposed negentropy-based methods have higher reconstruction accuracies under various SNRs. The performance gains further increase when the measurements are contaminated by non-Gaussian noise. As we can see from Figure 2b, the relative error of ISTA is within the level of 10^{-1}, but the relative error of the proposed algorithm can reach the scale of 10^{-2}, which means that the proposed algorithm model is more robust than conventional algorithms in non-Gaussian noise. Moreover, when weighted ℓ_1 norm or $\ell_{0.5}$ norm were adopted as sparse constraints, the recovery accuracy further increases, and the $\ell_{0.5}$ norm case has superior performance, under which lower reconstruction error can be achieved with high probability.
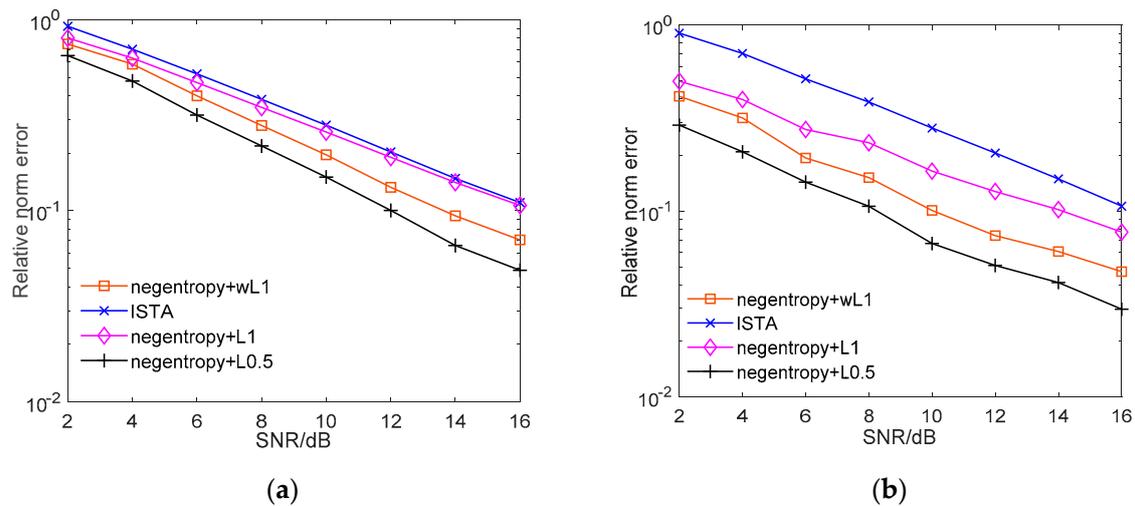
Figure 2. Relative norm error performance versus signal-to-noise ratio (SNR). (**a**) Comparison between ISTA and proposed algorithm with Gaussian noise; (**b**) comparison between ISTA and proposed algorithm with non-Gaussian noise.

Figure 3 further illustrates the reconstruction performance with non-Gaussian noisy measurements under various compression ratios. As shown in Figure 3, as the compression ratio decreases, despite the performance of the algorithms deteriorating, the negentropy algorithm always maintains a lower relative error. The $\ell_{0.5}$ norm-based negentropy algorithm presents the best performance, which indicates that the negentropy algorithm can exhibit more stable performance with different compression ratios and also proves that ℓ_p norm has superior sparse constraint effect to ℓ_1 norm.

The impact of sparsity on the reconstruction performance is further presented in Figure 4. As Figure 4 shows, the reconstruction error increases with the sparsity. The negentropy algorithms present better performance than ISTA. The reconstruction error of the negentropy algorithm based on ℓ_1 norm or weighted ℓ_1 norm can only reach the level of 10^{-1} when the signal has low sparsity levels. The negentropy algorithm based on $\ell_{0.5}$ norm has the best performance; the relative error is always kept within 10^{-2} level of magnitude, which shows better suitability for signals with low sparsity.

Figure 5 depicts the relative error with iterations. As the number of iterations increases, the relative error will gradually decrease and eventually converge. As we can see from Figure 5, the convergence performance can be significantly improved after convergence optimization. Taking the negentropy algorithm based on $\ell_{0.5}$ norm as an example, the curve tends to be gentle after 75 iterations without accelerating convergence. After the convergence rate is optimized, convergence is achieved in about 40 iterations. Thus, by means of optimizing the convergence rate based on FISTA, the negentropy algorithm also proves stronger convergence is guaranteed compared to ISTA.

(**a**)　　　　　　　　　　　　　　　　　　　　　(**b**)

(**c**)
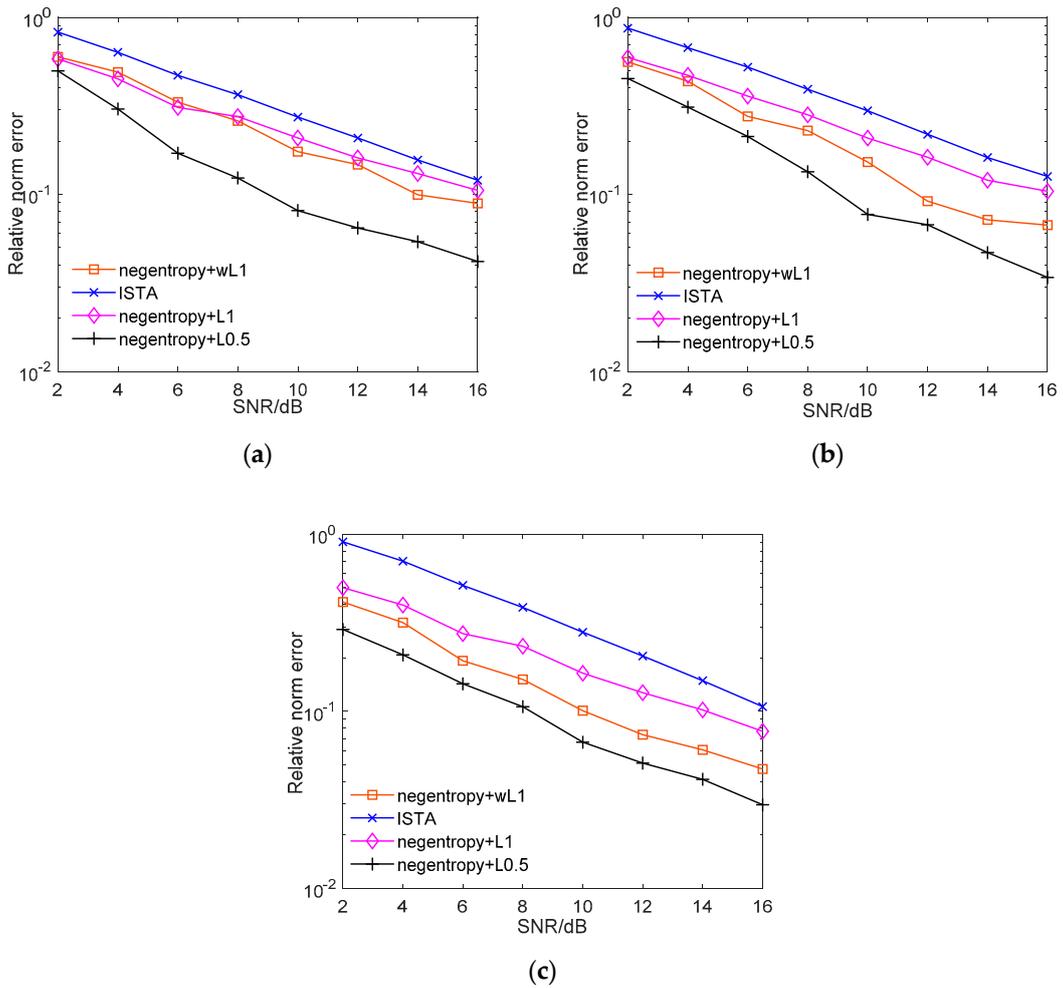
Figure 3. Relative norm error performance versus SNR. (**a**) Comparison between ISTA and proposed algorithm with 40% compression; (**b**) comparison between ISTA and proposed algorithm with 50% compression; (**c**) comparison between ISTA and proposed algorithm with 60% compression.
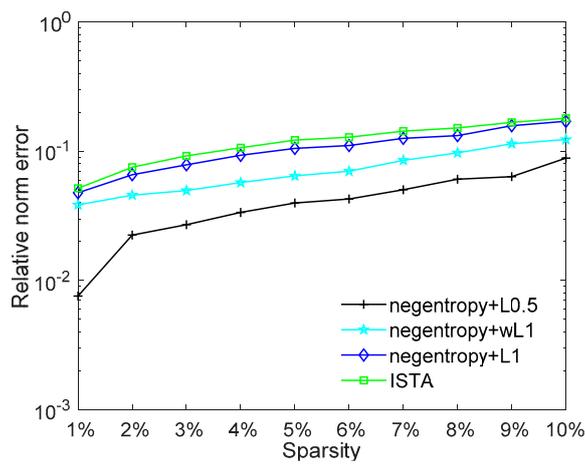
Figure 4. Relative norm error performance versus sparsity.
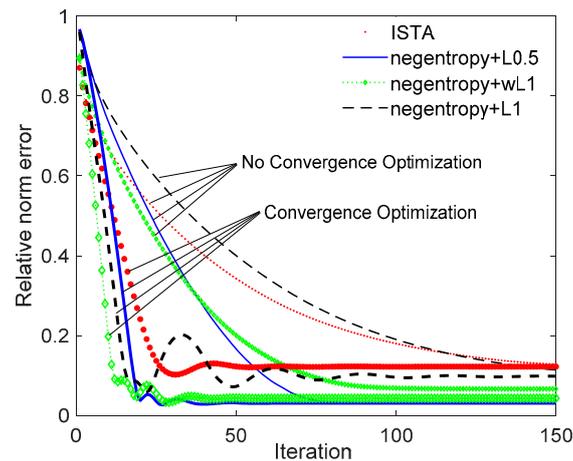
Figure 5. Relative norm error performance versus iterations.

(2) Image Denoising Experiments

In order to test the performance of the proposed algorithms in the application, we applied them to image denoising, which performs better compared with ISTA in these experiments. The details of the image denoising experiment are described below. We selected the image named 'Cameraman' as the target, which is a 512×512-pixel gray-scaled photograph. Gaussian noise and salt and pepper noise were added to the image, which was separated into 12×12-pixel small patches with an interval of 2 between patches to form an input set $\mathbf{X} \in \mathbb{R}^{144 \times 63001}$. According to the signal model in Equation (1), we generate a 144×256 overcomplete discrete cosine transform (DCT) distributed dictionary for the denoising task. The sparse representation of each patch in the dictionary is recovered by the proposed algorithm and ISTA, which removes the noise in the images. The parameter settings of the algorithms are summarized in Table 2. The peak signal-to-noise ratio (PSNR) of the recovered images and the structural similarity (SSIM) index between the recovered images and original images were used to evaluate the denoising performance.

Table 2. The parameter setting of algorithms in image denoising experiment.

Algorithm	Parameters
ISTA	$\lambda = 33, \eta = 5 \times 10^{-2}$
Fast iterative Negentropy ℓ_1 norm	$\lambda = 1.5, \eta = 0.3, c = 1$
Fast iterative Negentropy + weighted ℓ_1 norm	$\lambda = 0.65, \eta = 0.6, c = 1, \delta = 1 \times 10^{-7}, p = 0.9$
Fast iterative Negentropy + $\ell_{0.5}$ norm	$\lambda = 2.2, \eta = 0.6, c = 1$

Figure 6 shows the results of recovering images with 20.17 dB Gaussian noise using ISTA and the proposed algorithms. By comparing (c) and (d), we can see that the PSNR value of the image recovered by ISTA is 1 dB larger than that of the negentropy algorithm. However, from the perspective of visual effects, the image recovered by the negentropy algorithm based on ℓ_1 norm are relatively clear, and the image recovered by ISTA are somewhat blurred. Therefore, the negentropy algorithm has a certain denoising effect for images with Gaussian noise, but it has no obvious advantage over ISTA.
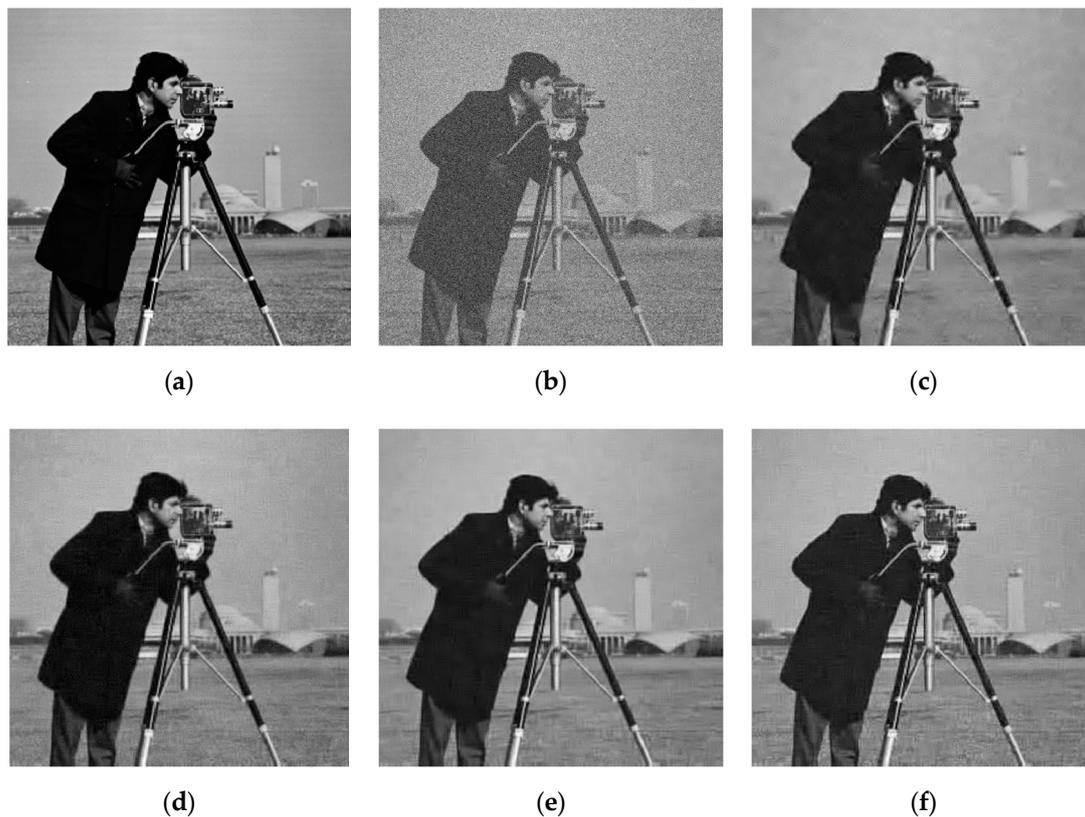
Figure 6. Image denoising result with Gaussian noise: (**a**) original image; (**b**) noised image; (**c**) ISTA; (**d**) negentropy + ℓ_1-norm; (**e**) negentropy + weighted ℓ_1 -norm; (**f**) negentropy + $\ell_{0.5}$ -norm.

Figure 7 is the results of recovering an image with salt and pepper noise with 30% density using ISTA and the proposed algorithms, which shows that the images recovered using ISTA are very blurred and the PSNR is only 20.24 dB. It can be judged that ISTA can hardly recover the images corrupted by strong salt and pepper noise. The proposed algorithms, however, reconstruct the original image clearly and the PSNR value is over 30 dB, which shows that the algorithm based on negentropy is able to adapt to the influence of non-Gaussian noise. Compared with (d–f), it can be seen that the PSNR value of the image recovered using the negentropy algorithm based on ℓ_p norm is 1.3 dB higher than that recovered using the negentropy algorithm based on ℓ_1 norm, and the visual effects are better, which demonstrates that ℓ_p norm, as a sparse constraint, can make the reconstruction accuracy of the algorithm higher and plays a key role in the restoration of image details. For the convenience of comparison, the PSNR results in two noise conditions are shown in Table 3.

The PSNR and SSIM index values of the denoised images are presented in Figure 8. As we can see, both the PSNR and the SSIM index decrease gradually as the Gaussian noise intensity increases. Despite the fact that the ISTA algorithm presents comparable performance with the proposed algorithms when the standard deviation of the Gaussian noise is around 20, its performance rapidly decreases with the noise intensity. When the Gaussian noise is strong, the proposed algorithm presents much better performance in both PSNR and SSIM, which indicates that the negentropy algorithms have strong adaptability to different noise intensity.
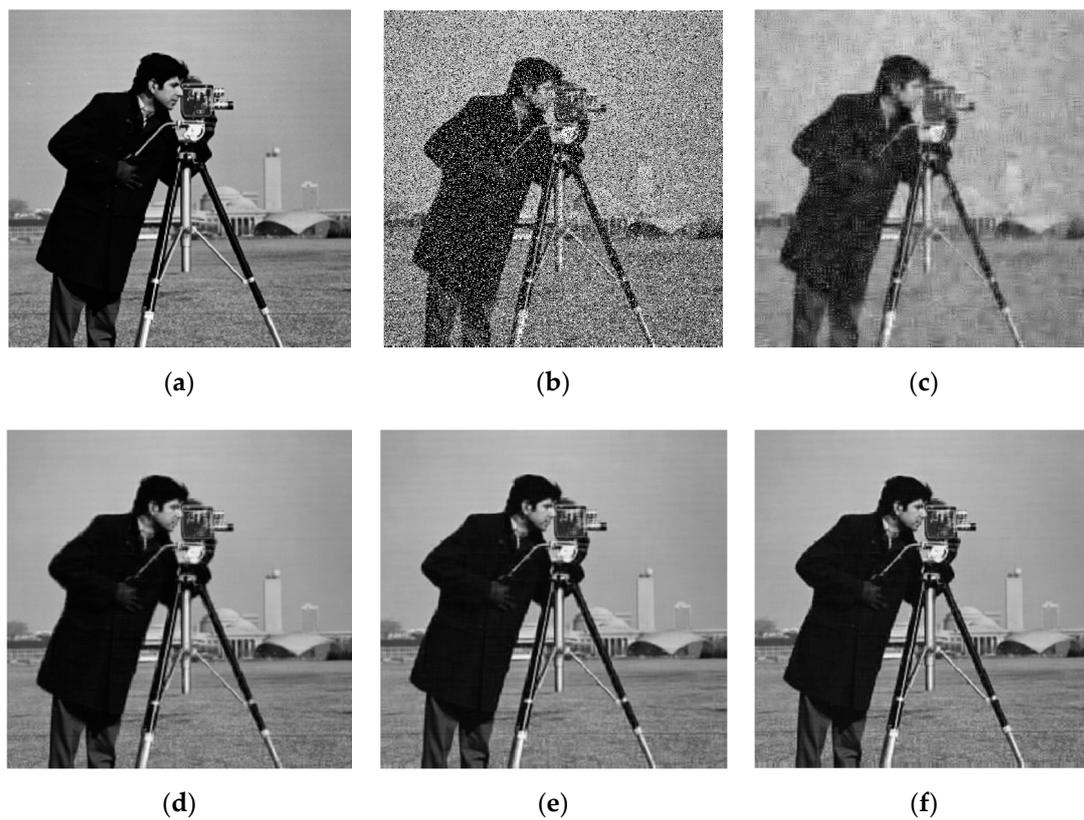
Figure 7. Image denoising result with salt and pepper noise: (**a**) original image; (**b**) noised image; (**c**) ISTA; (**d**) negentropy + ℓ_1-norm; (**e**) negentropy + weighted ℓ_1 -norm; (**f**) negentropy + $\ell_{0.5}$ -norm.

Table 3. Denoising results (peak signal-to-noise ratio (PSNR) in dB) of the different methods in two noise conditions.

Noise Conditions	Noised Image	ISTA	Negentropy + L1	Negentropy + wL1	Negentropy + L0.5
Gaussian noise	20.17	30.34	29.31	30.67	30.86
Salt and pepper noise	10.35	20.24	30.24	31.69	32.74
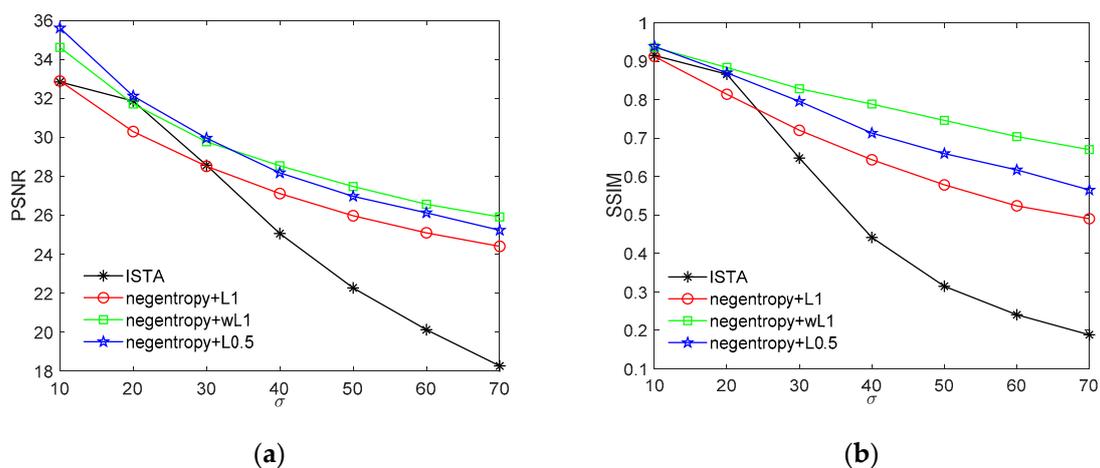
Figure 8. Comparison between ISTA and the proposed algorithm for recovering image with Gaussian noise with standard deviation ranging from 10 to 70. (**a**) PSNR value of recovered image versus noise intensity; (**b**) structural similarity (SSIM) value of recovered image versus noise intensity.

Figure 9 presents the denoising performance with various densities of salt and pepper noise. As we can observe from Figure 9, the ISTA algorithm can only recover images with salt and pepper noise density below 20%, and the recovered image is extremely blurred, where the maximum SSIM value only reaches 0.74. Furthermore, when the noise density in the image is higher, ISTA is unable to recover a clear image. The negentropy algorithm, on the other hand, can recover the original image with much higher accuracy. For instance, when the noise density in the image is less than 30%, all three negentropy-based methods have PSNR values of more than 30 dB and SSIM values of more than 0.9. The negentropy algorithm with $\ell_{0.5}$ norm achieves the best performance, with PSNR = 38 dB and SSIM = 0.97, when the noise density is 10%. With weighted ℓ_1 norm as constraint, although the PSNR value is lower than $\ell_{0.5}$ norm, the structure similarity between the recovered images and the original images is higher, leading to better visual effects.

(**a**)　　　　　　　　　　　　　　　　　　(**b**)
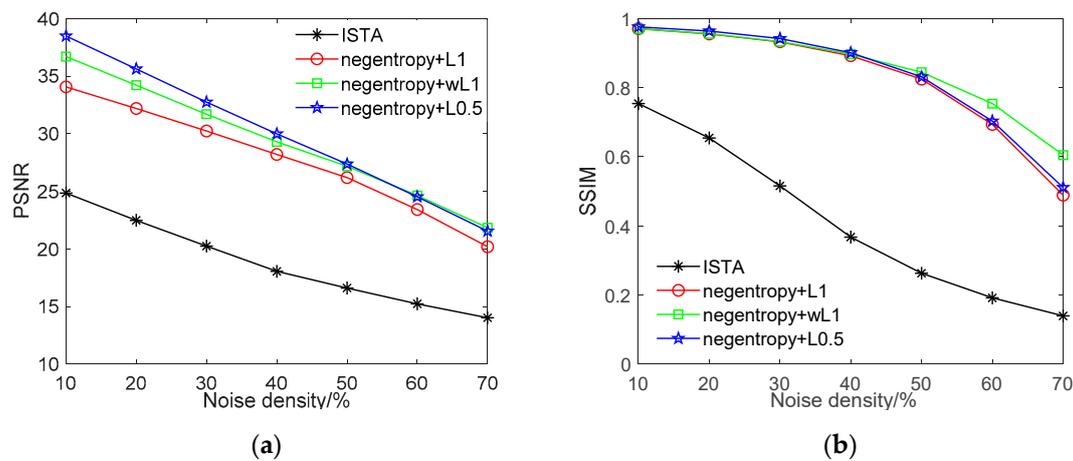
Figure 9. Comparison between ISTA and proposed algorithm for recovering image with salt and pepper noise with density from 10% to 70%. (**a**) PSNR value of recovered image versus noise density; (**b**) SSIM value between recovered images and original images versus noise density.

3.2. Comparison with Robust Methods

(1) Sparse Signal Recovery

In this part, we conduct sparse signal recovery experiments under non-Gaussian noise and use several well-known, robust sparse recovery algorithms for comparison, including Huber-FISTA [20], YALL1 [21] and LqLA-ADMM [22]. The construction of the simulated k-sparse signal is similar to that in Section 3.1. The length of the sparse signal and the number of measurements are set at N = 100, M = 60. An $M \times N$ orthonormal Gaussian random matrix is used as the measurement matrix. Impulse noise with SNR = 16 dB. Since the original sparse signal is unknown, we choose $\lambda = \xi \|\mathbf{D}^T \mathbf{x}\|_\infty$ (as proposed in [37]) with $\xi = 0.05$ for all algorithms.

Figure 10 presents the recovery performance of the compared algorithms versus sparsity with $k \in [1 \sim 10]$. All algorithms are conducted 300 times for each k. Figure 11 shows the relative error versus iterations with fixed k = 5. As shown in these figures, the negentropy algorithms (except the ℓ_1 norm-based negentropy algorithm in low sparsity) outperform other compared algorithms in terms of both recovery performance and convergence speed, especially in the case of $\ell_{0.5}$ norm.
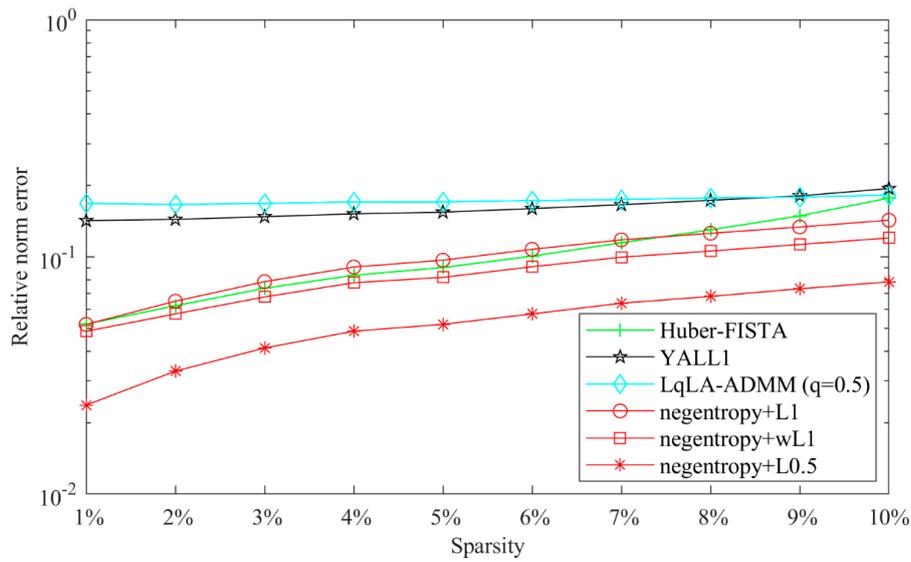
Figure 10. Relative norm error performance versus sparsity.
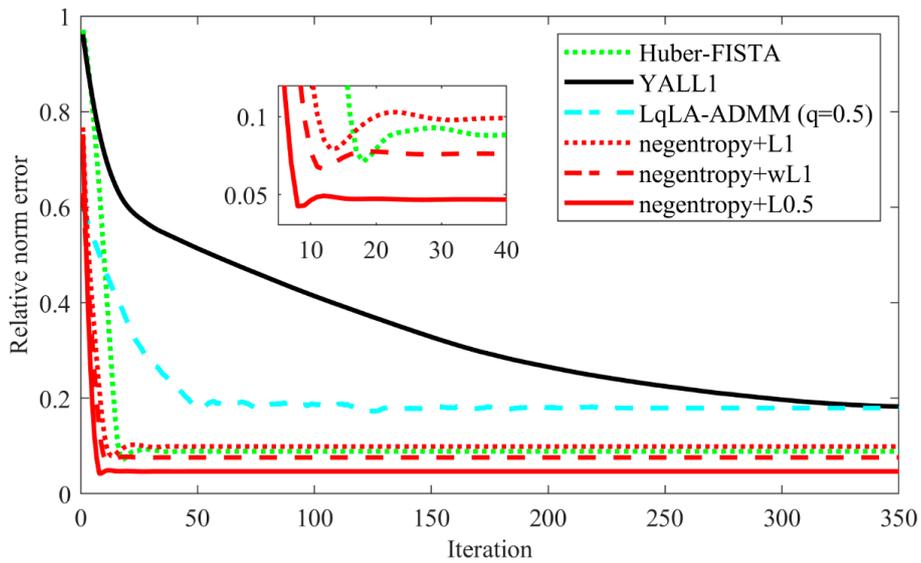
Figure 11. Relative norm error performance versus iterations.

(2) Magnetic Resonance Imaging Example

　　To further verify the effectiveness and practicability of the proposed algorithms, this part will focus on the reconstruction of medical images. The used MRI images are of size 256×256, for a brain MRI and sinusitis MRI, as shown in Figure 12, and evaluate the recovery performance by the value of PSNR. We utilize the Haar wavelets as the basis for sparse representation of the images. The compression ratio M/N is set as 0.4 and the measurement matrix is a partial DCT matrix. A Gaussian mixture model (GMM) $(f = \rho N(0, \sigma^2) + (1 - \rho)N(0, k\sigma^2))$ is taken to model the impulsive noise, where the parameters $\rho \in (0, 1)$ and $k > 1$ respectively control the proportion and the strength of outliers in the noise [19]. To ensure fairness of comparison, the λ in each algorithm is selected by providing the best performance in terms of relative error of recovery [22].
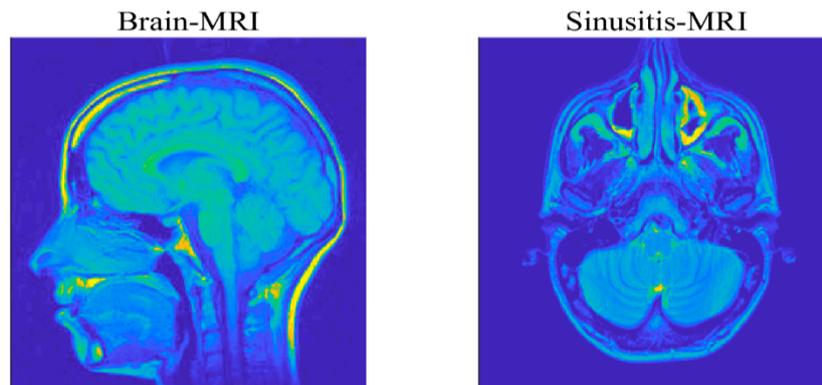
Figure 12. Two test MRI images.

Figure 13 shows the recovered MRI images of all algorithms under the GMM noise with $\rho = 0.9, k = 1000$ and SNR = 20 dB. The PSNR results are shown in Table 4. It can be seen that each algorithm can successfully reconstruct the two MRI images. Quantitatively, it is observed that the PSNR of the MRI images recovered using negentropy algorithms based on ℓ_p norm have a higher value than those recovered by other compared algorithms and the negentropy algorithm based on $\ell_{0.5}$ norm obtains the best recovery performance. Particularly, for the brain MRI image, the PSNR gains of negentropy + $\ell_{0.5}$ over Huber-FISTA, YALL1 and LqLA-ADMM (q = 0.5) are 0.53 dB, 1.13 dB and 0.31 dB, respectively. Furthermore, Figure 14 presents the convergence curve of PSNR versus iterations in recovering the brain MRI image. It can be seen that compared with LqLA-ADMM (q = 0.5) and Huber-FISTA, the negentropy algorithm requires less iterations to converge. Thus, the proposed algorithm is more efficient for the recovery of MRI images in terms of accuracy and convergence speed.

Table 4. Comparison of PSNR (dB) values obtained by different methods under Gaussian mixture model (GMM) noise.

Images	Huber-FISTA	YALL1	LqLA-ADMM (q = 0.5)	Negentropy + L1	Negentropy + wL1	Negentropy + L0.5
Brain MRI	25.89	25.29	26.11	25.89	26.19	26.42
Sinusitis MRI	28.36	27.93	28.42	28.36	28.55	28.57
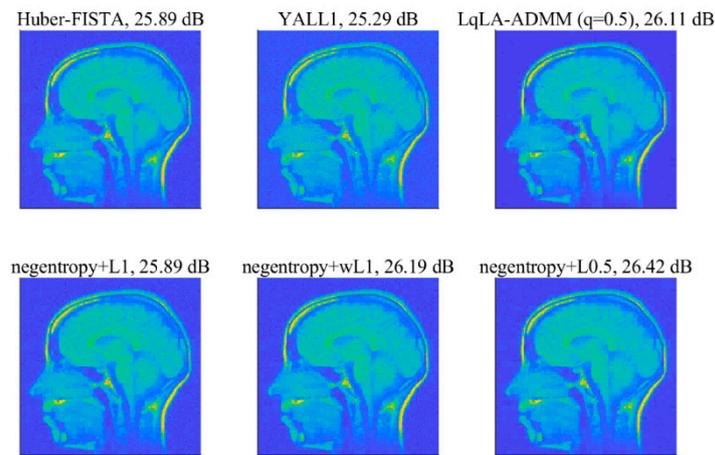
(**a**)
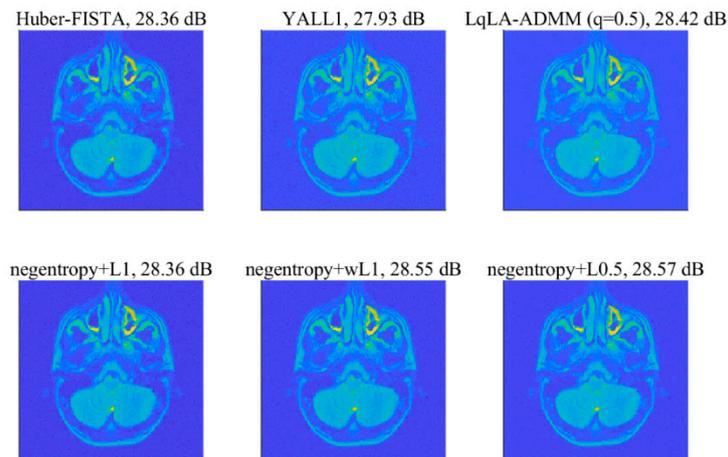
(**b**)

Figure 13. Recovery results of MRI images with the GMM noise: (**a**) brain MRI.; (**b**) sinusitis MRI.
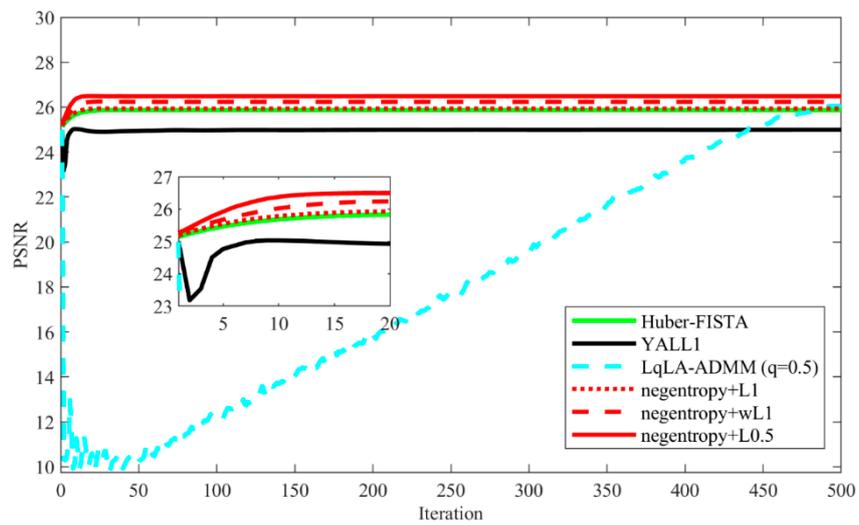
Figure 14. PSNR versus iterations in recovering the brain MRI image.

4. Conclusions

In this paper, we propose a sparse signal recovery model based on negentropy maximization. To improve the robustness of the sparse signal recovery algorithm under different types of noise interference, ℓ_p norm is adopted as the sparsity constraint. We presented two approaches to solve the optimization problem. One is to approximate the ℓ_p norm to the weighted ℓ_1 norm, which is a convex function and joins the information of the sparse signal from the previous iteration, and then to solve the minimum value using the corresponding approximation operator. The other is to use the approximate operator of IHTA to solve the minimum value of the problem to solve the problem with $\ell_{0.5}$ constraint. To improve the convergence performance, we further optimize the convergence rate of the proposed algorithm based on FISTA to reduce the number of iterations. Numerical experiments were conducted on sparse signal reconstruction, image denoising and MRI images recovery.

In the sparse signal reconstruction task, the proposed algorithms presented better reconstruction performance over ISTA, especially with non-Gaussian noise. Furthermore, the algorithm based on ℓ_p norm can recover the sparse signal more accurately with higher probability than all the other algorithms. In the image denoising task, the negentropy algorithm has a better denoising effect over the ISTA algorithm when the images are corrupted by non-Gaussian noise, which is more significant when salt and pepper noise is considered. Moreover, the image recovered by the ℓ_p norm-based algorithm is better than that by the ℓ_1 norm-based algorithm regarding both visual effects and objective evaluation criteria. In the MRI images recovery task, under the circumstance of GMM noise, the proposed negentropy algorithm based on $\ell_{0.5}$ norm can achieve the best recovery performance among the reference algorithms and requires less iterations to converge. In further work, we plan to develop a data-driven approach for adaptively adjusting the parameters of the proposed method and investigate the effectiveness of our approach for wireless sensor networks and the internet of things.

References

1. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306. [CrossRef]
2. Candes, E.; Romberg, J.; Tao, T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory* **2006**, *52*, 489–509. [CrossRef]
3. Boche, H.; Calderbank, R.; Kutyniok, G.; Vybiral, J. *Compressed Sensing and Its Applications*; Springer: Berlin/Heidelberg, Germany, 2015.
4. Rani, M.; Dhok, S.B.; Deshmukh, R.B. A systematic review of compressive sensing: Concepts, implementations and applications. *IEEE Access* **2018**, *6*, 4875–4894. [CrossRef]
5. Gao, Z.; Dai, L.; Han, S.; Chih-Lin, I.; Wang, Z.; Hanzo, L. Compressive sensing techniques for next-generation wireless communications. *IEEE Wirel. Commun.* **2018**, *25*, 144–153. [CrossRef]
6. Wang, L.; Yin, X.; Yue, H.; Xiang, J. A Regularized Weighted Smoothed L0 Norm Minimization Method for Underdetermined Blind Source Separation. *Sensors* **2018**, *18*, 4260. [CrossRef]
7. Goyal, V.K.; Fletcher, A.K.; Rangan, S. Compressive sampling and lossy compression. *IEEE Signal Process. Mag.* **2008**, *25*, 48–56. [CrossRef]
8. Wang, Q.; Meng, C.; Ma, W.; Wang, C.; Yu, L. Compressive sensing reconstruction for vibration signals based on the improved fast iterative shrinkage-thresholding algorithm. *Measurement* **2019**, *142*, 68–78. [CrossRef]
9. Hayashi, K.; Nagahara, M.; Tanaka, T. A user's guide to compressed sensing for communications systems. *IEICE Trans. Commun.* **2013**, *96*, 685–712. [CrossRef]

10. Wang, D.; Li, L.; Ji, Y.; Yan, Y. Model recovery for Hammerstein systems using the auxiliary model based orthogonal matching pursuit method. *Appl. Math. Model.* **2018**, *54*, 537–550. [CrossRef]
11. Needell, D.; Tropp, J.A. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.* **2009**, *26*, 301–321. [CrossRef]
12. Needell, D.; Vershynin, R. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. *Found. Comput. Math.* **2009**, *9*, 317–334. [CrossRef]
13. Bredies, K.; Lorenz, D.A. Linear convergence of iterative soft-thresholdong. *J. Fourier Anal. Appl.* **2008**, *14*, 813–837. [CrossRef]
14. Chen, S.S.; Donoho, D.L.; Saunders, M.A. Atomic decomposition by basis pursuit. *SIAM Rev.* **2001**, *43*, 129–159. [CrossRef]
15. Kim, N.; Byun, H.G.; You, Y.H.; Kwon, K. Blind signal processing for impulsive noise channels. *J. Commun. Netw.* **2012**, *14*, 27–33. [CrossRef]
16. Karthikeyan, P.; Vasuki, S. Efficient decision based algorithm for the removal of high density salt and pepper noise in images. *J. Commun. Technol. Electron.* **2016**, *61*, 963–970. [CrossRef]
17. Kuai, X.; Sun, H.; Zhou, S.; Cheng, E. Impulsive noise mitigation in underwater acoustic OFDM systems. *IEEE Trans. Veh. Technol.* **2016**, *65*, 8190–8202. [CrossRef]
18. Carrillo, R.E.; Barner, K.E. Lorentzian iterative hard thresholding: Robust compressed sensing with prior information. *IEEE Trans. Signal Process.* **2013**, *61*, 4822–4833. [CrossRef]
19. Pham, D.S.; Venkatesh, S. Improved image recovery from compressed data contaminated with impulsive noise. *IEEE Trans. Image Process.* **2012**, *21*, 397–405. [CrossRef]
20. Pham, D.S.; Venkatesh, S. Efficient algorithms for robust recovery of images from compressed data. *IEEE Trans. Image Process.* **2013**, *22*, 4724–4737. [CrossRef]
21. Yang, J.; Zhang, Y. Alternating direction algorithms for ℓ_1-problems in compressive sensing. *SIAM J. Sci. Comput.* **2011**, *33*, 250–278. [CrossRef]
22. Wen, F.; Pei, L.; Yang, Y.; Yu, W.; Liu, P. Efficient and robust recovery of sparse signal and image using generalized nonconvex regularization. *IEEE Trans. Comput. Imaging* **2017**, *3*, 566–579. [CrossRef]
23. Zhao, Y.; Liu, Z.; Wang, Y.; Wu, H.; Ding, S. Sparse Coding Algorithm with Negentropy and Weighted $\ell1$-Norm for Signal Reconstruction. *Entropy* **2017**, *19*, 599. [CrossRef]
24. Beck, A.; Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2009**, *2*, 183–202. [CrossRef]
25. Toh, K.C.; Yun, S. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pacific J. Optim.* **2010**, *6*, 615–640.
26. Sun, Y.; Babu, P.; Palomar, D.P. Majorization-minimization algorithms in signal processing, communications, and machine learning. *IEEE Trans. Signal Process.* **2017**, *65*, 794–816. [CrossRef]
27. Wei, Z.; Zhang, J.; Xu, Z.; Huang, Y.; Liu, Y.; Fan, X. Gradient Projection with Approximate L0 Norm Minimization for Sparse Reconstruction in Compressed Sensing. *Sensors* **2018**, *18*, 3373. [CrossRef] [PubMed]
28. Wen, F.; Chu, L.; Liu, P.; Qiu, R.C. A survey on nonconvex regularization-based sparse and low-rank recovery in signal processing, statistics, and machine learning. *IEEE Access* **2018**, *6*, 69883–69906. [CrossRef]
29. Zha, Z.; Zhang, X.; Wang, Q.; Tang, L.; Liu, X. Group-based sparse representation for image compressive sensing reconstruction with non-convex regularization. *Neurocomputing* **2018**, *296*, 55–63. [CrossRef]
30. Asif, M.S.; Romberg, J. Fast and accurate algorithms for re-weighted l1-norm minimization. *IEEE Trans. Signal Process.* **2013**, *61*, 5905–5916. [CrossRef]
31. Zhao, Y.B.; Li, D. Reweighted ℓ_1-minimization for sparse solution to underdetermined linear systems. *SIAM J. Optim.* **2012**, *22*, 1065–1088. [CrossRef]
32. Xu, Z.; Chang, X.; Xu, F.; Zhang, H. $L_{1/2}$ regularization: A thresholding representation theory and a fast solver. *IEEE Trans. Neural Netw. Learn. Syst.* **2012**, *23*, 1013–1027. [PubMed]
33. Zeng, J.; Lin, S.; Wang, Y.; Xu, Z. $L_{1/2}$ regularization: Convergence of iterative half thresholding algorithm. *IEEE Trans. Signal Process.* **2014**, *62*, 2317–2329. [CrossRef]
34. Yue, C.; Liang, J.; Qu, B.; Han, Y.; Zhu, Y.; Crisalle, O.D. A Novel Multiobjective Optimization Algorithm for Sparse Signal Reconstruction. *Signal Process.* **2020**, *167*, 107292. [CrossRef]
35. Wu, X.; Xiong, Y.; Huang, W.; Shen, H.; Li, M. An efficient compressive data gathering routing scheme for large-scale wireless sensor networks. *Comput. Electr. Eng.* **2013**, *39*, 1935–1946. [CrossRef]

36. Bershad, N.J. Comments on "A recursive least M-estimate algorithm for robust adaptive filtering in impulsive noise: Fast algorithm and convergence performance analysis". *IEEE Trans. Signal Process.* **2008**, *57*, 388–389. [CrossRef]

37. Wright, S.J.; Nowak, R.D.; Figueiredo, M.A. Sparse reconstruction by separable approximation. *IEEE Trans. Signal Process.* **2009**, *57*, 2479–2493. [CrossRef]