

## RESEARCH ARTICLE

# Entropy-Regularized Iterative Weighted Shrinkage-Thresholding Algorithm (ERIWSTA) for inverse problems in imaging

Limin Ma<sup>1</sup>, Bingxue Wu<sup>1</sup>, Yudong Yao<sup>2</sup>, Yueyang Teng<sup>1,3\*</sup>

**1** College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, Liaoning Province, China, **2** Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, United States of America, **3** Key Laboratory of Intelligent Computing in Medical Image, Ministry of Education, Shenyang, Liaoning Province, China

\* [tengyy@bmie.neu.edu.cn](mailto:tengyy@bmie.neu.edu.cn)

## OPEN ACCESS

**Citation:** Ma L, Wu B, Yao Y, Teng Y (2024) Entropy-Regularized Iterative Weighted Shrinkage-Thresholding Algorithm (ERIWSTA) for inverse problems in imaging. PLoS ONE 19(12): e0311227. <https://doi.org/10.1371/journal.pone.0311227>

**Editor:** Zi Xu, Shanghai University, CHINA

**Received:** August 30, 2023

**Accepted:** September 16, 2024

**Published:** December 27, 2024

**Copyright:** © 2024 Ma et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data are shared on Dryad and the link is as follows: <https://datadryad.org/stash/share/5rW5F45qSTpPSoudl9dJr66mWAAd3KoNMbJ8Sho5EII0>.

**Funding:** This work was supported by the Natural Science Foundation of Liaoning Province (2002-MS-114). The funder had a role in study conceptualization, formal analysis, investigation, methodology, supervision, and writing-review and editing of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Abstract

The iterative shrinkage-thresholding algorithm (ISTA) is a classic optimization algorithm for solving ill-posed linear inverse problems. Recently, this algorithm has continued to improve, and the iterative weighted shrinkage-thresholding algorithm (IWSTA) is one of the improved versions with a more evident advantage over the ISTA. It processes features with different weights, making different features have different contributions. However, the weights of the existing IWSTA do not conform to the usual definition of weights: their sum is not 1, and they are distributed over an extensive range. These problems may make it challenging to interpret and analyze the weights, leading to inaccurate solution results. Therefore, this paper proposes a new IWSTA, namely, the entropy-regularized IWSTA (ERIWSTA), with weights that are easy to calculate and interpret. The weights automatically fall within the range of [0, 1] and are guaranteed to sum to 1. At this point, considering the weights as the probabilities of the contributions of different attributes to the model can enhance the interpretation ability of the algorithm. Specifically, we add an entropy regularization term to the objective function of the problem model and then use the Lagrange multiplier method to solve the weights. Experimental results of a computed tomography (CT) image reconstruction task show that the ERIWSTA outperforms the existing methods in terms of convergence speed and recovery accuracy.

## Introduction

Inverse problems in imaging are essential for physical and biomedical sciences [1]. They have been widely used in optical and radar systems, including X-ray computed tomography (CT), positron emission tomography (PET), and electrical tomography (ET). An inverse problem in imaging aims to estimate an unknown image from the given measurements by solving an optimization problem with a regularizer. Mathematically, it is to infer an original signal  $x \in R^n$  from its measurements  $b = Ax \in R^m$ . Here,  $A \in R^{m \times n}$  is a linear random projection (matrix). Because  $m \ll n$ , this inverse problem is typically ill-posed.

Therefore, researchers typically construct an optimization problem and utilize a regularizer to solve it. However, the employed regularizer is usually nonsmooth, so it cannot be solved in a straightforward manner. Therefore, many first-order iterative proximal gradient methods have also been proposed, e.g., the iterative shrinkage-thresholding algorithm (ISTA) [2], two-step ISTA (TwISTA) [3] and fast ISTA (FISTA) [4, 5]. The ISTA is one of the most popular methods for solving such problems, and its advantage is its simplicity. However, the ISTA has also been recognized as a slow method. TWIST, as a faster method than the ISTA, also has good effectiveness. The FISTA has low complexity, fast convergence, and moderate recovery accuracy. It has been proven that the FISTA can also be faster than TWIST by several orders of magnitude. In addition, the alternating direction method of multipliers (ADMM) [6] and primal-dual hybrid gradient (PDHG) algorithm [7] are important methods. These methods solve inverse problems with nonsmooth regularizers that possess high computational efficiency. Unfortunately, these methods treat every attribute equally, which can result in inaccurate optimal solutions.

Recently, the iterative weighted shrinkage-thresholding algorithm (IWSTA) [8] and its variants have been developed to extend the range of practical applications. These methods often outperform unweighted algorithms. Nasser *et al.* [9] proposed the weighted FISTA (W-FISTA), which has higher estimation efficiency than the original FISTA but the same complexity. Candes *et al.* [10] proposed an iterative algorithm for reweighted  $L_1$  minimization (IRL1) to penalize better nonzero coefficients. This method solves the imbalance between the  $L_1$  and  $L_0$  norms. Chartrand *et al.* [11] proposed an iterative reweighted least-squares (IRLS) algorithm to attain an improved ability to recover sparse signals. IRLS and IRL1 are known for their state-of-the-art reconstruction rates for noiseless and noisy measurements. Foucart *et al.* [12] proposed a weighted method to obtain the solution of a system with the minimal  $L_q$  quasi-norm (WLQ). This method generalizes and improves the result obtained with the  $L_1$  norm. Wipf *et al.* [13] presented distinct detail-separable and nonseparable iterative reweighting algorithms and introduced mainly nonnegative sparse coding examples via reweighted  $L_1$  minimization (IRW) for solving linear inverse problems. However, the weights of these algorithms do not conform to the usual weight definition. Their sums are not one, and these weights are distributed over an enormous range. Such weights are difficult to explain and may lead to inaccurate results. Table 1 lists some IWTAs and their details.

This paper proposes a new IWSTA based on entropy regularization called the entropy-regularized IWSTA (ERIWSTA). This algorithm makes the weights easy to calculate, and it has good interpretability. Additionally, the iterative formula update process becomes simple. Experimental results obtained in synthetic CT image denoising tasks show that the proposed method is feasible and effective.

## Related work

We introduce the general model of an inverse problem and its solving method (the ISTA). For the convenience of description, we present the symbolic notations. Matrices are represented as capital letters. For a matrix  $A$ ,  $A_{*i}$ ,  $A_{i*}$  and  $A_{ij}$  denote the  $i$ -th column, the  $i$ -th row and the  $(i, j)$ -th element of  $A$ , respectively;  $\|\cdot\|_i$  represents the  $i$ -norm of a vector. All the vectors are column vectors unless transposed to a row vector by a prime superscript  $T$ .

The equation of an inverse problem in imaging can be expressed as follows [14]:

$$b = Ax + \epsilon, \quad (1)$$

where  $x \in R^n$  denotes an unknown image,  $b \in R^m$  denotes the given measurements,  $A \in R^{m \times n}$  is called the system matrix (usually,  $m \ll n$ ), and  $\epsilon$  is the unknown disturbance term or noise.

Table 1. Variants of weighted methods.

Algorithm	Reference	Weights	Min.	Max.	Regularizer
ISTA	Daubechies <i>et al.</i> [2]	1	1	1	$\sum_{i=1}^n  x_i $
IRL1	Candes <i>et al.</i> [10]	$\frac{1}{ x_i^{k-1}  + \delta}$	0	$\frac{1}{\delta}$	$\sum_{i=1}^n \log( x_i  + \delta)$
WLQ	Foucart <i>et al.</i> [12]	$\frac{1}{( x_i^{k-1}  + \delta)^{1-p}}$	0	$\frac{1}{\delta^{1-p}}$	$\sum_{i=1}^n ( x_i  + \delta)^p$
IRW	Wipf <i>et al.</i> [13]	$\frac{1 + ( x_i  + \delta)^{p+1}}{( x_i  + \delta)^{p+1}}$	0	1	$\sum_{i=1}^n ( x_i  - \frac{1}{(x_i + \delta)^p})$

<https://doi.org/10.1371/journal.pone.0311227.t001>

Note that the system is underdetermined. To reconstruct the original image  $x$ , researchers usually construct the following optimization problem [15]:

$$\min_x \frac{1}{2} \| Ax - b \|^2_2 \tag{2}$$

Generally, the least-squares method is used to solve the problem.

To suppress overfitting, some scholars [16–18] have introduced the  $L_0$  norm to Eq (2) as sparse prior information. Therefore, to obtain the solution of Eq (2), we must solve an optimization problem that minimizes the cost function [19]:

$$\min_x \frac{1}{2} \| Ax - b \|^2_2 + \beta \| x \|_0, \tag{3}$$

where  $\|x\|_0$ , the number of nonzero components of  $x$ , is the regularizer that imposes prior knowledge (sparsity);  $\beta > 0$ , the regularization coefficient, is a hyperparameter used to control the tradeoff between accuracy and sparsity. Eq (3) is an NP-hard optimization problem [20], which is highly discrete, so it is challenging to precisely solve this problem. Thus, we must seek an effective approximation solution for this problem. The  $L_1$  norm regularizer is introduced as a substitute for the  $L_0$  norm. That is [2],

$$\min_x \frac{1}{2} \| Ax - b \|^2_2 + \beta \| x \|_1. \tag{4}$$

This is a convex continuous optimization problem with a sole nondifferentiable point ( $x = 0$ ). The classic method for solving the problem is the ISTA proposed by Chambolle *et al.* [21, 22]. The ISTA updates  $x$  through the following shrinkage and soft thresholding operation in each iteration:

$$x^{k+1} = \text{soft}_{\beta t} [x^k - 2tA^T(Ax^k - b)], \tag{5}$$

where  $k$  represents the  $k$ -th iteration,  $t$  is an appropriate step size and *soft* is the soft threshold operation function. The *soft* function has the following form:

$$\text{soft}_\theta(x_i) = \text{sign}(x_i)(|x_i| - \theta), \tag{6}$$

where  $\text{sign}(x_i)$  is the sign function of  $x_i$  and  $\theta$  is the threshold.

Recently, the IWSTA has attracted more interest than the ISTA, as it outperforms its unweighted counterparts in most cases. In these methods, decision variables and weights are optimized in an alternating manner, or decision variables are optimized under heuristically chosen weights. Usually, the objective function of this type of algorithm has the following form

[13]:

$$\min_{x, w \geq 0} \frac{1}{2} \|Ax - b\|_2^2 + \beta \sum_{i=1}^n w_i |x_i|_1. \tag{7}$$

In this paper, we improve the second term in Eq (7) to obtain a variant. Then, we propose an iterative update law for the variant.

### Methodology

The main idea of IWSTA-type algorithms is to define a weight for each attribute based on the current iteration  $x^k$  and then use the defined weights to obtain a new  $x$ . In this section, we introduce an entropy regularizer to the cost function and obtain the following optimization model:

$$\begin{aligned} \min \quad & \Phi_{\beta, \gamma}(x, w) = F(x) + \beta G_\gamma(x, w) \\ \text{s. t.} \quad & w_i \geq 0, \sum_{i=1}^n w_i = 1 \\ \text{where} \quad & F(x) = \frac{1}{2} \|Ax - b\|_2^2 \\ & G_\gamma(x, w) = \sum_{i=1}^n w_i |x_i| + \gamma \sum_{i=1}^n w_i \ln w_i \end{aligned} \tag{8}$$

where  $\gamma \geq 0$  is a given hyperparameter, and the inspiration for adding a term with logarithms comes from the literature [23]. It is worth noting that if we do not use the entropy regularizer,  $w$  can easily be solved as  $w_i = 1$  when  $|x_i| = \min\{|x_1|, \dots, |x_n|\}$ , and otherwise is 0. The update rule can be easily explained by an example as

$$\begin{aligned} \min \{4, 1, 5\} &= \min 4w_1 + 1w_2 + 5w_3 \\ \text{s. t. } w_1, w_2, w_3 &\geq 0 \\ w_1 + w_2 + w_3 &= 1 \end{aligned}$$

The solutions are  $w_1 = 0, w_2 = 1$  and  $w_3 = 0$ , among which  $w_2$  corresponds to the minimum value of  $\{4, 1, 5\}$ . This is very similar to the weight computation in the k-means algorithm. This shows the simple fact that only one element of  $w$  is 1, and the others are 0, which is grossly incompatible with the actual problem. Then, we add the negative entropy of the weights to measure the uncertainty of the weights and stimulate more attributes to help with signal reconstruction because it is well known that  $\sum_{i=1}^n w_i \ln w_i$  is minimized in information theory when

$$w_1 = w_2 = \dots = w_n \tag{9}$$

We alternatively solve  $w$  and  $x$  in Eq (8) as follows.

### Update rule for $w$

To solve  $w$ , we introduce the Lagrange multiplier  $\lambda$  and obtain the following Lagrange function. For  $w$ ,  $F(x)$  is a constant, so we only construct a Lagrange function on  $G(x)$ , which can be

expressed as follows:

$$L(w, \lambda) = G_\gamma(x, w) + \lambda(\sum_{i=1}^n w_i - 1). \tag{10}$$

We set the partial derivatives of  $L(w, \lambda)$  with respect to  $w_i$  and  $\lambda$  to zero and then obtain the following two equations:

$$\frac{\partial L(w, \lambda)}{\partial w_i} = |x_i| + \gamma(1 + \ln w_i) + \lambda = 0, \tag{11}$$

$$\frac{\partial L(w, \lambda)}{\partial \lambda} = \sum_{i=1}^n w_i - 1 = 0. \tag{12}$$

From Eq (11), we know that

$$w_i = \exp\left(-\frac{\lambda}{\gamma}\right) \exp\left(-\frac{|x_i|}{\gamma}\right) \exp(-1). \tag{13}$$

Substituting Eq (13) into Eq (12), we have

$$\sum_{i=1}^n w_i = \exp\left(-\frac{\lambda}{\gamma}\right) \sum_{i=1}^n \exp\left(-\frac{|x_i|}{\gamma}\right) \exp(-1) = 1. \tag{14}$$

It follows that

$$\exp\left(-\frac{\lambda}{\gamma}\right) = \frac{\exp^1}{\sum_{i=1}^n \exp\left(-\frac{|x_i|}{\gamma}\right)}. \tag{15}$$

Substituting this expression into Eq (13), we obtain that

$$w_i = \frac{\exp\left(-\frac{|x_i|}{\gamma} + 1\right)}{\sum_{i=1}^n \exp\left(-\frac{|x_i|}{\gamma}\right)}. \tag{16}$$

Such weights certainly satisfy the constraints that  $w_i \geq 0$  and  $\sum_{i=1}^n w_i = 1$ .

### Update rule for $x$

Inspired by the work concerning the ISTA, we adopt a similar approach for the iterative update process of  $x$ . The construction of a majorization is an important step toward obtaining the update rule.

**Definition 0.1** (Majorization) Assume  $\Psi(x)$  is an  $n$ -dimensional real valued function about vector  $x$ , we can denote  $\psi^k(x; x^k)$  as a majorization for  $\Psi(x)$  at  $x^k$  (fixed) if  $\psi^k(x^k) = \Psi(x^k)$  and  $\psi^k(x) \geq \Psi(x)$ .

Clearly,  $\Psi(x)$  is nonincreasing under the update rule  $x^{k+1} = \min_x \psi(x|x^k)$  because

$$\Psi(x^{k+1}) \leq \psi(x^{k+1}|x^k) \leq \psi(x^k|x^k) = \Psi(x^k). \tag{17}$$

Then, we can construct a majorization for  $F(x)$ .

**Remark 0.1.** Conditions for constructing surrogate function: A basic principle of optimization algorithm is to construct an easily minimized function  $\psi^k(x; x^k)$  to replace the original function  $\Psi$

(x). Then calculate the minimum value of the function  $\psi^k(x; x^k)$ , and use the minimum value point as the new iteration point (i.e.  $x^{k+1} = \operatorname{argmin} \psi^k(x; x^k)$ ). By continuously repeating the two steps of constructing the surrogate function and finding the minimum value of the surrogate function, an estimated sequence  $\psi^k$  of  $\Psi$  can be obtained. The estimated sequence  $\psi^k$  makes  $\Psi(x^k)$  monotonically decreasing when the surrogate function satisfies the following conditions [24, 25]:

$$\psi^k(x^k) = \Psi(x^k) \tag{18a}$$

$$\psi^k(x) \geq \Psi(x), x \in D \tag{18b}$$

**Proposition 0.1** Notably,  $F(x)$  is a Lipschitz-continuous and differentiable convex function, which has a majorization function at the fixed current iteration  $x^k$  as follows:

$$f(x, x^k) = F(x^k) + [\nabla F(x^k)]^T(x - x^k) + \frac{L}{2} \|x - x^k\|_2^2, \tag{19}$$

where  $L$  is larger than or equal to the maximum eigenvalue of  $A^T A$ .

*Proof.* It is well known that

$$F(x) = \frac{1}{2} \|Ax - b\|_2^2 = F(x^k) + [\nabla F(x^k)]^T(x - x^k) + \frac{1}{2}(x - x^k)^T A^T A(x - x^k) \tag{20}$$

Comparing  $F(x)$  with  $F(x, x^k)$ , we find that their last items are different. By performing singular value decomposition (SVD) on a symmetric definite matrix, we know that  $A^T A = Q^T \Sigma Q$ , in which  $Q$  is an orthogonal matrix consisting of all eigenvectors and  $\Sigma$  is a diagonal consisting of all eigenvalues. Let  $z = x - x^k$ ; then,

$$z^T(A^T A)z = z^T Q^T \Sigma Q z \leq L \|Qz\|_2^2 = L \|z\|_2^2 \tag{21}$$

It is also certain that  $z^T A^T A z = L \|z\|_2^2 = 0$  if  $x = x^k$ . Thus, the proof is established. Now, we obtain the majorization for the cost function  $\Phi(x, w)$  on  $x$ .

$$\phi(x, x^k) = f(x, x^k) + \beta G_i(x, w), \tag{22}$$

which can be reorganized as:

$$\begin{aligned} \phi(x, x^k) &= \frac{L}{2} \|x - \left[ x^k - \frac{1}{L} \nabla F(x^k) \right]\|_2^2 + \beta G_i(x, w) \\ &= \sum_{i=1}^n \left\{ \frac{L}{2} \|x_i - \left[ x^k - \frac{1}{L} \nabla F(x^k) \right]_i\|_2^2 + \beta w_i |x_i| \right\} + \text{constant}. \end{aligned} \tag{23}$$

Then,

$$x^{k+1} = \operatorname{argmin} \sum_{i=1}^n \frac{L}{2} \{x_i - \left[ x^k - \frac{1}{L} \nabla F(x^k) \right]_i\}_2^2 + \beta w_i |x_i|. \tag{24}$$

where  $(\cdot)_i$  denotes the  $i$ -th element of vector  $(\cdot)$ .

Let

$$r^k := x^k - \frac{1}{L} \nabla F(x^k), \tag{25}$$

so we know that

$$x^{k+1} = \operatorname{argmin} \sum_{i=1}^n \frac{L}{2} (x_i - r_i^k)^2 + \beta w_i |x_i|. \tag{26}$$

Let

$$g(x_i) = \frac{L}{2} (x_i - r_i^k)^2 + \beta w_i |x_i|. \tag{27}$$

Let  $x_i^k$  be the value of  $x_i$  when the partial derivative  $\frac{\partial g(x_i)}{\partial x_i}$  is 0, that is:

$$\frac{\partial g(x_i)}{\partial x_i} = L(x_i^k - r_i^k) + \beta w_i \operatorname{sign}(x_i^k) = 0, \tag{28}$$

By transferring terms, we can obtain the following equation:

$$r_i^k = x_i^k + \frac{1}{L} \beta w_i \operatorname{sign}(x_i^k) \tag{29}$$

That is,

$$x_i^{(k+1)} = x_i^k + \frac{1}{L} \Delta F(x_i^k) + \frac{1}{L} \beta w_i \operatorname{sign}(x_i^k) \tag{30}$$

Due to  $t = \frac{1}{L}$  and  $\Delta F(x) = 2A^T(Ax - b)$ , the iterative formula can be easily obtained as follows:

$$x_i^{k+1} = \operatorname{soft}_{\beta t w_i}(r_i^k) = \operatorname{soft}_{\beta t w_i}[x_i^k - 2tA^T(Ax_i^k - b)] \tag{31}$$

Therefore, the update of  $x$  is completed.

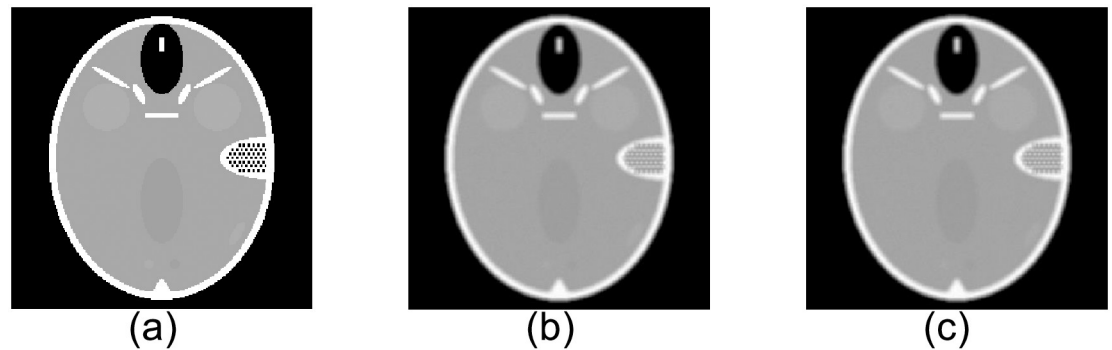
## Experiments

### Experimental description

We use a simulated CT dataset and a real PET dataset to evaluate the performance of the ERWISTA. The simulated CT dataset consists of a Shepp-Logan phantom with  $256 \times 256$  pixels. The use of phantom data brings many advantages, including the fact that we have accurate a priori knowledge of the pixel values and the freedom to add noise to them as needed. We blur the image ( $b$ ) using a uniform  $5 \times 5$  kernel (applied by the `fspecial` MATLAB function) and then add Gaussian noise with a mean of 0 and variances of  $10^{-2}$  and  $10^{-3}$ . The original and blurred and noisy images are shown in Fig 1. The real dataset comes from the cooperating hospital. The dataset includes low-dose and normal-dose PET images with size of  $256 \times 256$ . All experiments are performed on an HP computer with a 2.5 GHz Intel(R) Core(TM) i7-4710MQ CPU with 12 GB of memory using MATLAB R2019a for coding.

### Evaluation standard

This paper uses the mean absolute error (MAE) to measure the similarity between the reconstructed and true images. The value of the MAE is calculated by taking the average of the squared differences between the restored pixel values and the true pixel values. Let  $x$  represent the ground truth,  $\hat{x}$  represent the reconstructed image and  $N$  denote the number of voxels.



**Fig 1. Original and noisy head phantom images.** (a) Head phantom with  $256 \times 256$  pixels; (b) and (c) blurred images with  $5 \times 5$  uniform kernels and additive Gaussian noise with variances of  $\sigma = 10^{-2}$  and  $\sigma = 10^{-3}$ , respectively.

<https://doi.org/10.1371/journal.pone.0311227.g001>

The MAE is defined as follows:

$$MAE = \frac{1}{N} \|x - \hat{x}\|_1. \quad (32)$$

Generally, lower MAE values indicate better reconstructed image quality.

### Experimental results of simulated dataset

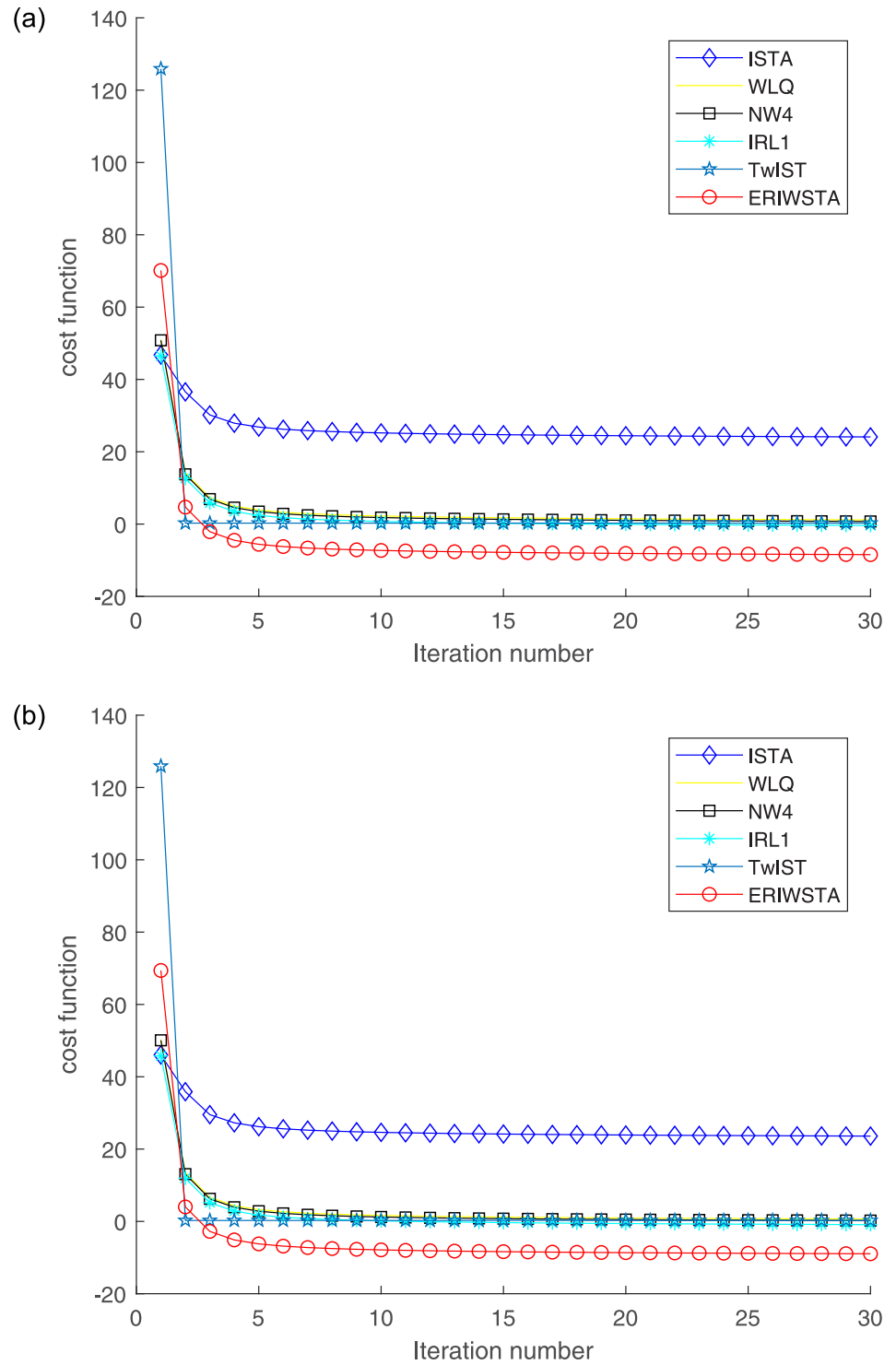
To illustrate the performance of the proposed method, we provide some visual results. Fig 2 displays the cost function curves produced by six algorithms during training. They are the ISTA, WLQ, IRW, IRL1, TwIST and the ERIWSTA. As is well known, we should not compare the values of different cost functions. However, we can visually compare their convergence speeds, in which an algorithm shows faster convergence when the corresponding curve becomes flatter within fewer iterations. We find that the six algorithms tend to converge after approximately several iterations and that the proposed algorithm has a fast convergence speed. The ERIWSTA arrives at the stable status early.

Fig 3 shows the MAE curves of the six algorithms versus the number of iterations. Sub-graphs (a) and (b) denote the curves produced in the cases with Gaussian noise levels of  $\sigma = 10^{-2}$  and  $\sigma = 10^{-3}$ , respectively. We can observe that each algorithm has a small MAE value, indicating that all methods have considerable denoising capabilities. However, compared with the other algorithms, the ERIWSTA obtains the minimum MAE value after each iteration to obtain a clearer denoised image by observing an enlarged detail image.

Figs 4 and 5 show the denoising results obtained by six algorithms under the given noise level. As seen, the image reconstructed by the ERIWSTA has a better noise removal effect than those of the other algorithms, and it can more accurately recover the edge information and texture features of brain images.

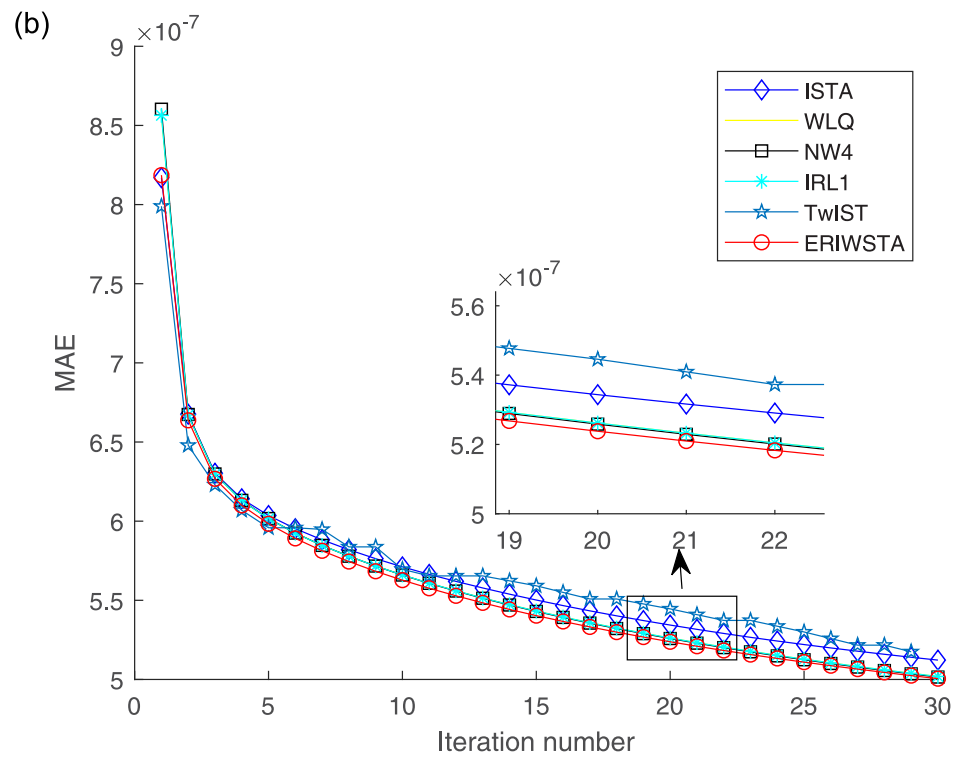
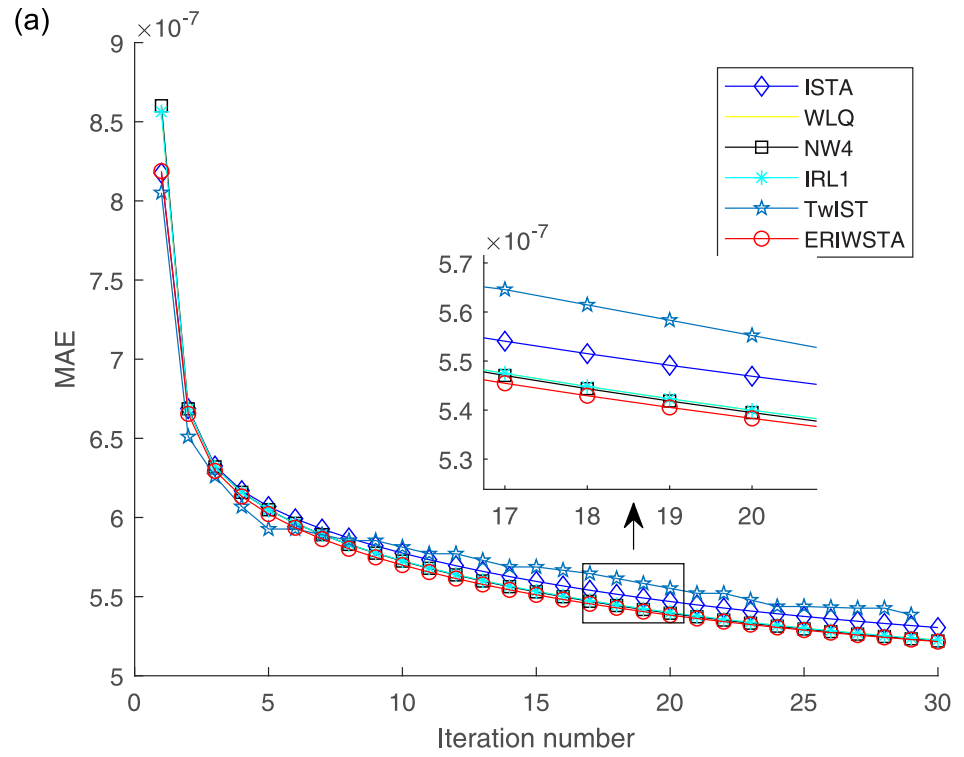
In addition, we further verify the effectiveness of the proposed algorithm. We select the same row and column of four reconstructed images for one slice and compare them with the corresponding rows and columns of the real image by their pixel values. Figs 6 and 7 show the comparisons among the horizontal and vertical center profiles of the restored images, respectively. IRL1 yields significant deviations from the true values in both the horizontal and vertical regions. The ISTA, IRW, WLQ and TwIST recover well in the horizontal region but sometimes have abnormal states in the vertical region. These methods cannot guarantee a stable denoising effect. Compared with the other algorithms, the ERIWSTA does not guarantee the lowest denoising errors in all intervals, but it can fit the true values more accurately overall.





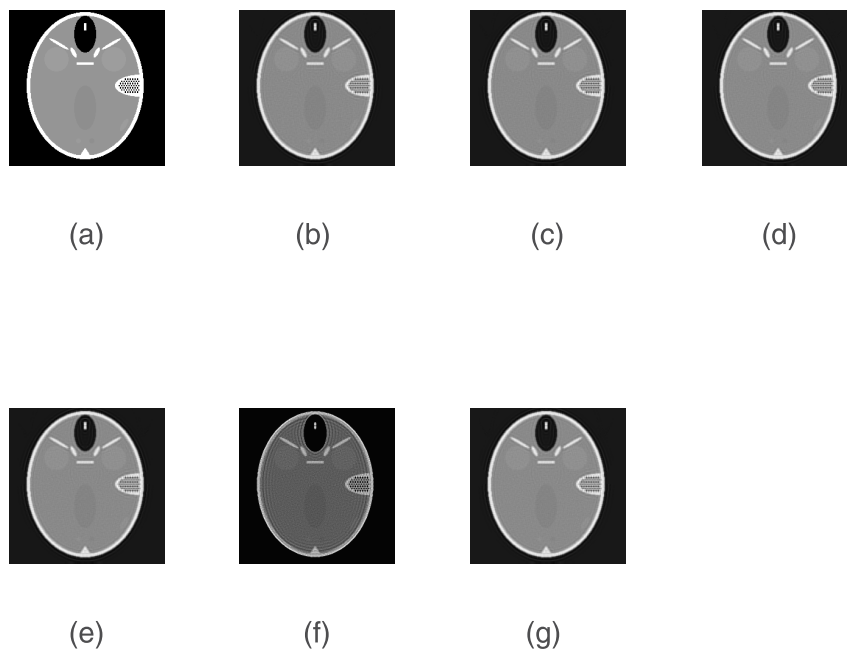
**Fig 2. The cost function  $\Phi(x, w)$  versus the number of iterations for different Gaussian noise levels: (a)  $\sigma = 10^{-2}$  and (b)  $\sigma = 10^{-3}$ .**

<https://doi.org/10.1371/journal.pone.0311227.g002>



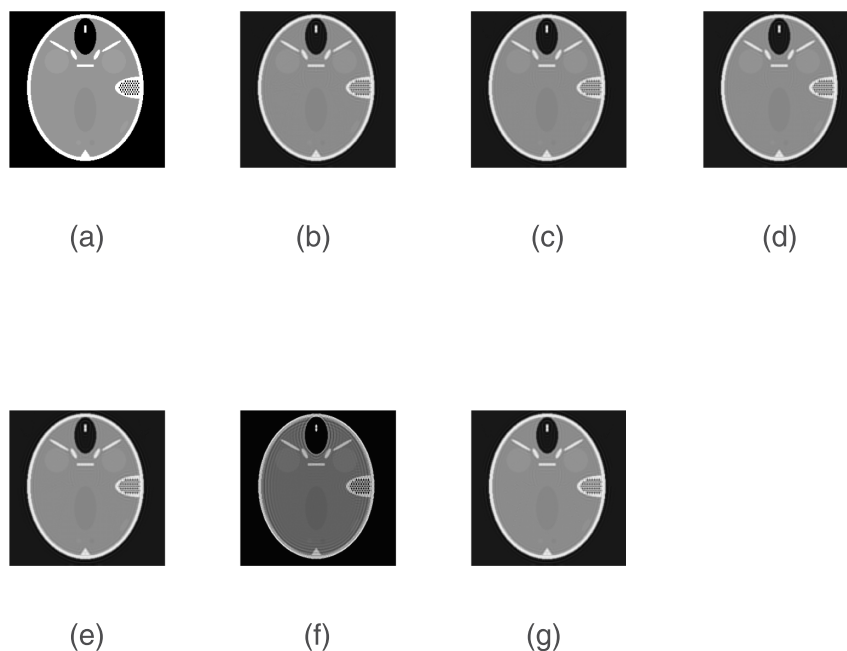
**Fig 3. MAE versus the number of iterations for different Gaussian noise levels: (a)  $\sigma = 10^{-2}$  and (b)  $\sigma = 10^{-3}$ .**

<https://doi.org/10.1371/journal.pone.0311227.g003>



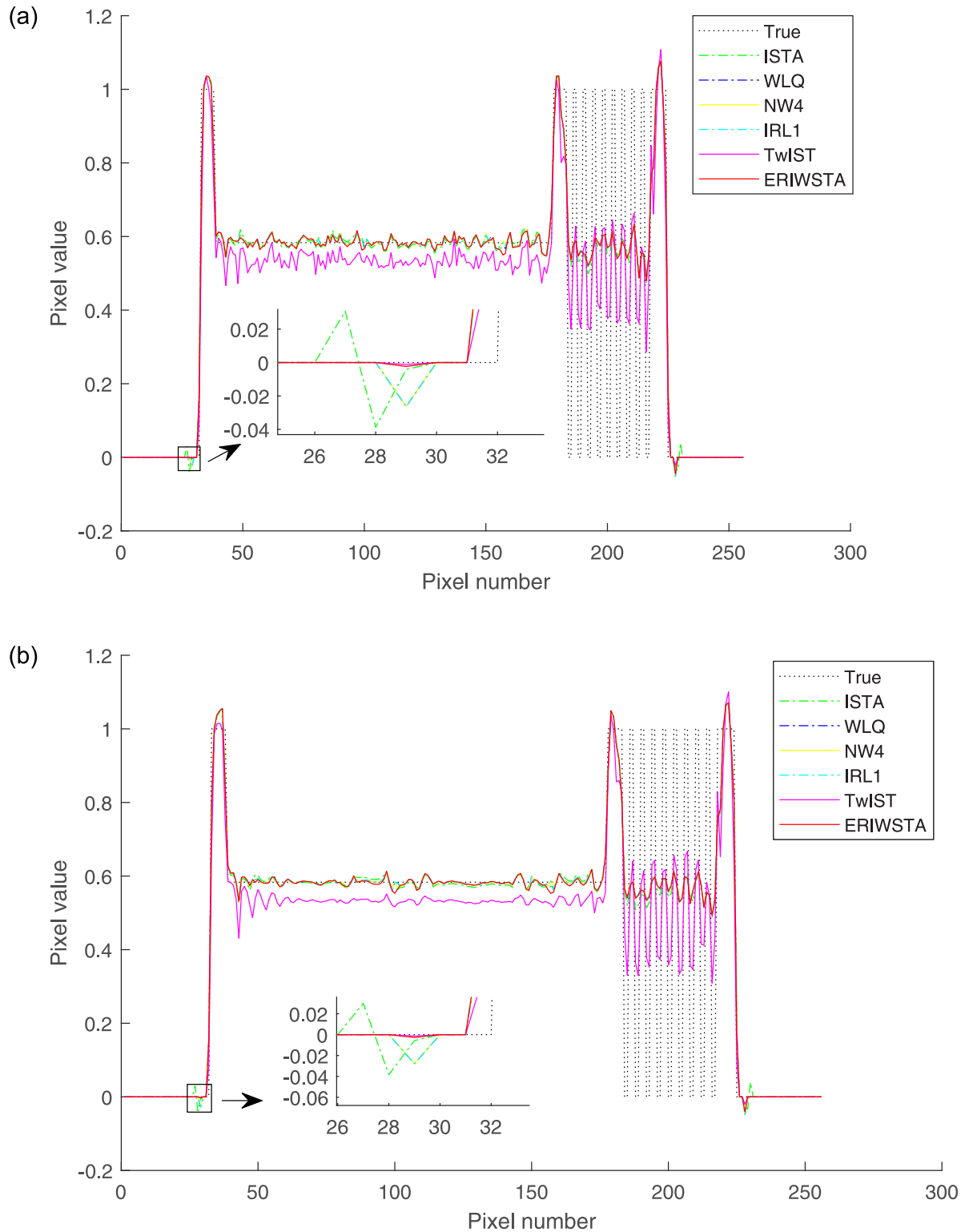
**Fig 4.** The denoising results yielded by different algorithms on a dataset with Gaussian noise at a level of  $\sigma = 10^{-2}$  after 30 iterations. (a) denotes the original image, and (b)–(g) denote the denoised images produced by the ISTA, WLQ, IRW, IRL1, TwIST and the ERIWSTA, respectively.

<https://doi.org/10.1371/journal.pone.0311227.g004>



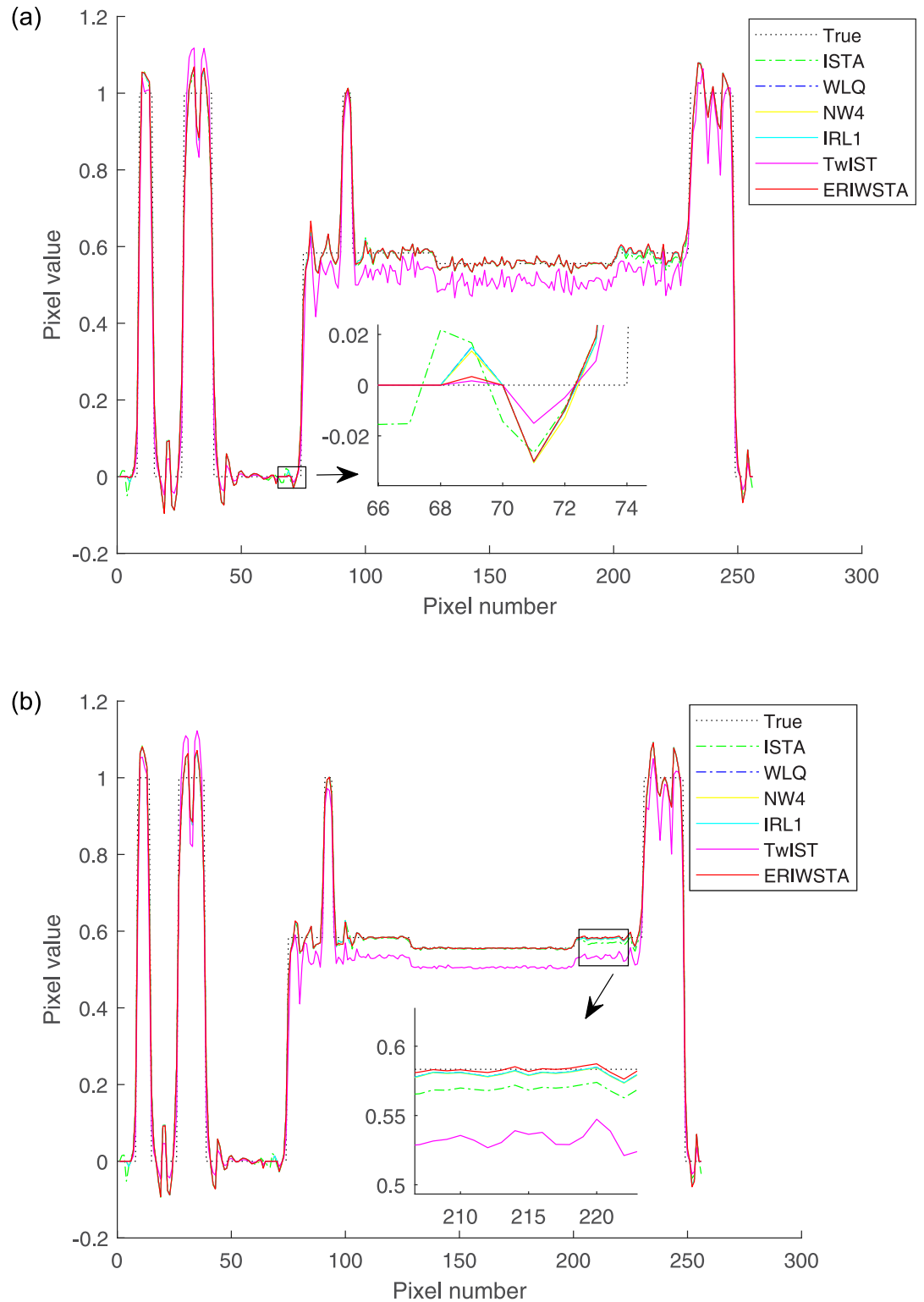
**Fig 5.** The denoising results yielded by different algorithms on a dataset with Gaussian noise at a level of  $\sigma = 10^{-3}$  after 30 iterations. (a) denotes the original image, and (b)–(g) denote the denoised images produced by the ISTA, WLQ, IRW, IRL1, TwIST and the ERIWSTA, respectively.

<https://doi.org/10.1371/journal.pone.0311227.g005>



**Fig 6. Horizontal central profiles produced for the 128-th row of the restored images with different Gaussian noise levels: (a)  $\sigma = 10^{-2}$  and (b)  $\sigma = 10^{-3}$ .**

<https://doi.org/10.1371/journal.pone.0311227.g006>



**Fig 7. Vertical central profiles produced for the 128-th column of the restored images with different Gaussian noise levels: (a)  $\sigma = 10^{-2}$  and (b)  $\sigma = 10^{-3}$ .**

<https://doi.org/10.1371/journal.pone.0311227.g007>

The ERIWSTA also has the best stability and robustness, can guarantee good denoising effects in areas with minor pixel value variations, and can track the true profile more accurately.

**Hyperparameter selection.** We conduct hyperparameter experiments to select the optimal penalty hyperparameter  $\beta$  and the optimal entropy-weighted hyperparameter  $\gamma$ . We provide a parameter range from  $10^{-10}$  to  $10^{10}$  and choose the MAE as the evaluation index. The number of experimental iterations is 100, and we plot a three-dimensional line graph for the ERIWSTA with the vertical axis representing the MAE values corresponding to different hyperparameters, as shown in Fig 8. The optimal values of  $\beta$  and  $\gamma$  can be chosen from wide ranges, and the ERIWSTA has stable lower MAE values for both, which also shows the excellent robustness of the ERIWSTA. In addition, to ensure the validity of the experimental results, we choose the optimal  $\beta$  and  $\gamma$  values for all the compared algorithms. For details, Tables 2 and 3 show the optimal values of all the algorithms'  $\beta$  and  $\gamma$  parameters when conducting experiments on the datasets with noise levels of  $\sigma = 10^{-2}$  and  $\sigma = 10^{-3}$ , respectively. An interesting observation is that, regardless of whether low or high noise levels are used, the restoration accuracy of our algorithm is always better than that of the other approaches.

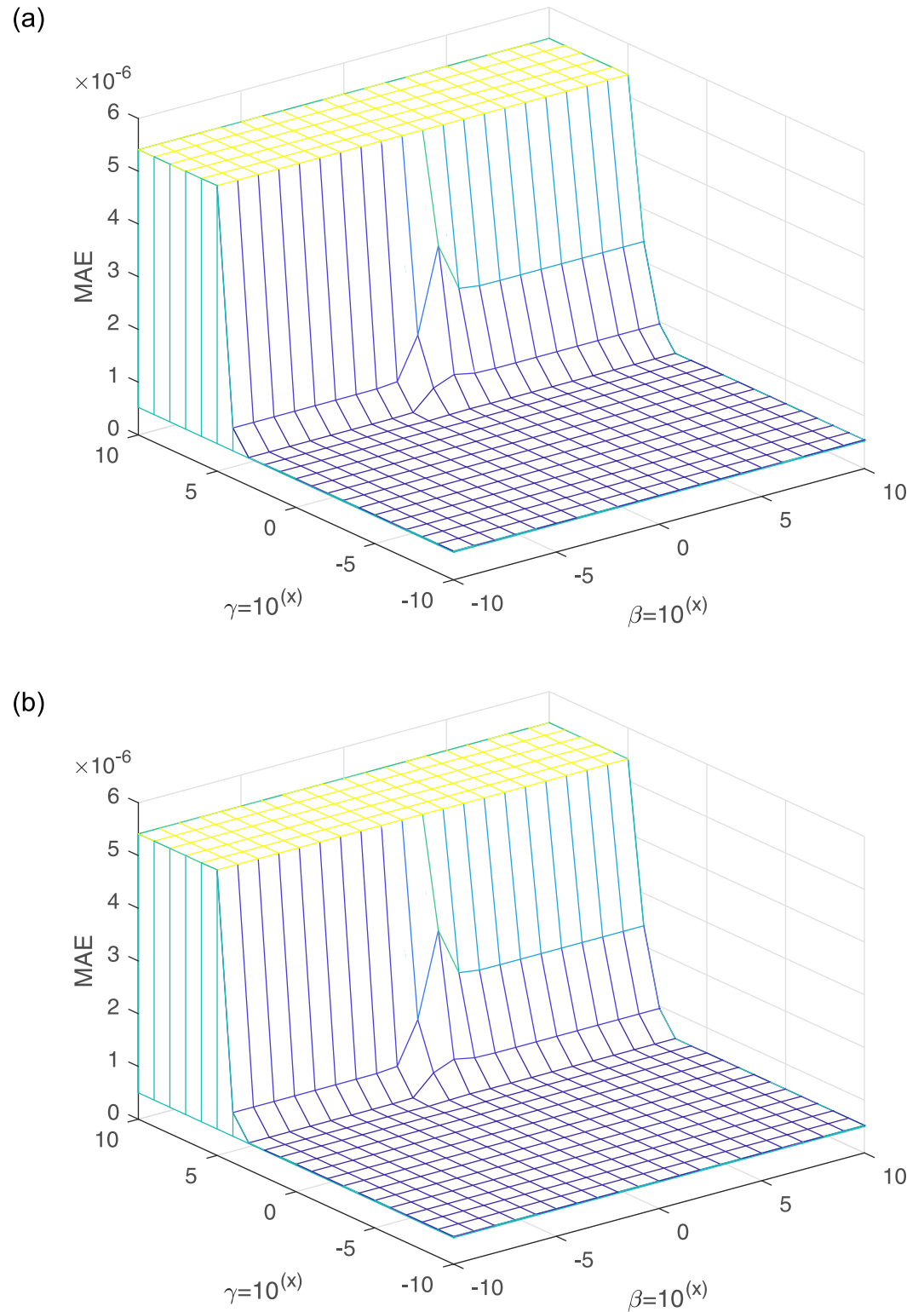
### Experimental results of real PET dataset

To make the results more convincing, we added the experiment result of real PET image. Fig 9 shows the denoised images and the second to sixth columns of Fig 9 are the images reconstructed using ISTA, WLQ, IRW, IRL1 and the proposed method, respectively. It is seen that all methods have considerable denoising ability, and the proposed algorithm has a better ability to denoise. Fig 10 shows the comparisons among the horizontal and vertical center profiles of the restored PET images, respectively. The ISTA, IRW and WLQ recover well in the horizontal region but sometimes have abnormal states in the vertical region. Compared with the other algorithms, the ERIWSTA does not guarantee the lowest denoising errors in all intervals, but it can fit the true values more accurately overall. The ERIWSTA also has the best stability and robustness.

### Conclusions

This paper proposes a new IWSTA for solving linear inverse problems based on entropy regularization. The main innovation of the algorithm is its introduction of an entropy regularization term in the loss function of the classic ISTA. The weights of the new algorithm are easy to calculate and have good interpretability. In addition, the iterative process of updating the formula becomes simple. Finally, we demonstrate the algorithm's effectiveness in CT and PET data denoising experiments.

Although the experimental results show that the proposed method effectively removes noise, this work still needs to be improved. The hyperparameters of all algorithms in this paper are obtained by manually adjusting them according to the resulting imaging quality. This selection method requires cyclic experiments for the selected parameter intervals to choose the optimal parameters. This method could be more efficient and select only the optimal parameters in the given interval. Therefore, in the future, we will study a method for adaptively setting the parameters to accurately select the optimal parameters of each algorithm, thus ensuring the algorithm's effectiveness and further improving its noise reduction ability. In addition, the data used in this experiment are phantom data. In algorithmic analyses, the use of simulated data provides good properties. It is easy to add noise and is conducive to the theoretical study of an algorithm's effectiveness. However, the proposed algorithm still must be verified in more practice. Therefore, verifying the algorithm's effectiveness on more patients' clinical data and effectively improving the algorithm based on an actual application would be meaningful.



**Fig 8. 3D profiles of  $\beta$  and  $\lambda$  with respect to the MAE for different Gaussian noise levels: (a)  $\sigma = 10^{-2}$  and (b)  $\sigma = 10^{-3}$ .**

<https://doi.org/10.1371/journal.pone.0311227.g008>

Table 2. The optimal MAE values and corresponding hyperparameters (Gaussian noise with  $\sigma = 10^{-2}$ ).

Termed	$\beta$	$\gamma$	$\delta$	MAE
ISTA	$10^{-3}$	–	–	$5.312077 * 10^{-7}$
WLQ	$10^{-5}$	$10^{-10}$	$10^{-3}$	$5.228672 * 10^{-7}$
IRW	$10^{-5}$	$10^{-2}$	$10^{-3}$	$5.410231 * 10^{-7}$
IRL1	$10^{-5}$	–	$10^{-3}$	$5.228672 * 10^{-7}$
TwIST	$10^{-5}$	–	$10^{-5}$	$5.333443 * 10^{-7}$
ERIWSTA	$10^2$	$10^{-2}$	–	$5.218246 * 10^{-7}$

<https://doi.org/10.1371/journal.pone.0311227.t002>

Table 3. The optimal MAE values and corresponding hyperparameters (Gaussian noise with  $\sigma = 10^{-3}$ ).

Termed	$\beta$	$\gamma$	$\delta$	MAE
ISTA	$10^{-3}$	–	–	$5.122013 * 10^{-7}$
WLQ	$10^{-5}$	$10^{-5}$	$10^{-3}$	$5.018339 * 10^{-7}$
IRW	$10^{-5}$	$10^{-2}$	$10^{-3}$	$5.410231 * 10^{-7}$
IRL1	$10^{-5}$	–	$10^{-3}$	$5.018340 * 10^{-7}$
TwIST	$10^{-5}$	–	$10^{-5}$	$5.311041 * 10^{-7}$
ERIWSTA	$10^2$	$10^{-2}$	–	$5.005524 * 10^{-7}$

<https://doi.org/10.1371/journal.pone.0311227.t003>

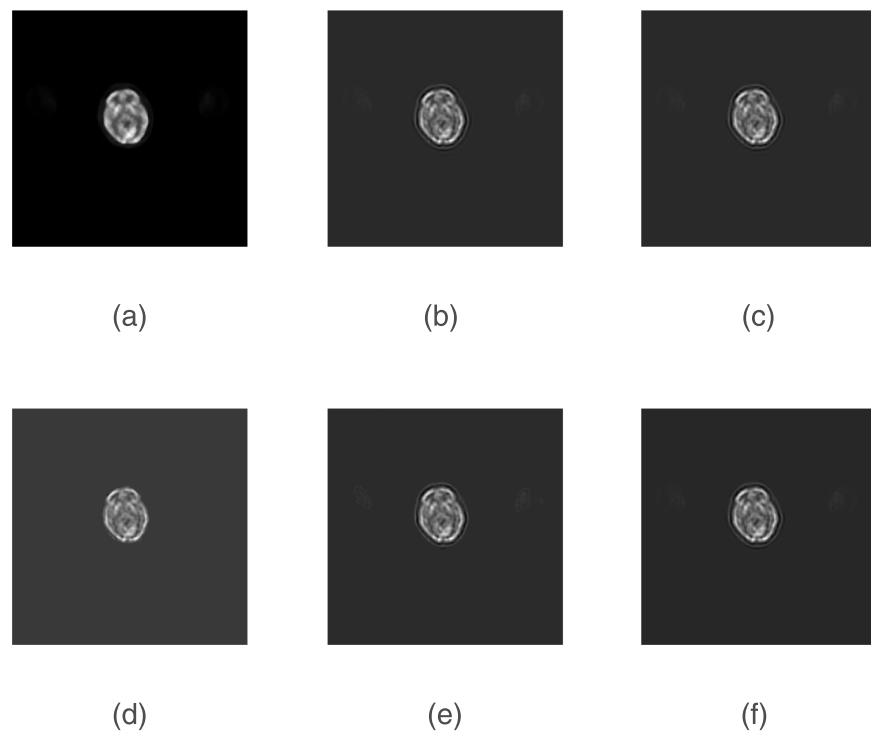
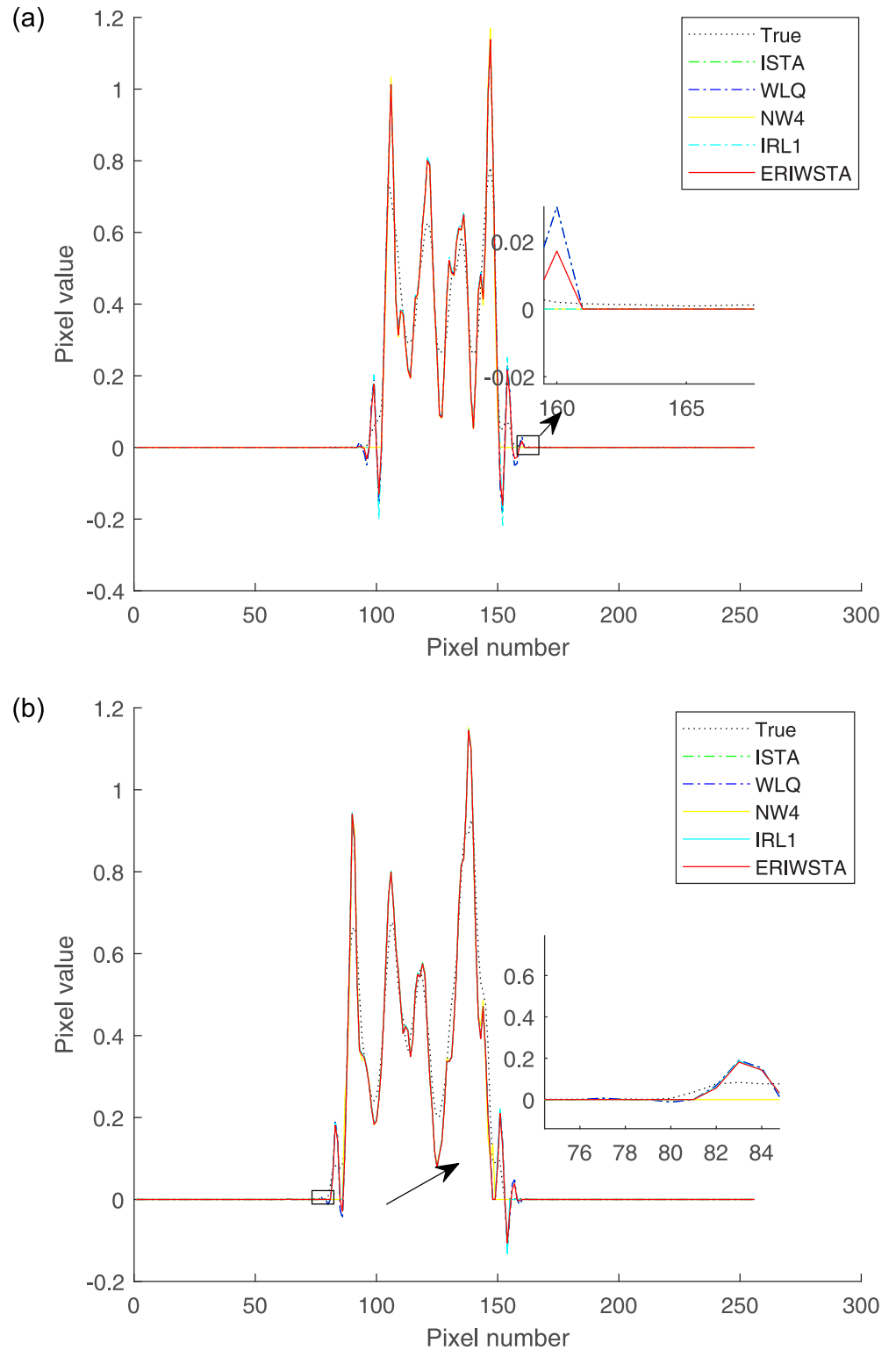


Fig 9. The denoising results yielded by different algorithms. (a) denotes the ground truth, and (b)–(f) denote the denoised images produced by the ISTA, WLQ, IRW, IRL1 and the ERIWSTA, respectively.

<https://doi.org/10.1371/journal.pone.0311227.g009>





**Fig 10.** The comparison between PET images reconstructed by different algorithms and the label images focuses on specific pixel profiles: (a) shows the pixel values along the 128th row of the corresponding images. (b) shows the pixel values along the 128th column of the corresponding images.

<https://doi.org/10.1371/journal.pone.0311227.g010>

## Supporting information

### S1 Graphical abstract.

(PDF)

## Author Contributions

**Data curation:** Bingxue Wu.

**Funding acquisition:** Yueyang Teng.

**Investigation:** Yueyang Teng.

**Methodology:** Limin Ma, Yueyang Teng.

**Supervision:** Yudong Yao, Yueyang Teng.

**Validation:** Limin Ma.

**Writing – original draft:** Limin Ma, Bingxue Wu.

## References

1. Xiang JX, Dong YG, Yang YJ. FISTA-net: Learning a fast iterative shrinkage thresholding network for inverse problems in imaging. *IEEE Transactions on Medical Imaging*. 2021 Jan; 40(5):1329–1339. <https://doi.org/10.1109/TMI.2021.3054167> PMID: 33493113
2. Daubechies I, Defrise M, Demol C. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*. 2004 Aug; 57(11):1413–1457. <https://doi.org/10.1002/cpa.20042>
3. Bioucas-Dias JM, Figueiredo MA. A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image Processing*. 2007 Nov; 16(12):2992–3004. <https://doi.org/10.1109/TIP.2007.909319> PMID: 18092598
4. Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*. 2009 Mar; 2(1):183–202. <https://doi.org/10.1137/080716542>
5. Vonesch C, Unser M. A fast iterative thresholding algorithm for wavelet-regularized deconvolution. In: *Wavelets XII*. SPIE, 2007: 135–139.
6. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*. 2011 Jul; 3(1):1–122. <https://doi.org/10.1561/2200000016>
7. Chambolle A, Pock T. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*. 2011 Dec; 40:120–145. <https://doi.org/10.1007/s10851-010-0251-1>
8. Zhu WX, Huang ZL, Chen JL, Peng Z. Iterative weighted thresholding method for sparse solution of underdetermined linear equations. *Sci. China Math*. 2020 Jul; 64(3):639–664. <https://doi.org/10.1007/s11425-018-9467-7>
9. Nasser A, Elsabrouty M, Muta O. Weighted fast iterative shrinkage thresholding for 3D massive MIMO channel estimation. In: *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Montreal, QC, Canada. IEEE, 2017: 1–5.
10. Candes EJ, Wakin MB, Boyd SP. Enhancing sparsity by reweighted  $L_1$  minimization. *Journal of Fourier analysis and applications*. 2008 Oct; 14:877–905. <https://doi.org/10.1007/s00041-008-9045-x>
11. Chartrand R, Yin W. Iteratively reweighted algorithms for compressive sensing. In: *2008 IEEE international conference on acoustics, speech and signal processing*, Las Vegas, NV, USA. IEEE, 2008: 3869–3872.
12. Foucart S, Lai MJ. Sparsest solutions of undetermined linear systems via  $L_q$ -minimization for  $0 < q \leq 1$ . *Applied and Computational Harmonic Analysis*. 2009 May; 26(3):395–407. <https://doi.org/10.1016/j.acha.2008.09.001>
13. Wipf D, Nagarajan S. Iterative reweighted  $L_1$  and  $L_2$  methods for finding sparse solutions. *IEEE Journal of Selected Topics in Signal Processing*. 2010; 4(2):317–329. <https://doi.org/10.1109/JSTSP.2010.2042413>

14. Chung JL, Nagy JG. An efficient iterative approach for large-scale separable nonlinear inverse problems. *SIAM Journal on Scientific Computing*. 2010; 31(6):4654–4674. <https://doi.org/10.1137/080732213>
15. Donoho DL. Compressed sensing. *IEEE Transactions on Information Theory*. 2006 Apr; 52(4): 1289–1306. <https://doi.org/10.1109/TIT.2006.871582>
16. Elad M, Figueiredo MT, Ma Y. On the role of sparse and redundant representations in image processing. *Proceedings of the IEEE*. 2010 Feb; 98(6):972–982. <https://doi.org/10.1109/JPROC.2009.2037655>
17. Wright J, Ma Y, Mairal J, Sapiro G, Huang TS, Yan SC. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*. 2010 Jun; 98(6):1031–1044. <https://doi.org/10.1109/JPROC.2010.2044470>
18. Gong MG, Liu J, Li H, Cai Q, Su LZ. A multiobjective sparse feature learning model for deep neural networks. *IEEE transactions on neural networks and learning systems*. 2015 Aug; 26(12):3263–3277. <https://doi.org/10.1109/TNNLS.2015.2469673> PMID: 26340790
19. Barbie D, Lucibello C, Saglietti L, Krzakala F, Zdeborova L. Compressed sensing with l0-norm: statistical physics analysis and algorithms for signal recovery. *arXiv:2304.12127v1*. 2023.
20. Natarajan BK. Sparse approximate solutions to linear systems. *SIAM journal on computing*. 1995; 24(2):227–234. <https://doi.org/10.1137/S0097539792240406>
21. Figueiredo MA, Nowak RD. An EM algorithm for wavelet-based image restoration. *IEEE Transactions on Image Processing*. 2003 Aug; 12(8):906–916. <https://doi.org/10.1109/TIP.2003.814255> PMID: 18237964
22. Chambolle A, Devore RA, Lee NY, Lucier BJ. Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Transactions on Image Processing*. 1998 Mar; 7(3):319–335. <https://doi.org/10.1109/83.661182>
23. James MQ. Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Oakland, CA, USA. 1967: 281–297.
24. Teng YY, Zhang YN, Li H, Kang Y. A convergent non-negative deconvolution algorithm with Tikhonov regularization. *Inverse Problems*. 2015 Mar; 31(3):035002. <https://doi.org/10.1088/0266-5611/31/3/035002>
25. Jacobson MW, Fessler JA. An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms. *IEEE Transactions on Image Processing*. 2007 Oct; 16(10):2411–2422. <https://doi.org/10.1109/tip.2007.904387> PMID: 17926925