

IMPULSE NOISE REMOVAL BY L1 WEIGHTED NUCLEAR NORM MINIMIZATION*

Jian Lu

*Shenzhen Key Laboratory of Advanced Machine Learning and Applications,
College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, China
Guangdong Key Laboratory of Intelligent Information Processing,
Pazhou Lab, Guangzhou 510335, China
Email: jianlu@szu.edu.cn*

Yuting Ye

*Shenzhen Key Laboratory of Advanced Machine Learning and Applications,
College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, China
Email: 1084541754@qq.com*

Yiqiu Dong¹⁾

*Department of Applied Mathematics and Computer Science,
Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
Email: yido@dtu.dk*

Xiaoxia Liu²⁾ and Yuru Zou

*Shenzhen Key Laboratory of Advanced Machine Learning and Applications,
College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, China
Email: xiaoxia_liu_math@outlook.com, yuruzou@szu.edu.cn*

Abstract

In recent years, the nuclear norm minimization (NNM) as a convex relaxation of the rank minimization has attracted great research interest. By assigning different weights to singular values, the weighted nuclear norm minimization (WNNM) has been utilized in many applications. However, most of the work on WNNM is combined with the l^2 -data-fidelity term, which is under additive Gaussian noise assumption. In this paper, we introduce the L1-WNNM model, which incorporates the l^1 -data-fidelity term and the regularization from WNNM. We apply the alternating direction method of multipliers (ADMM) to solve the non-convex minimization problem in this model. We exploit the low rank prior on the patch matrices extracted based on the image non-local self-similarity and apply the L1-WNNM model on patch matrices to restore the image corrupted by impulse noise. Numerical results show that our method can effectively remove impulse noise.

Mathematics subject classification: 68U10, 94A08, 90C26, 15A03, 46N10.

Key words: Image denoising, Weighted nuclear norm minimization, l^1 -data-fidelity term, Low rank analysis, Impulse noise.

1. Introduction

With the rapid development of technologies in image processing, many effective image denoising methods have been proposed based on the low rank matrix approximation (LRMA)

* Received June 24, 2021 / Revised version received October 8, 2021 / Accepted January 27, 2022 /
Published online September 29, 2022 /

¹⁾ Corresponding authors.

²⁾ Corresponding authors.

that aims to restore a low rank matrix from its noisy or incomplete observation, e.g., in [1–3]. Generally, LRMA methods can be sorted into two categories: the nuclear norm minimization (NNM) methods, see [4–7], and the low rank matrix factorization (LRMF) methods, see [1, 2, 8]. In this paper, we focus on the first type. The NNM methods aim to seek a low rank solution by minimizing the nuclear norm and the work in [3] shows that many NNM-based problems can be solved via the nuclear norm proximal (NNP) that is defined as

$$\hat{X} = \arg \min_{X \in \mathbb{R}^{m \times n}} \frac{1}{2} \|X - Y\|_F^2 + \lambda \|X\|_*, \quad (1.1)$$

where $Y \in \mathbb{R}^{m \times n}$ denotes the given observation, $\|\cdot\|_F$ denotes the Frobenius norm, $\|X\|_* = \sum_{i=1}^l \sigma_i(X)$ is the nuclear norm of X with $\sigma_i(X)$ as the i -th largest singular value of X , $l = \min(m, n)$, and $\lambda > 0$ is the regularization parameter. According to the work in [9], \hat{X} defined in (1.1) has a closed form, which can be obtained by using a soft-thresholding operation on the singular values of the observation matrix Y , that is,

$$\hat{X} = \text{prox}_{\lambda \|\cdot\|_*}(Y) = U \mathcal{D}_\lambda(\Sigma) V^T,$$

where $Y = U \Sigma V^T$ denotes the singular value decomposition (SVD) of Y , U and V are, respectively, $m \times l$ and $n \times l$ matrices with orthonormal columns, Σ is an $l \times l$ diagonal matrix with the main diagonal $[\sigma_1(Y), \sigma_2(Y), \dots, \sigma_l(Y)]^T$, and $\mathcal{D}_\lambda : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$ is an operator which applies the soft-thresholding on each element with parameter λ . Since all elements in Σ are non-negative, we have

$$(\mathcal{D}_\lambda(\Sigma))_{i,i} = \max(\Sigma_{i,i} - \lambda, 0).$$

The main limitation of NNM methods is that all singular values are weighted equally, which may not be reasonable in some applications. As an example in image denoising, larger singular values are usually associated with the major image patterns and textures, while smaller singular values are usually associated with random noise. Thus, when we use NNM as regularization, the larger singular values should be weighted less in order to preserve major data components, while the smaller singular values should be weighted more in order to remove noise. In [10, 11] the weight nuclear norm $\|\cdot\|_{\omega,*}$ was proposed, which is defined as follows:

$$\|X\|_{\omega,*} = \sum_{i=1}^l \omega_i \sigma_i(X),$$

where $\omega = [\omega_1, \omega_2, \dots, \omega_l]^T$ with $\omega_i \geq 0$ for all $i = 1, \dots, l$ includes all weights. Combined with the l^2 -data-fidelity term, a weighted nuclear norm minimization (WNNM) model was further proposed as follows:

$$\hat{X} = \arg \min_{X \in \mathbb{R}^{m \times n}} \frac{1}{2} \|X - Y\|_F^2 + \|X\|_{\omega,*}. \quad (1.2)$$

The minimization problem defined in (1.2) is also called the weighted nuclear norm proximal (WNNP) problem. Its solution is a low rank approximation to Y and can be obtained efficiently as shown in [10, 11].

For the the l^2 -data-fidelity term used in (1.2), it potentially assumes that the noise in Y is additive white Gaussian noise. However, in many applications different data-fidelity terms are considered to remove non-Gaussian noise [12–15]. For example, the l^1 -data-fidelity term is usually used to remove impulse noise like the salt-and-pepper noise and the Laplace noise [12, 16–19]. In this paper, we combine the weighted nuclear norm with the l^1 -data-fidelity term

and propose a new WNNM model, called the L1-WNNM model. To solve the non-convex optimization problem in the proposed model, we apply the alternating direction method of multipliers (ADMM) [24] and study its convergence. According to the image non-local self-similarity [20], patch matrices formed by non-local similar patches extracted from the image should be low rank. Then we take advantage of this low rank prior and apply the L1-WNNM model on patch matrices to restore the image corrupted by impulse noise. In addition, we illustrate the performance of the new method for impulse noise removal through some numerical experiments.

The rest paper is organized as follows. In Section 2, we investigate the L1-WNNM model (2.1) in details, develop an optimization method to solve it and provide some convergence results. In Section 3, we apply the L1-WNNM model to remove impulse noise. Experimental results are presented in Section 4, and we conclude our paper in Section 5.

2. The L1-WNNM Model

In this section, we introduce an L1-WNNM model that can be used to remove impulse noise. Then we develop an algorithm to solve the minimization problem in this model and provide some convergence results.

By combining the weighted nuclear norm with the l^1 -data-fidelity term, we introduce *the L1-WNNM model*:

$$\min_{X \in \mathcal{C}} \sum_{i=1}^m \sum_{j=1}^n |X_{i,j} - Y_{i,j}| + \|X\|_{\omega,*}, \quad (2.1)$$

where $Y \in \mathbb{R}^{m \times n}$ denotes the noisy image, $\mathcal{C} \subseteq \mathbb{R}^{m \times n}$ denotes a nonempty closed convex set that describes feasible constraints on X , and the weight vector ω is sorted in a non-descending order, that is, $0 \leq \omega_1 \leq \omega_2 \leq \dots \leq \omega_l$, $l = \min(m, n)$. Since one of the main characteristics of impulse noise is that it only corrupts part of the image, many denoising methods are equipped with a noise detector, see [21–23]. In this case, associated with the noisy image $Y \in \mathbb{R}^{m \times n}$, the convex set \mathcal{C} can be chosen as

$$\mathcal{C} = \{X \in \mathbb{R}^{m \times n} : X_{i,j} = Y_{i,j} \text{ with } (i, j) \in \mathcal{U}\},$$

where \mathcal{U} is a subset of $\{1, \dots, m\} \times \{1, \dots, n\}$ that marks all noise-free pixels according to a noise detection result. In other words, the convex set \mathcal{C} can be described as

$$\mathcal{C} = \{X \in \mathbb{R}^{m \times n} : P_{\mathcal{U}} \odot X = P_{\mathcal{U}} \odot Y\}, \quad (2.2)$$

where $P_{\mathcal{U}} \in \mathbb{R}^{m \times n}$ denotes an indicator matrix with element 1 on the pixels marked in \mathcal{U} and element 0 elsewhere, and \odot denotes the matrix pointwise multiplication. Under the same constraint as defined in (2.2), the L1-WNNM model can also be used for image inpainting problems.

In the L1-WNNM model (2.1), the l^1 -data-fidelity term is non-differentiable and the regularization term $\|\cdot\|_{\omega,*}$ is non-smooth. And in this paper we do not consider equal weights, which implies that $\|\cdot\|_{\omega,*}$ is non-convex. Thus, it is very challenging to solve the minimization problem in the model (2.1).

2.1. Optimization algorithm

To solve the non-convex non-smooth optimization problem in the model (2.1), we reformulate this model and apply the alternating direction method of multipliers (ADMM) [24].

First, we transform the constrained problem (2.1) into an unconstrained problem by introducing the indicator function $\iota_{\mathcal{C}}(X)$ of the nonempty closed convex set \mathcal{C} . The resulting unconstrained problem is

$$\min_{X \in \mathbb{R}^{m \times n}} \sum_{i=1}^m \sum_{j=1}^n |X_{i,j} - Y_{i,j}| + \|X\|_{\omega,*} + \iota_{\mathcal{C}}(X),$$

where the indicator function $\iota_{\mathcal{C}}(X)$ is defined as

$$\iota_{\mathcal{C}}(X) = \begin{cases} 0, & X \in \mathcal{C}, \\ +\infty, & X \notin \mathcal{C}. \end{cases}$$

Then we split the data-fidelity term and the regularization term by introducing a new variable $E = Y - X$, that is,

$$\begin{aligned} \min_{X, E \in \mathbb{R}^{m \times n}} \quad & \sum_{i=1}^m \sum_{j=1}^n |E_{i,j}| + \|X\|_{\omega,*} + \iota_{\mathcal{C}}(X) \\ \text{s.t.} \quad & E_{i,j} = Y_{i,j} - X_{i,j} \quad \text{for all } 1 \leq i \leq m, \quad 1 \leq j \leq n. \end{aligned} \tag{2.3}$$

The corresponding augmented Lagrange function is

$$\Psi(E, X, L, \mu) = \sum_{i=1}^m \sum_{j=1}^n |E_{i,j}| + \|X\|_{\omega,*} + \iota_{\mathcal{C}}(X) + \langle L, Y - X - E \rangle + \frac{\mu}{2} \|Y - X - E\|_F^2, \tag{2.4}$$

where $L \in \mathbb{R}^{m \times n}$ is the Lagrange multiplier, and $\mu > 0$ is a penalty parameter that controls the convergence speed of the ADMM algorithm.

By alternately minimizing the augmented Lagrangian function Ψ with respect to E and X and then updating the Lagrange multiplier L , the ADMM iterations with an adaptive penalty parameter μ^k are given by

$$\begin{aligned} E^{k+1} &= \arg \min_E \Psi(E, X^k, L^k, \mu^k) \\ &= \arg \min_E \sum_{i=1}^m \sum_{j=1}^n |E_{i,j}| + \frac{\mu^k}{2} \left\| Y + \frac{1}{\mu^k} L^k - X^k - E \right\|_F^2, \end{aligned} \tag{2.5}$$

$$\begin{aligned} X^{k+1} &= \arg \min_X \Psi(E^{k+1}, X, L^k, \mu^k) \\ &= \arg \min_X \|X\|_{\omega,*} + \iota_{\mathcal{C}}(X) + \frac{\mu^k}{2} \left\| Y + \frac{1}{\mu^k} L^k - E^{k+1} - X \right\|_F^2, \end{aligned} \tag{2.6}$$

$$L^{k+1} = L^k + \mu^k (Y - X^{k+1} - E^{k+1}), \tag{2.7}$$

$$\mu^{k+1} = \rho \mu^k, \tag{2.8}$$

where $\rho > 1$.

For the subproblem (2.5) to E^{k+1} , the solution has a closed form

$$E^{k+1} = \mathcal{S}_{\frac{1}{\mu^k}} \left(Y + \frac{1}{\mu^k} L^k - X^k \right), \tag{2.9}$$

where $\mathcal{S}_{\frac{1}{\mu^k}} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$ is the soft-thresholding operator with the parameter $\frac{1}{\mu^k}$ and defined as

$$\left(\mathcal{S}_{\frac{1}{\mu^k}}(X) \right)_{i,j} = \max \left(\left| X_{i,j} \right| - \frac{1}{\mu^k}, 0 \right) \cdot \text{sign}(X_{i,j}).$$

For the subproblem to X^{k+1} , since the convex set \mathcal{C} has the structure defined in (2.2), we can rewrite (2.6) as

$$X^{k+1} = P_{\mathcal{U}} \odot Y + (\mathbf{1} - P_{\mathcal{U}}) \odot \arg \min_X \left\{ \|X\|_{\omega,*} + \frac{\mu^k}{2} \left\| Y + \frac{1}{\mu^k} L^k - E^{k+1} - X \right\|_F^2 \right\}, \quad (2.10)$$

where $\mathbf{1}$ denotes an all 1-element matrix with size of $m \times n$. Note that the minimization problem in (2.10) is in the same form as (1.2), so according to the closed form solution to (1.2) given in the following theorem we obtain the solution to (2.10).

Theorem 2.1 (Theorem 1 in [10]). *Given $Y \in \mathbb{R}^{m \times n}$, let $Y = U\Sigma V^T$ be the SVD of Y . If the weights are sorted in a non-descending order, that is, $0 \leq \omega_1 \leq \omega_2 \leq \dots \leq \omega_l$, $l = \min(m, n)$, then the global optimum of WNNP problem (1.2) can be expressed as*

$$\hat{X} = \text{prox}_{\|\cdot\|_{\omega,*}}(Y) = U\mathcal{D}_{\omega}(\Sigma)V^T,$$

where $\mathcal{D}_{\omega}(\Sigma)$ is the generalized soft-thresholding operator with the weight vector ω that applies on diagonal matrix Σ and returns a diagonal matrix with the (i, i) -entry given by

$$(\mathcal{D}_{\omega}(\Sigma))_{i,i} = \max(\Sigma_{i,i} - \omega_i, 0).$$

Then, the closed-form solution to (2.10) is given by

$$X^{k+1} = P_{\mathcal{U}} \odot Y + (\mathbf{1} - P_{\mathcal{U}}) \odot (U^k \Sigma^k (V^k)^T), \quad (2.11)$$

where $U^k \Sigma^k (V^k)^T$ is SVD of the matrix $(Y + \frac{1}{\mu^k} L^k - E^{k+1})$ and $\Sigma^k = \mathcal{D}_{\frac{\omega}{\mu^k}}(\Lambda^k)$.

The overall algorithm for solving the minimization problem in the L1-WNNM model (2.1) is given in Algorithm 2.1.

Algorithm 2.1. ADMM Algorithm for the L1-WNNM model (2.1)

- 1: **Input:** the noisy image Y and the weight vector ω
- 2: **Initialize:** $X^0 = Y, L^0=0, \mu^0 > 0, \theta > 0, \rho > 1, k=0$.
- 3: **repeat**
- 4: Update E^{k+1} via (2.9)
- 5: Update X^{k+1} via (2.11)
- 6: Update L^{k+1} via (2.7)
- 7: Update μ^{k+1} via (2.8)
- 8: $k \leftarrow k + 1$
- 9: **until:** $\|Y - E^{k+1} - X^{k+1}\|_F / \|Y\|_F < \theta$
- 10: **Output:** the restored image X^{k+1} .

2.2. Convergence analysis

For convex models, ADMM based methods can ensure optimal solutions if the penalty parameter μ^k is fixed or bounded [25]. However, for the proposed non-convex model, ADMM

with a fixed μ^k may converge very slowly and rely heavily on the choice of μ^k [26]. Therefore, in this paper, we consider an increasing sequence $\{\mu^k\}$ to improve the practical performance and get some convergence results using the unboundedness of $\{\mu^k\}$.

We first prove the boundedness of the sequences $\{E^k\}$, $\{X^k\}$ and $\{L^k\}$ generated by Algorithm 2.1 in the following proposition.

Proposition 2.1. *If the weights are sorted in a non-descending order, then the sequences $\{E^k\}$, $\{X^k\}$ and $\{L^k\}$ generated by Algorithm 2.1 are bounded.*

Proof. First, we show that $\{L^k\}$ is bounded. Recall that the SVD of the matrix $(Y + \frac{1}{\mu^k}L^k - E^{k+1})$ at the $(k + 1)$ -th iteration is defined as

$$Y + \frac{1}{\mu^k}L^k - E^{k+1} = U^k \Lambda^k (V^k)^T, \tag{2.12}$$

where Λ^k is the diagonal singular value matrix with non-negative elements, and U^k and V^k are, respectively, $m \times l$ and $n \times l$ matrices with orthonormal columns, $l = \min(m, n)$. Then, based on the update of L^{k+1} in (2.7) together with (2.12) and the update of X^{k+1} in (2.11), we have

$$\begin{aligned} \|L^{k+1}\|_F &= \mu^k \left\| \frac{1}{\mu^k}L^k + Y - E^{k+1} - X^{k+1} \right\|_F \\ &= \mu^k \left\| U^k \Lambda^k (V^k)^T - (P_U \odot Y + (\mathbf{1} - P_U) \odot (U^k \Sigma^k (V^k)^T)) \right\|_F \\ &\leq \mu^k \left\| U^k \Lambda^k (V^k)^T - U^k \Sigma^k (V^k)^T \right\|_F + \mu^k \left\| P_U \odot (Y - U^k \Sigma^k (V^k)^T) \right\|_F \\ &= \mu^k \|\Lambda^k - \Sigma^k\|_F + \mu^k \left\| P_U \odot \left(U^k \Lambda^k (V^k)^T + E^{k+1} - \frac{1}{\mu^k}L^k - U^k \Sigma^k (V^k)^T \right) \right\|_F \\ &= \mu^k \left\| \Lambda^k - \mathcal{D}_{\frac{\omega}{\mu^k}}(\Lambda^k) \right\|_F + \mu^k \left\| P_U \odot \left(E^{k+1} - \frac{1}{\mu^k}L^k \right) \right\|_F \\ &\quad + \mu^k \left\| P_U \odot \left(\Lambda^k - \mathcal{D}_{\frac{\omega}{\mu^k}}(\Lambda^k) \right) \right\|_F. \end{aligned}$$

We note that

$$\mu^k \left\| P_U \odot \left(\Lambda^k - \mathcal{D}_{\frac{\omega}{\mu^k}}(\Lambda^k) \right) \right\|_F \leq \mu^k \left\| \Lambda^k - \mathcal{D}_{\frac{\omega}{\mu^k}}(\Lambda^k) \right\|_F \leq \mu^k \sqrt{\sum_{i=1}^l \left(\frac{\omega_i}{\mu^k} \right)^2} = \sqrt{\sum_{i=1}^l (\omega_i)^2}. \tag{2.13}$$

Since the update of E^{k+1} has a closed form as (2.9), we have

$$\begin{aligned} &\left\| P_U \odot \left(E^{k+1} - \frac{1}{\mu^k}L^k \right) \right\|_F \\ &= \left\| P_U \odot \left(S_{\frac{1}{\mu^k}} \left(Y + \frac{1}{\mu^k}L^k - X^k \right) - \left(Y + \frac{1}{\mu^k}L^k - X^k \right) \right) + P_U \odot (Y - X^k) \right\|_F \\ &= \left\| P_U \odot \left(S_{\frac{1}{\mu^k}} \left(Y + \frac{1}{\mu^k}L^k - X^k \right) - \left(Y + \frac{1}{\mu^k}L^k - X^k \right) \right) \right\|_F \leq \frac{mn}{\mu^k}. \end{aligned} \tag{2.14}$$

Thus, we get

$$\|L^{k+1}\|_F \leq 2 \sqrt{\sum_{i=1}^l (\omega_i)^2 + mn}.$$

Before showing that $\{E^k\}$ and $\{X^k\}$ are bounded, we show that $\{\Psi(E^{k+1}, X^{k+1}, L^k, \mu^k)\}$ with Ψ defined in (2.4) is upper bounded. According to the update of E^{k+1} in (2.5) and the update of X^{k+1} in (2.6), we have

$$\Psi(E^{k+1}, X^{k+1}, L^k, \mu^k) \leq \Psi(E^{k+1}, X^k, L^k, \mu^k) \leq \Psi(E^k, X^k, L^k, \mu^k).$$

By using the update of L^k , we obtain

$$\begin{aligned} \Psi(E^k, X^k, L^k, \mu^k) &= \Psi(E^k, X^k, L^{k-1}, \mu^{k-1}) + \langle L^k - L^{k-1}, Y - X^k - E^k \rangle \\ &\quad + \frac{\mu^k - \mu^{k-1}}{2} \|Y - X^k - E^k\|_F^2 \\ &= \Psi(E^k, X^k, L^{k-1}, \mu^{k-1}) + \langle L^k - L^{k-1}, \frac{1}{\mu^{k-1}}(L^k - L^{k-1}) \rangle \\ &\quad + \frac{\mu^k - \mu^{k-1}}{2} \left\| \frac{1}{\mu^{k-1}}(L^k - L^{k-1}) \right\|_F^2 \\ &= \Psi(E^k, X^k, L^{k-1}, \mu^{k-1}) + \frac{\mu^k + \mu^{k-1}}{2(\mu^{k-1})^2} \|L^k - L^{k-1}\|_F^2. \end{aligned}$$

Thus,

$$\Psi(E^{k+1}, X^{k+1}, L^k, \mu^k) \leq \Psi(E^1, X^1, L^0, \mu^0) + 4M \sum_{j=1}^k \frac{\mu^j + \mu^{j-1}}{2(\mu^{j-1})^2},$$

where M is the upper bound of $\|L^k\|_F^2$ for all $k = 1, 2, \dots$. Recall that $\mu^{k+1} = \rho\mu^k$ with $\rho > 1$ and $\mu^k > 0$, then we get

$$\sum_{j=1}^{\infty} \frac{\mu^j + \mu^{j-1}}{2(\mu^{j-1})^2} < \infty.$$

Hence, the upper boundedness of $\{\Psi(E^{k+1}, X^{k+1}, L^k, \mu^k)\}$ is proved.

Finally, we show that $\{E^k\}$ and $\{X^k\}$ are bounded. By definition of $\Psi(E^k, X^k, L^{k-1}, \mu^{k-1})$, we have

$$\begin{aligned} &\sum_{i=1}^m \sum_{j=1}^n |E_{i,j}^k| + \|X^k\|_{\omega,*} + \iota_C(X^k) \\ &= \Psi(E^k, X^k, L^{k-1}, \mu^{k-1}) + \frac{\mu^{k-1}}{2} \left(\frac{1}{(\mu^{k-1})^2} \|L^{k-1}\|_F^2 - \|Y - X^k - E^k\|_F^2 + \frac{1}{\mu^{k-1}} \|L^{k-1}\|_F^2 \right) \\ &= \Psi(E^k, X^k, L^{k-1}, \mu^{k-1}) + \frac{1}{2\mu^{k-1}} (\|L^{k-1}\|_F^2 - \|L^k\|_F^2) \\ &\leq \Psi(E^k, X^k, L^{k-1}, \mu^{k-1}) + \frac{M}{\mu^0}. \end{aligned}$$

Since the right-hand side has been proved to be bounded, we directly get that $\{E^k\}$ and $\{X^k\}$ are bounded. □

The boundedness of the sequences $\{E^k\}$, $\{X^k\}$ and $\{L^k\}$ implies the existence of accumulation points of $\{(E^k, X^k, L^k)\}$. And the following theorem regarding the primal residual convergence, meaning $r^{k+1} = Y - X^{k+1} - E^{k+1} \rightarrow 0$ as $k \rightarrow \infty$, implies that any accumulation point of $\{(E^k, X^k, L^k)\}$ is a feasible solution of the proposed model.

Theorem 2.2. *If the weights are sorted in a non-descending order, then the sequences $\{E^k\}$ and $\{X^k\}$ generated by Algorithm 2.1 satisfy*

- (i) $\lim_{k \rightarrow \infty} \|E^{k+1} - E^k\|_F = 0,$
- (ii) $\lim_{k \rightarrow \infty} \|X^{k+1} - X^k\|_F = 0,$
- (iii) $\lim_{k \rightarrow \infty} \|Y - E^{k+1} - X^{k+1}\|_F = 0.$

Proof. (i) According to (2.7) and (2.9), we obtain

$$\begin{aligned} & \|E^{k+1} - E^k\|_F \\ &= \left\| S_{\frac{1}{\mu^k}} \left(Y + \frac{1}{\mu^k} L^k - X^k \right) - \left(Y + \frac{1}{\mu^k} L^k - X^k \right) + \frac{1}{\mu^k} L^k + \frac{1}{\mu^{k-1}} L^k - \frac{1}{\mu^{k-1}} L^{k-1} \right\|_F \\ &\leq \frac{mn}{\mu^k} + \frac{1}{\mu^k} \|L^k\|_F + \frac{1}{\mu^{k-1}} \|L^k\|_F + \frac{1}{\mu^{k-1}} \|L^{k-1}\|_F. \end{aligned}$$

Recall that $\mu^{k+1} = \rho\mu^k$ with $\rho > 1$ and $\mu^k > 0$, and $\{L^k\}$ is bounded. Let k go to infinity on the both sides, then we get $\lim_{k \rightarrow \infty} \|E^{k+1} - E^k\|_F = 0.$

(ii) Based on the update of L^{k+1} in (2.7), we have

$$X^{k+1} = Y + \frac{1}{\mu^k} L^k - E^{k+1} - \frac{1}{\mu^k} L^{k+1}.$$

By using (2.11), we get

$$\begin{aligned} & \|X^{k+1} - X^k\|_F \\ &= \left\| Y + \frac{1}{\mu^k} L^k - E^{k+1} - \frac{1}{\mu^k} L^{k+1} - \left(P_U \odot Y + (\mathbf{1} - P_U) \odot (U^{k-1} \Sigma^{k-1} (V^{k-1})^T) \right) \right\|_F \\ &\leq \left\| Y + \frac{1}{\mu^{k-1}} L^{k-1} - E^k - U^{k-1} \Sigma^{k-1} (V^{k-1})^T \right\|_F + \left\| P_U \odot \left(Y - U^{k-1} \Sigma^{k-1} (V^{k-1})^T \right) \right\|_F \\ &\quad + \left\| E^k - E^{k+1} + \frac{1}{\mu^k} L^k - \frac{1}{\mu^k} L^{k+1} - \frac{1}{\mu^{k-1}} L^{k-1} \right\|_F. \end{aligned}$$

Based on (2.12), $U^{k-1} \Lambda^{k-1} (V^{k-1})^T$ is SVD of the matrix $(Y + \frac{1}{\mu^{k-1}} L^{k-1} - E^k)$ in the k -th iteration, thus

$$\begin{aligned} & \|X^{k+1} - X^k\|_F \\ &\leq \left\| P_U \odot \left(U^{k-1} \Lambda^{k-1} (V^{k-1})^T + E^k - \frac{1}{\mu^{k-1}} L^{k-1} - U^{k-1} \Sigma^{k-1} (V^{k-1})^T \right) \right\|_F \\ &\quad + \left\| \Lambda^{k-1} - \mathcal{D}_{\frac{\omega}{\mu^{k-1}}}(\Lambda^{k-1}) \right\|_F + \|E^k - E^{k+1}\|_F + \left\| \frac{1}{\mu^k} L^k - \frac{1}{\mu^k} L^{k+1} - \frac{1}{\mu^{k-1}} L^{k-1} \right\|_F \\ &\leq \left\| P_U \odot \left(U^{k-1} (\Lambda^{k-1} - \Sigma^{k-1}) (V^{k-1})^T \right) \right\|_F + \left\| \Lambda^{k-1} - \mathcal{D}_{\frac{\omega}{\mu^{k-1}}}(\Lambda^{k-1}) \right\|_F \\ &\quad + \left\| P_U \odot \left(E^k - \frac{1}{\mu^{k-1}} L^{k-1} \right) \right\|_F + \|E^k - E^{k+1}\|_F + \left\| \frac{1}{\mu^k} L^k - \frac{1}{\mu^k} L^{k+1} - \frac{1}{\mu^{k-1}} L^{k-1} \right\|_F. \end{aligned}$$

Recall the result from (2.13) and (2.14), then we obtain

$$\begin{aligned} & \|X^{k+1} - X^k\|_F \\ & \leq 2\sqrt{\frac{\sum_{i=1}^l \omega_i^2}{\mu^{k-1}}} + \frac{mn}{\mu^{k-1}} + \|E^k - E^{k+1}\|_F + \left\| \frac{1}{\mu^k}L^k - \frac{1}{\mu^k}L^{k+1} - \frac{1}{\mu^{k-1}}L^{k-1} \right\|_F. \end{aligned}$$

Since $\lim_{k \rightarrow \infty} \frac{1}{\mu^k} = 0$, combining with the first result of (ii), we get

$$\lim_{k \rightarrow \infty} \|X^{k+1} - X^k\|_F = 0.$$

(iii) By the update of L^{k+1} in (2.7), we have $Y - X^{k+1} - E^{k+1} = \frac{1}{\mu^k}(L^{k+1} - L^k)$. Since $\|L^{k+1}\|_F$ is bounded and $\lim_{k \rightarrow \infty} \frac{1}{\mu^k} = 0$, we obtain

$$\lim_{k \rightarrow \infty} \|Y - X^{k+1} - E^{k+1}\|_F = \lim_{k \rightarrow \infty} \frac{1}{\mu^k} \|L^{k+1} - L^k\|_F = 0. \quad \square$$

3. Application of the L1-WNNM Model to Impulse Noise Removal

Impulse noise widely occurs in digital images due to pixel failures in the camera sensors, timing errors in analog-to-digital conversion, and errors in data transmission and data storage [27]. Impulse noise removal is a fundamental step to the subsequent image processing tasks, such as object recognition and image segmentation. Thus, we are motivated to remove impulse noise by applying the L1-WNNM model (2.1) that consists of an l^1 -data-fidelity term that has been shown suitable for impulse noise removal [12], a weighted nuclear norm regularization term that aims to approximate low rank solutions and a set constraint that aims to identify the noisy pixels. In this section, inspired by the work introduced in [10], we take advantage of the non-local self-similarity (NSS) of natural images and apply the L1-WNNM model on patch matrices to remove impulse noise.

Algorithm 3.1. L1-WNNM Impulse Noise Removal

- 1: **Input:** the noisy image Y .
- 2: **Noise detection:** Use a noise detector to detect noise and save the indices of all noise-free pixels in \mathcal{U} .
- 3: **Initialize** $\hat{X}^0 = Y$ and $\theta > 0$.
- 4: **for** $k = 1, 2, \dots, K$ **do**
- 5: Calculate $Y^k = \hat{X}^{k-1} + \theta(Y - \hat{X}^{k-1})$
- 6: **for** each patch y_s in Y^k **do**
- 7: Find similar patches via block matching and stack patches as the matrix Y_s .
- 8: Estimate the weight vector ω_s via (3.2)
- 9: Estimate X_s by solving the L1-WNNM model (3.1) via Algorithm 2.1
- 10: **end for**
- 11: Aggregate all similar patches $\{X_s\}$ to obtain the restored image \hat{X}^k .
- 12: **end for**
- 13: **Output:** the restored image \hat{X}^K .

According to NSS, for a local patch $y_s \in \mathbb{R}^p$ of size $\sqrt{p} \times \sqrt{p}$ at position s in image Y , we can always find enough non-local patches within a local window that share similar patterns with y_s using block matching [20, 28–31], in which the similarity is measure in terms of the Euclidean distance. By stacking all non-local similar patches, we obtain a matrix $Y_s \in \mathbb{R}^{p \times q}$, where p and q denote the number of pixels in each patch and the number of similar patches, respectively. To remove impulse noise, we apply the L1-WNNM model on the non-local similar patch matrix Y_s instead of the whole image Y , and estimate the corresponding clean low rank patch matrix X_s that preserves image details (i.e., textures and structures). After solving the L1-WNNM model for each patch matrix, we aggregate all the estimated patch matrices X_s , meaning averaging the estimated values for each pixel that may belong to several patch matrices, and obtain the restored image X . In practice, we repeat these procedures several times to improve the denoising performance. The whole denoising algorithm is summarized in Algorithm 3.1 and some details of the proposed algorithm are discussed below.

In Algorithm 3.1, the L1-WNNM model in terms of patch matrices is

$$\min_{X_s \in \mathcal{C}_s} \sum_{i=1}^p \sum_{j=1}^q |(X_s)_{i,j} - (Y_s)_{i,j}| + \|X_s\|_{\omega_s, *}, \quad (3.1)$$

where $\mathcal{C}_s := \{X_s \in \mathbb{R}^{p \times q} : (X_s)_{i,j} = (Y_s)_{i,j} \text{ with } (i,j) \in \mathcal{U}_s\}$ marks all noise-free pixels in the similar patches. The weight vector ω_s plays an important role, which should be determined in advance. From the general prior knowledge, the smaller singular values of X_s are less important than the larger ones in the application of image denoising. Therefore, the weight vector is chosen to be in a non-descending order defined as

$$(\omega_s)_i = \frac{c}{\sigma_i(X_s^*) + \varepsilon}, \quad (3.2)$$

where c is a compromising constant that is adjusted manually in our experiments, ε is a small positive number to avoid dividing by 0, and X_s^* is a low rank solution that approximates the given noisy patch matrix Y_s . In particular, we compute X_s^* explicitly using the following lemma in [10] with $Y = Y_s$ and $X^* = X_s^*$.

Lemma 3.1. *Let $Y = U\Sigma V^T$ be SVD of Y . Suppose that the regularization parameter c is positive and the positive value ε is small enough to make the inequality $\varepsilon < \min(\sqrt{c}, \frac{c}{\sigma_1(Y)})$ hold. Then the sequence $\{X^k\}$ generated by*

$$X^{k+1} = \operatorname{argmin}_X \frac{1}{2} \|Y - X\|_F + \|X\|_{\omega^{k+1}, *}$$

with the reweighted formula

$$\omega_i^{k+1} = \frac{c}{\sigma_i(X^k) + \varepsilon}$$

and initial estimation $X^0 = Y$, converges to the closed-form solution: $X^* = U\tilde{\Sigma}V^T$, where $\tilde{\Sigma}$ is a diagonal matrix whose (i, i) -entry is given by $\tilde{\Sigma}_{i,i} = \sigma_i(X^*)$,

$$\sigma_i(X^*) = \begin{cases} 0, & \text{if } d_i < 0, \\ \frac{c_i + \sqrt{d_i}}{2}, & \text{if } d_i \geq 0, \end{cases}$$

$c_i = \sigma_i(Y) - \varepsilon$ and $d_i = (\sigma_i(Y) + \varepsilon)^2 - 4c$.

4. Numerical Results

In this section, we show the numerical results on restoring images corrupted by two types of impulse noise, i.e., the salt-and-pepper noise (SPN) and the random-valued impulse noise

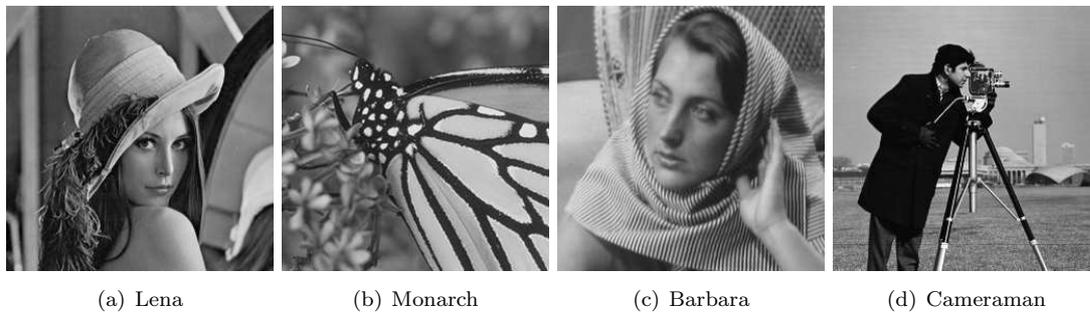


Fig. 4.1. The test images.

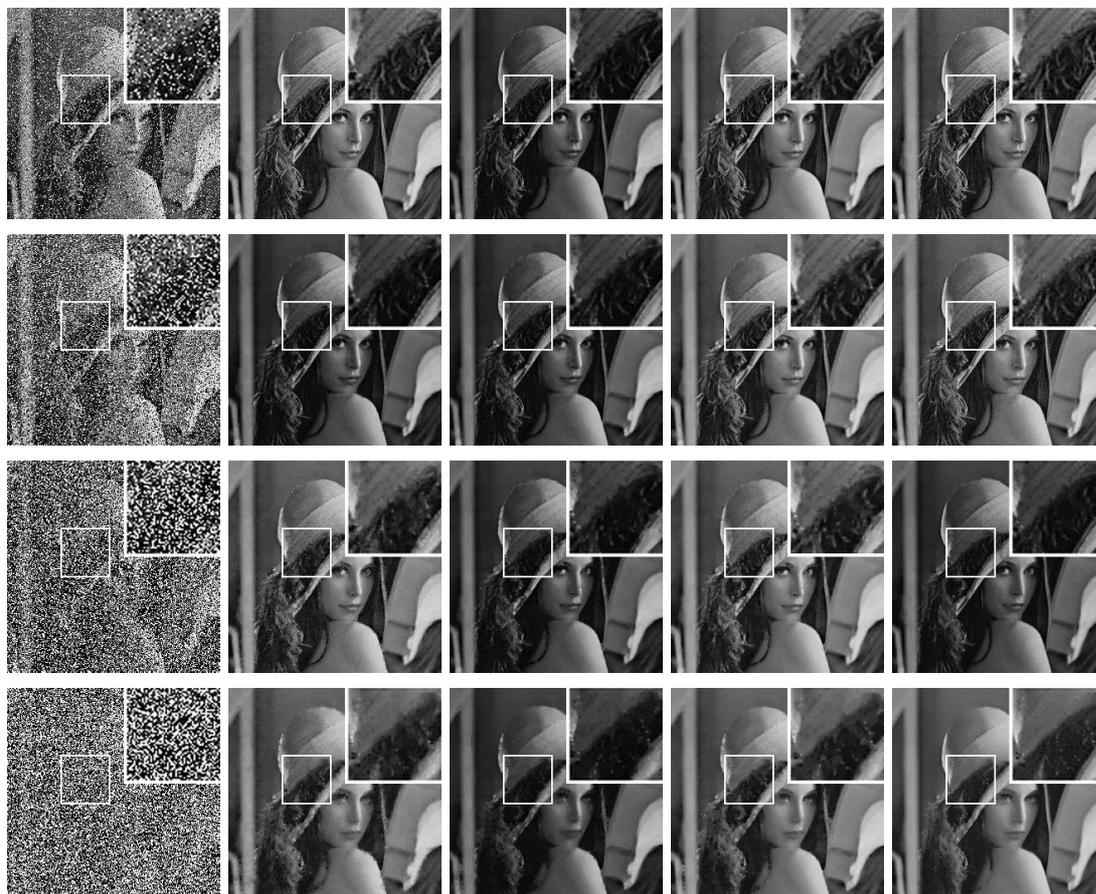


Fig. 4.2. The restored results for “Lena” from the noise level $r = 20\%, 40\%, 60\%, 80\%$ (top to bottom). The first column: the noisy images; the second column: the results from L1-TV; the third column: the results from L0-TV; the fourth column: the results from F2TV; the fifth column: the results from our method.

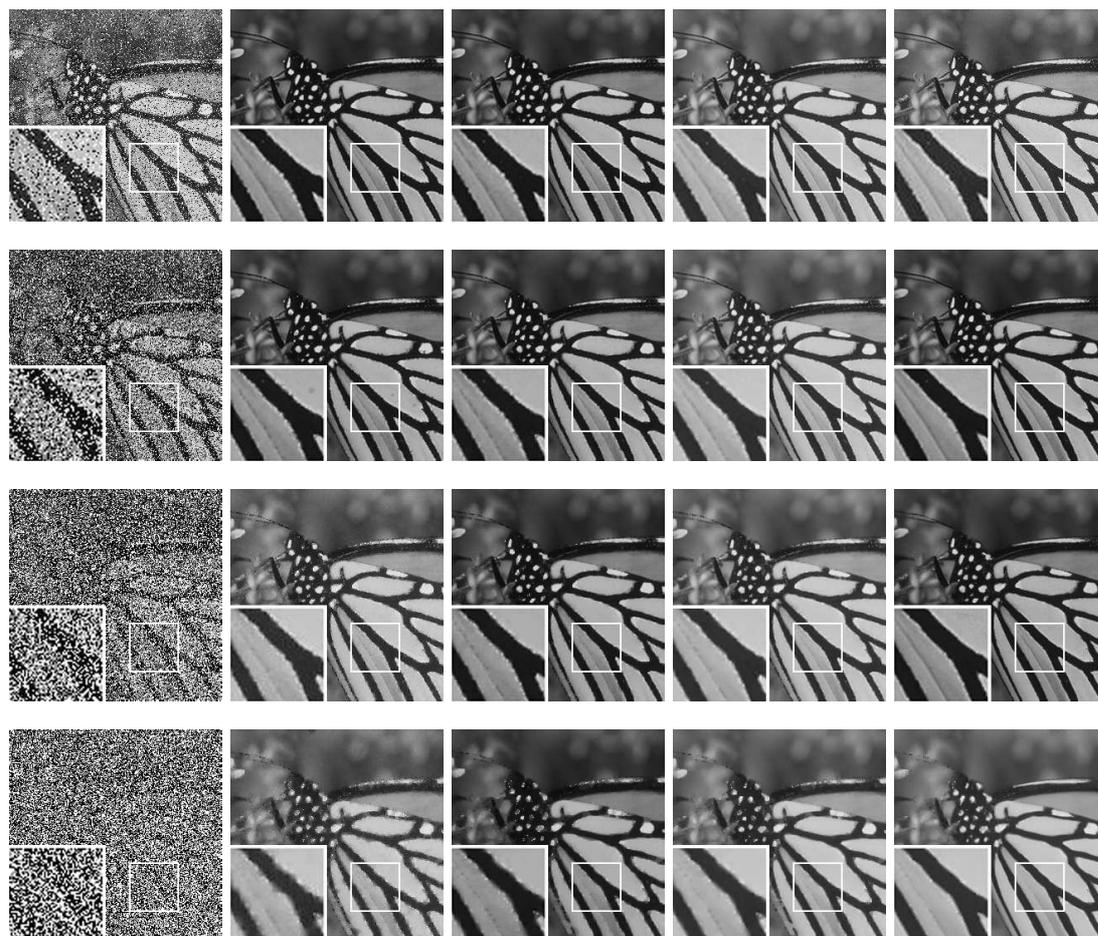


Fig. 4.3. The restored results for “Monarch” from the noise level $r = 20\%$, 40% , 60% , 80% (top to bottom). The first column: the noisy images; the second column: the results from L1-TV; the third column: the results from L0-TV; the fourth column: the results from F2TV; the fifth column: the results from our method.

(RVIN). In our experiments, we use 256×256 8-bit grayscale images “Lena”, “Monarch”, “Barbara” and “Cameraman” as test images, which are shown in Fig. 4.1. In order to evaluate the performance of the proposed method, we utilize the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) [32] to evaluate the results. All the experiments are performed under Windows 7 and MATLAB 7.6 (R2017a) running on a PC equipped with an Intel(R) Pentium(R) CPU G645 at 2.90 GHz.

4.1. Removal of the salt-and-pepper noise

The salt-and-pepper noise (SPN) can be considered as a special case of the random-valued impulse noise, where the noise value only can be the minimum or the maximum of the intensity range $[d_{\min}, d_{\max}]$. The noise level r of SPN denotes the probability that a pixel is corrupted by the noise, and the value of the noisy pixel is either d_{\min} or d_{\max} with the same probability $1/2$.

In the numerical experiments, we compare our proposed method with the L1-TV, L0-TV and F2TV methods proposed in [33], [34], and [35] respectively. All of them are two-stage



Fig. 4.4. The restored results for “Barbara” from the noise level $r = 20\%$, 40% , 60% , 80% (top to bottom). The first column: the noisy images; the second column: the results from L1-TV; the third column: the results from L0-TV; the fourth column: the results from F2TV; the fifth column: the results from our method.

methods, i.e., first noisy pixels are detected, then total-variation-based variational methods are applied to restore the noisy pixels. In all three methods, we use the adaptive median filter (AMF) [27] as noise detector. And we test all the methods on the test images corrupted by SPN with the noise levels $r = 20\%$, 40% , 60% and 80% .

In Algorithm 2.1, we set $\rho = 1.05$, and set $\theta = 0.02$ for the noise level $r \leq 50\%$, and $\theta = 0.0001$ for $r > 50\%$. In Algorithm 3.1, the patch sizes are set by experience respectively 6×6 for the cases of $r \leq 20\%$, 7×7 for $20\% < r \leq 60\%$, and 8×8 for $r > 60\%$; and the maximum iteration number K is set as 18 for $r \leq 60\%$ and 25 for $r > 60\%$, respectively.

In Figs. 4.2-4.5, we show the restored results from the image corrupted by SPN with different noise level. It is clear that all four methods are effective for removing the noise, and our method preserves the details, especially the textures, much better, see the textures on the butterfly in “Monarch” and in the scarf in “Barbara”. In Table 4.1, we list the PSNR and SSIM values from all four methods. It is obvious that the PSNR values from our method are higher than the other three. Taking “Lena” as an example, we can see that the improvements of PSNR

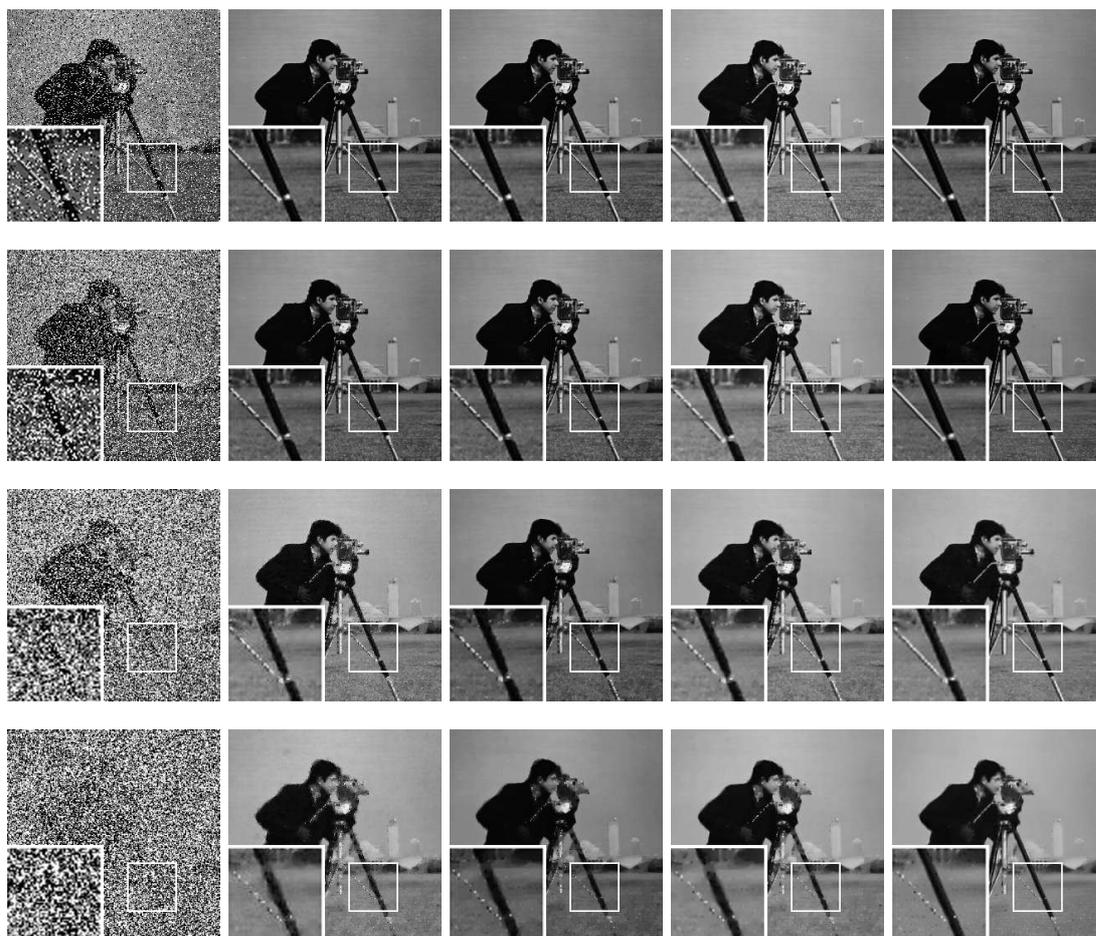


Fig. 4.5. The restored results for “Cameraman” from the noise level $r = 20\%$, 40% , 60% , 80% (top to bottom). The first column: the noisy images; the second column: the results from L1-TV; the third column: the results from L0-TV; the fourth column: the results from F2TV; the fifth column: the results from our method.

values are 0.07dB to 1.78dB. For “Barbara”, PSNR from our method is even more than 2dB larger than others in every noise level. The main reason here is that there are clear pattern structures like on the scarf and trousers in “Barbara”, and it is perfectly fit for applying the proposed patch-based method.

4.2. Removal of Random-Valued Impulse Noise

The random-valued impulse noise (RVIN) can randomly take values from the intensity range $[d_{\min}, d_{\max}]$. The noise level r of RVIN denotes the probability of each pixel being corrupted by noise, and the noise value is uniformly distributed on $[d_{\min}, d_{\max}]$.

In the numerical experiments, we compare our method with other two methods. One is the fast total-variation-based two-phase denoising (F2TV) method introduced in [35], where, as well as in our method, we use marked/known noisy pixels. The other is the L0-TV method [34],

Table 4.1: The comparison of different methods for removing the salt-and-pepper noise with different noise levels.

Img	Level	Noisy		L1TV		L0TV		F2TV		Ours	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Lena	20%	12.15	0.1331	36.42	0.9797	36.73	0.9805	36.92	0.9812	37.30	0.9808
	40%	9.12	0.0610	32.04	0.9507	32.53	0.9514	32.66	0.9543	32.73	0.9499
	60%	7.40	0.0316	29.12	0.9097	28.98	0.9051	29.39	0.9118	30.76	0.9218
	80%	6.14	0.0147	25.56	0.8116	25.30	0.8070	25.72	0.8139	26.45	0.8210
Monarch	20%	12.33	0.1890	35.68	0.9869	35.89	0.9874	36.23	0.9888	36.33	0.9861
	40%	9.28	0.0943	30.27	0.9612	31.27	0.9618	31.92	0.9699	32.05	0.9649
	60%	7.48	0.0512	27.44	0.9249	27.32	0.9242	28.00	0.9335	29.46	0.9445
	80%	6.26	0.0241	23.25	0.8246	23.00	0.8225	23.51	0.8432	24.86	0.8483
Barbara	20%	12.43	0.1580	33.22	0.9736	33.11	0.9731	33.36	0.9748	38.76	0.9865
	40%	9.40	0.0713	28.78	0.9255	28.61	0.9242	28.89	0.9290	34.33	0.9637
	60%	7.62	0.0345	25.86	0.8482	25.67	0.8429	25.93	0.8529	32.24	0.9400
	80%	6.35	0.0156	23.59	0.7233	23.41	0.7102	23.61	0.7283	25.46	0.7667
Camera.	20%	12.01	0.1330	33.22	0.9745	33.32	0.9746	33.46	0.9750	35.29	0.9750
	40%	9.06	0.0664	28.82	0.9387	29.15	0.9403	29.53	0.9437	31.10	0.9382
	60%	7.31	0.0364	26.05	0.8868	25.92	0.8855	26.21	0.8913	28.25	0.9025
	80%	6.02	0.0166	23.12	0.7905	22.96	0.7909	23.18	0.8022	24.20	0.8035

Table 4.2: The comparison of different methods for removing the random-valued impulse noise with different noise levels.

Img	Level	Noisy		L0TV		F2TV		Ours	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Lena	20%	15.54	0.2372	27.62	0.8983	36.54	0.9795	37.84	0.9805
	40%	12.60	0.1244	24.92	0.7818	32.88	0.9540	33.96	0.9545
	60%	10.81	0.0709	21.94	0.7099	29.43	0.9083	31.46	0.9302
	80%	9.58	0.0340	18.69	0.5445	25.77	0.8195	26.53	0.8341
Monarch	20%	15.85	0.3021	24.38	0.8937	36.30	0.9885	37.70	0.9882
	40%	12.83	0.1719	22.17	0.7256	31.83	0.9692	33.34	0.9699
	60%	11.13	0.1022	19.38	0.6447	27.68	0.9337	30.36	0.9550
	80%	9.85	0.0504	16.34	0.4717	23.17	0.8410	24.85	0.8657
Barbara	20%	16.12	0.2899	25.67	0.8362	33.74	0.9761	40.49	0.9885
	40%	13.21	0.1546	22.76	0.7281	29.21	0.9306	35.47	0.9688
	60%	11.35	0.0822	21.25	0.6155	25.92	0.8520	32.24	0.9400
	80%	10.06	0.0413	19.82	0.4758	23.75	0.7274	27.83	0.8507
Camera.	20%	15.55	0.2312	23.77	0.8697	33.94	0.9762	36.01	0.9774
	40%	12.47	0.1278	22.17	0.7887	29.22	0.9422	31.59	0.9382
	60%	10.65	0.0715	19.86	0.6898	26.27	0.8892	28.77	0.9093
	80%	9.37	0.0353	17.20	0.5522	23.45	0.8058	23.85	0.7976

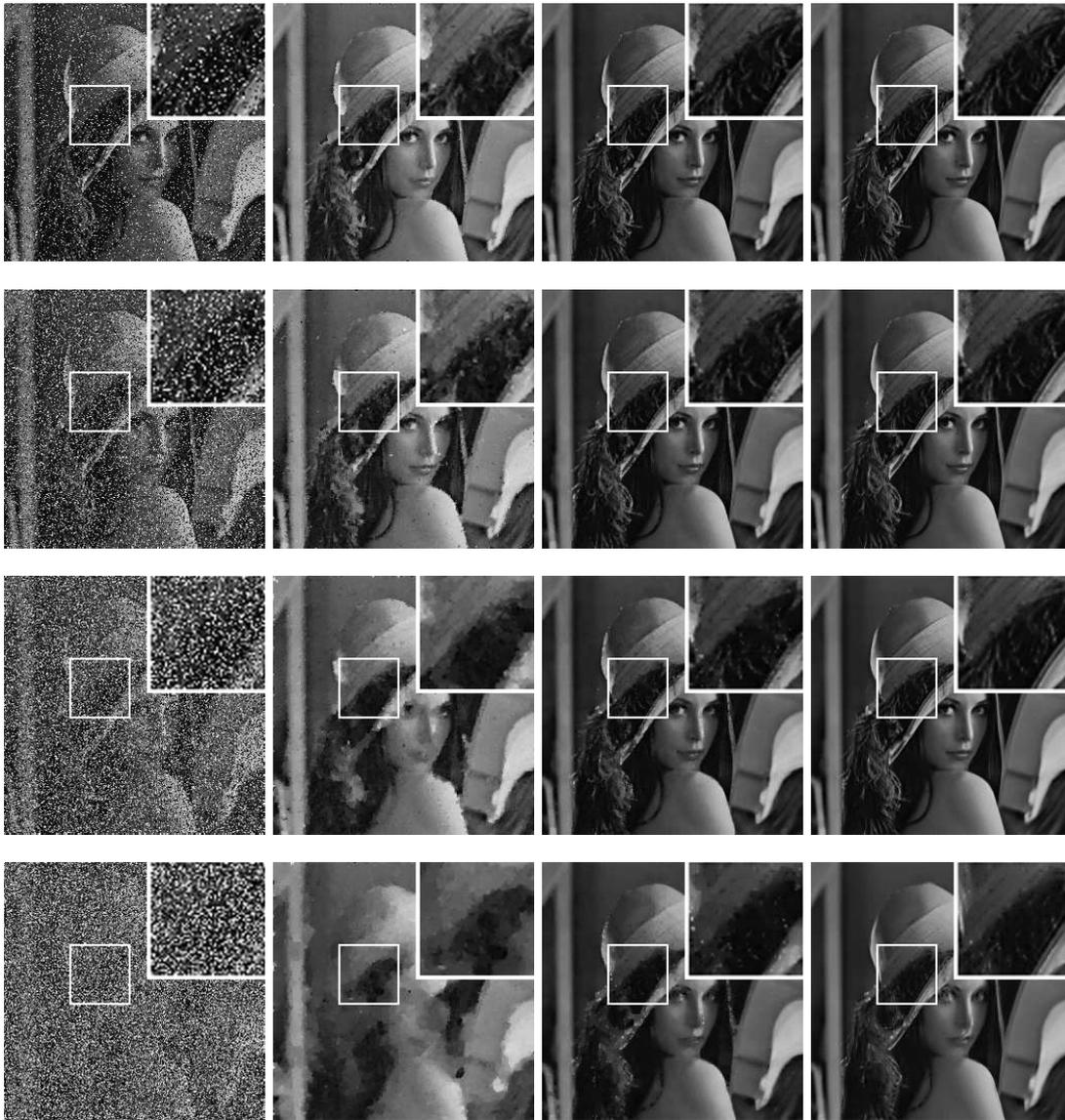


Fig. 4.6. The restored results for “Lena” from the noise level $r = 20\%$, 40% , 60% , 80% (top to bottom). The first column: the noisy images; the second column: the results by applying the L0-TV method; the third column: the results by applying the F2TV method; the fourth column: the results by applying our method.

where does not use noise detection for random-valued impulse noise. Since in RVIN the noise can be any values in the intensity range and some of it can be very close to the original values, RVIN is much more difficult to detect, especially under high noise levels. However, the capability of the two-phase method strongly depends on the accuracy of the noise detection. Here, we test both methods on the test images corrupted by RVIN with the noise level 20%, 40%, 60% and 80%.

In Algorithm 2.1, we set $\rho = 1.05$, $\theta = 0.02$ for the noise level $r \leq 50\%$, and $\theta = 0.0001$ for $r > 50\%$. In Algorithm 3.1, the patch sizes are set by experience respectively 7×7 for $r \leq 60\%$,

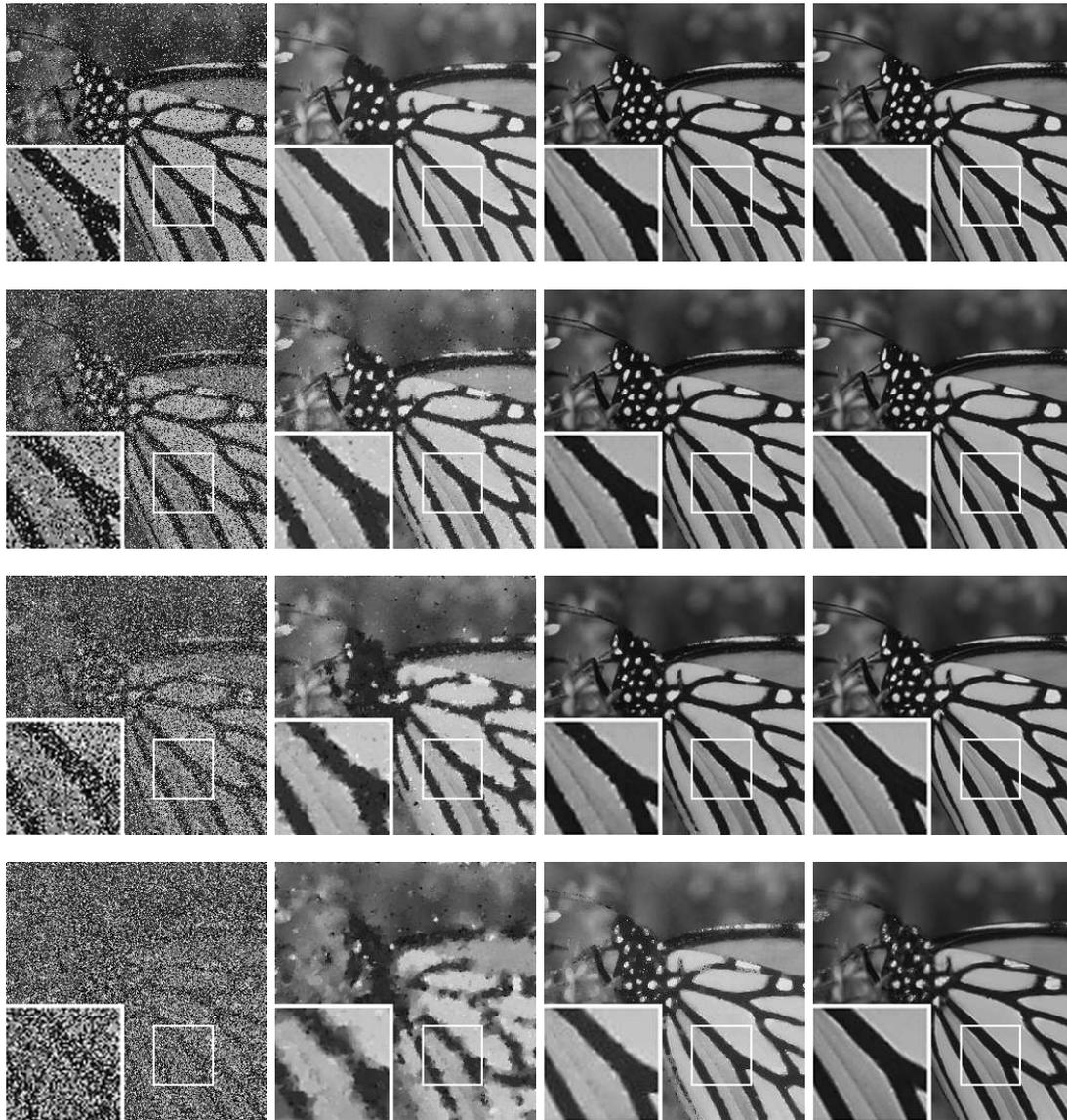


Fig. 4.7. The restored results for “Monarch” from the noise level $r = 20\%, 40\%, 60\%, 80\%$ (top to bottom). The first column: the noisy images; the second column: the results by applying the L0-TV method; the third column: the results by applying the F2TV method; the fourth column: the results by applying our method.

and 12×12 for $r > 60\%$; and the number of the iterations K is set as 18 for $r \leq 60\%$ and 22 for $r > 60\%$, respectively.

In Figs. 4.6-4.9, we show the denoising results from all the methods. It is obvious that our method provides the best results visually, especially to the images full of textures under high noise level, e.g., “Monarch” and “Barbara”. In Table 4.2, we list the PSNR and SSIM values for both methods to restore different images corrupted by RVIN with the noise level 20% to 80%. We can see that the PSNR values from our method are much higher than those from the



Fig. 4.8. The restored results for “Barbara” from the noise level $r = 20\%, 40\%, 60\%, 80\%$ (top to bottom). The first column: the noisy images; the second column: the results by applying the L0-TV method; the third column: the results by applying the F2TV method; the fourth column: the results by applying our method.

L0-TV and F2TV methods. For example, for “Barbara”, the improvement of the PSNR values by using our method reaches to 4dB to 15dB.

5. Conclusions

In this paper, we introduce a new variational model, L1-WNNM, which combines the l^1 -data-fidelity term with the weighted nuclear norm for removing impulse noise in images. Furthermore, we propose a numerical algorithm for solving the minimization problem in the L1-WNNM

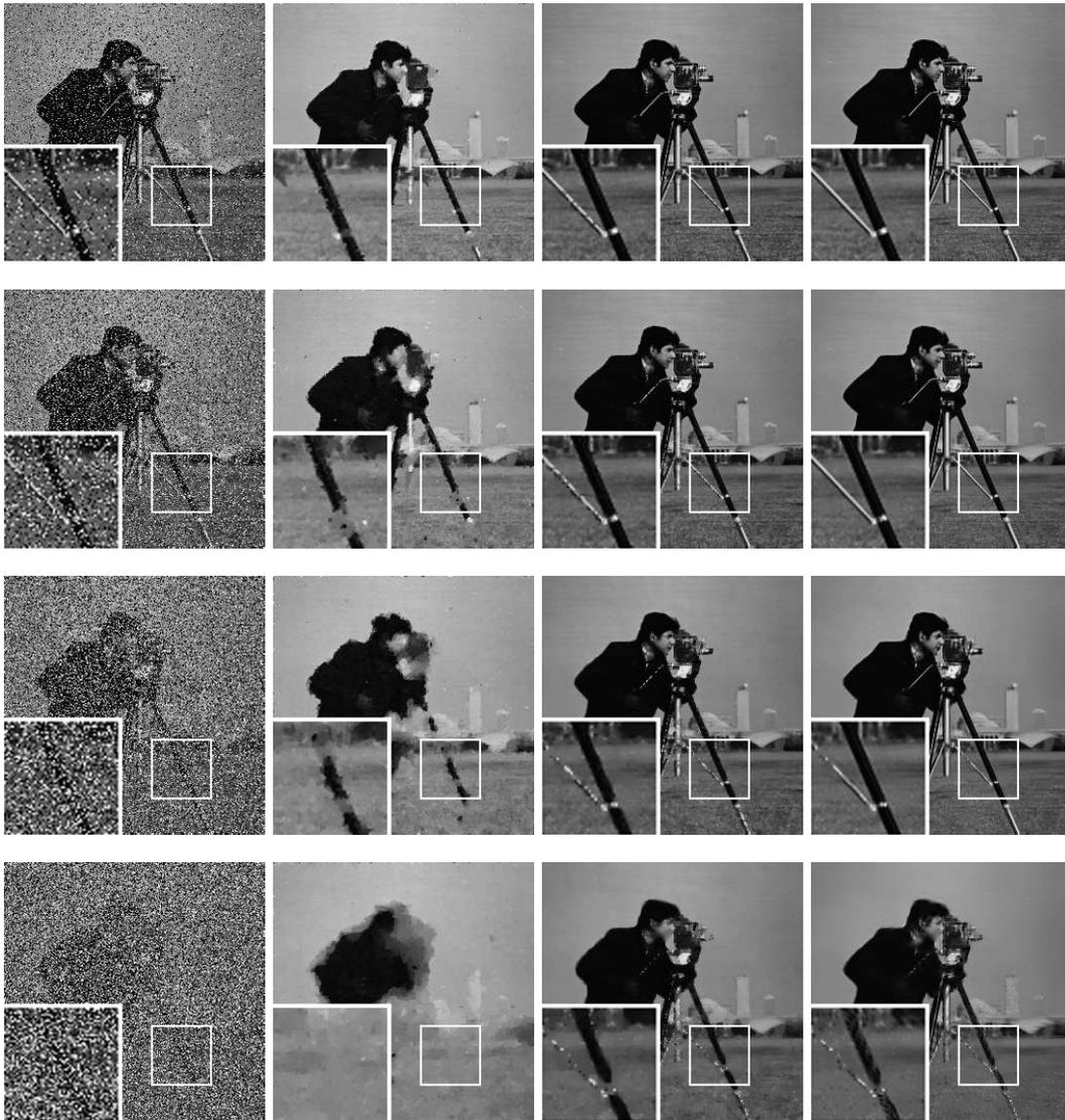


Fig. 4.9. The restored results for “Cameraman” from the noise level $r = 20\%$, 40% , 60% , 80% (top to bottom). The first column: the noisy images; the second column: the results by applying the L0-TV method; the third column: the results by applying the F2TV method; the fourth column: the results by applying our method.

model, and study its convergence. Since the new model is non-convex, we only provide the primal residual convergence results. Through numerical experiments, we show that the method can provide better denoising results in terms of PSNR and SSIM values as well as visually.

Acknowledgments. This work was supported by the National Natural Science Foundation of China under grants U21A20455, 61972265, 11871348 and 11701388, by the Natural Science Foundation of Guangdong Province of China under grant 2020B1515310008, by the Educational Commission of Guangdong Province of China under grant 2019KZDZX1007.

References

- [1] A. Eriksson and A. Van Den Hengel, Efficient computation of robust low-rank matrix approximations in the presence of missing data using the L1 norm, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (2010), 771–778.
- [2] A.M. Buchanan and A.W. Fitzgibbon, Damped Newton Algorithms for Matrix Factorization with Missing Data, *Computer Vision and Pattern Recognition*, **2** (2005), 316–322.
- [3] E. J Candès and B. Recht, Exact Matrix Completion via Convex Optimization, *Foundations of Computational Mathematics*, **9:6** (2009), 259–268.
- [4] M. Fazel, H. Hindi and S. Boyd, A rank minimization heuristic with application to minimum order system approximation, *Proceedings of the American Control Conference*, **6:6** (2001), 4734–4739.
- [5] E.J. Candès, X. Li, Y. Ma and J. Wright, Robust Principal Component Analysis? *Journal of the ACM*, **58:3** (2011), 11:1–11:37.
- [6] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu and Y. Ma, Robust Subspace Segmentation by Low-Rank Representation, *International Conference on Machine Learning*, **6** (2010), 21–24
- [7] J. Wright, A. Ganesh, S. Rao and Y. Ma, Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices, *Proceedings of Advances in Neural Information Processing Systems*, (2009), 2080–2088.
- [8] N. Srebro and T. Jaakkola, Weighted Low-Rank Approximations, *International Conference on Machine Learning*, **8:21–24** (2003), 720–727.
- [9] J.F. Cai, E.J. Candès and Z. Shen, A Singular Value Thresholding Algorithm for Matrix Completion, *SIAM Journal on Optimization*, **20:4** (2010), 1956–1982.
- [10] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng and L. Zhang, Weighted Nuclear Norm Minimization and Its Applications to Low Level Vision, *in: International Journal of Computer Vision*, **121:2** (2017), 183–208.
- [11] S. Gu, L. Zhang, W. Zuo and X. Feng, Weighted Nuclear Norm Minimization with Application to Image Denoising, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2014), 2862–2869.
- [12] M. Nikolova, A Variational Approach to Remove Outliers and Impulse Noise, *Journal of Mathematical Imaging and Vision*, **20:1–2** (2004), 99–120.
- [13] Y. Dong and T. Zeng, A Convex Variational Model for Restoring Blurred Images with Multiplicative Noise, *SIAM Journal on Imaging Sciences*, **6** (2013), 1598–1625.
- [14] T. Jeong, H. Woo and S. Yun, Frame-based Poisson image restoration using a proximal linearized alternating direction method, *Inverse Problems*, **29:7** (2013), 075007.
- [15] F. Sciacchitano, Y. Dong and T. Zeng, Variational approach for restoring blurred images with Cauchy noise, *SIAM Journal on Imaging Sciences*, **8:3** (2015), 1894–1922.
- [16] P. Bloomfield and W. Steiger, Least Absolute Deviations: Theory, Applications, and Algorithms, *Birkhäuser, Boston-Basel*, 1983.
- [17] C. Clason, B. Jin and K. Kunisch, A Duality-Based Splitting Method for L1-TV Image Restoration with Automatic Regularization Parameter Choice, *SIAM Journal on Scientific Computing*, **32** (2009), 1484–1505.
- [18] J. Yang, Y. Zhang and W. Yin, An Efficient TVL1 Algorithm for Deblurring Multichannel Images Corrupted by Impulsive Noise, *SIAM Journal on Scientific Computing*, **31** (2009), 2842–2865.
- [19] L. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica D*, **60** (1992), 259–26.,
- [20] A. Buades, B. Coll and J. M. Morel, A Non-Local Algorithm for Image Denoising, *Computer Vision and Pattern Recognition*, **2:7** (2005), 60–65.
- [21] R.H. Chan, C.W. Ho and M. Nikolova, Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization, *IEEE Transactions on Image Processing*, **14:10** (2005), 1479–1485.

- [22] R.H. Chan, C. Hu and M. Nikolova, An Iterative Procedure for Removing Random-Valued Impulse Noise, *IEEE Signal Processing Letters*, **11**:12 (2012), 921–924.
- [23] R. H. Chan, Y. Dong and M. Hintermüller, An efficient two-phase L1-TV method for restoring blurred images with impulse noise, *IEEE Transactions on Image Processing*, **19**:7 (2010), 1731–1739.
- [24] Y. Wang, W. Yin and J. Zeng, Global Convergence of ADMM in Nonconvex Nonsmooth Optimization, *Journal of Scientific Computing*, **78**:1 (2015), 29–63.
- [25] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, *Foundations and Trends in Machine Learning*, **3**:1 (2005), 1–122.
- [26] Z. Lin, R. Liu and H. Li, Linearized alternating direction method with parallel splitting and adaptive penalty for separable convex programs in machine learning, *Machine Learning*, **99**:2 (2015), 287–325.
- [27] R. Gonzalez and R. Woods, Digital Image Processing, *Addison-Wesley, Boston*, 1993.
- [28] D. Zoran and Y. Weiss, From learning models of natural image patches to whole image restoration, *Proceedings of the IEEE International Conference on Computer Vision*, (2011), 479–486.
- [29] A. Levin, B. Nadler, F. Durand and W. T. Freeman, Patch Complexity, Finite Pixel Correlations and Optimal Denoising, *Proceedings of the European Conference on Computer Vision*, (2012), 73–86.
- [30] H. Talebi and P. Milanfar, Global Image Denoising, *IEEE Transactions on Image Processing*, **23**:2 (2014), 755–768.
- [31] P. Chatterjee and P. Milanfar, Patch-Based Near-Optimal Image Denoising, *IEEE Transactions on Image Processing*, **21**:4 (2012), 1635–1649.
- [32] Z. Wang, A. Bovik, H. Sheikh and E. Simoncelli, Image Quality Assessment: From Error Visibility to Structural Similarity, *IEEE Transactions on Image Processing*, **13**:4 (2004), 600–612.
- [33] C. A. Micchelli, L. Shen, Y. Xu and X. Zeng, Proximity algorithms for the L1/TV image denoising model, *Advances in Computational Mathematics*, **38**:2 (2013), 401–426.
- [34] G. Yuan and B. Ghanem, L0TV: A Sparse Optimization Method for Impulse Noise Image Restoration, *Advances in Computational Mathematics*, **41**:2 (2019), 352–364.
- [35] J.F. Cai, R.H. Chan and M. Nikolova, Fast two-phase image deblurring under impulse noise, *Journal of Mathematical Imaging and Vision*, **36**:1 (2010), 46–53.
- [36] T. Chen and H. R. Wu, Adaptive impulse detection using center-weighted median filters, *IEEE Signal Processing Letters*, **8**:1 (2001), 1–3.