

## Journal Pre-proof

Image denoising based on the adaptive weighted TV<sup>P</sup> regularization

Zhi-Feng Pang, Hui-Li Zhang, Shousheng Luo, Tieyong Zeng

PII: S0165-1684(19)30378-0

DOI: <https://doi.org/10.1016/j.sigpro.2019.107325>

Reference: SIGPRO 107325

To appear in: *Signal Processing*

Received date: 6 February 2019

Revised date: 12 July 2019

Accepted date: 30 September 2019

Please cite this article as: Zhi-Feng Pang, Hui-Li Zhang, Shousheng Luo, Tieyong Zeng, Image denoising based on the adaptive weighted TV<sup>P</sup> regularization, *Signal Processing* (2019), doi: <https://doi.org/10.1016/j.sigpro.2019.107325>



This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 Published by Elsevier B.V.

**Highlights**

- This paper propose a new image denoising model by coupling the rotation matrix and the weighted matrix into the  $TV^p$  quasinorm regularization.
- Since the proposed model is nonconvex and non-Lipschitz, we employe the ADMM to solve it with the help of the half-quadratic scheme.
- The convergence of the proposed numerical methods is also discussed.
- Numerical comparisons show that the proposed model leaded to considerable performance gains when testing on several denoising tasks.



# Image denoising based on the adaptive weighted $TV^p$ regularization

Zhi-Feng Pang<sup>†,\*</sup>, Hui-Li Zhang<sup>‡</sup>, Shousheng Luo<sup>†</sup>, and Tieyong Zeng<sup>‡</sup>

<sup>†</sup> School of Mathematics and Statistics, Henan University, Kaifeng, China.

<sup>‡</sup> School of Medicine, Shanghai University, Shanghai, China

<sup>‡</sup> Department of Mathematics, The Chinese University of Hong Kong, Shatin, Hong Kong

## Abstract

Image denoising problem still remains an active research field in the image processing. To improve the denoising quality, it is very important to describe the local structure of the image in the proposed model. This fact motivates us to introduce an adaptive weighted  $TV^p$  regularization-based denoising model, where the rotation matrix and the weighted matrix depend on the local structure of the image. Specially, these two matrices can enhance the diffusion of the responding Euler-Lagrangian equation along with the tangential direction of the edge. This procedure offers more control over the regularization and then allows more denoising in smooth regions and less denoising when processing edge regions. In addition, since the proposed model is nonsmooth and non-Lipschitz, we employ the alternating direction method of multipliers (ADMM) to solve it with the help of using the half-quadratic scheme to solve the related  $\ell^2 - \ell^p$  subproblem. In particular, we also provide the convergence analysis of the used numerical methods. Some numerical comparisons show that the proposed model leads to considerable performance gains when tested on several denoising tasks.

**Keywords:** Image denoising; Adaptive weighted  $TV^p$ ; Half-quadratic scheme; Alternating direction method of multipliers (ADMM); Orientation field estimation (OFE).

## 1. Introduction

Image denoising plays an important role in the field of image processing. It is viewed as an inverse problem in which the clean image  $u : \Omega \rightarrow \mathbb{R}$  is estimated from the noisy observation  $f : \Omega \rightarrow \mathbb{R}$ . Here the image domain  $\Omega$  denotes a connected bounded open subset of  $\mathbb{R}^2$  with the Lipschitz boundary. However, the course of directly restoring noisy image is ill-posed due to the lacking of some prior information. Many regularization techniques [49, 3] have been presented to deal with this practical problem in the past decades. Specially, these techniques can be written as the following optimization problem

$$\min_u \left\{ \frac{\lambda}{2} \|u - f\|_2^2 + \mathcal{R}(u) \right\}, \quad (1)$$

where the fidelity term  $\|u - f\|_2^2$ <sup>1</sup> penalizes that the restored image  $u$  is very far away from the noisy observation  $f$ . The regularization term  $\mathcal{R}(u)$  mainly describes image priors such as continuity, smoothness or bounded variation. The regularization parameter  $\lambda$  can balance the effect of the data fitting term and the regularization term.

Email address: zhibfengpang@163.com (\*Corresponding author); huilizhang92@163.com; sluo@henu.edu.cn; zeng@math.cuhk.edu.hk. (Zhi-Feng Pang<sup>†,\*</sup>, Hui-Li Zhang<sup>‡</sup>, Shousheng Luo<sup>†</sup>, and Tieyong Zeng<sup>‡</sup>)

<sup>1</sup>This paper only considers to restore the image contaminated by the additive white Gaussian noise. Based on the classical maximum a posteriori (MAP) scheme [11], it is easy to get the reasonability of using the  $\ell^2$ -norm as the fitting term.

In the model (1), how to choose an effective regularization term  $\mathcal{R}(u)$  is a key to improve the denoising ability. Toward this end, many efforts have been proposed through seeking some suitable image priors to capture intrinsic characteristics such as learning-based models [33, 25, 58, 59, 18, 39], wavelet-based models [16, 51, 60], and variation-based models [49, 3, 11, 6, 38, 40, 19]. Here we mainly focus on variation-based models, especially for the models based on the total variation (TV) regularization. Since their introduction in a classic paper by Rudin, Osher and Fatemi [47], the TV-based models have become one of the most popular and successful methodology for image restoration, inpainting, blind deconvolution and vector-valued images [43, 37, 36, 8, 6, 27]. Although these TV-based models were reported to give better restored performance, they still may not be ideally suited for the regularizer when describing local details of the image. One reason is that the difference grid of the classic TV-based models only depends on the horizontal and perpendicular directions in the numerical implementation. This choice is impossible to keep the diffusion of the corresponding Euler-Lagrangian equation along with the tangential direction of the edge. In order to overcome above drawback, the directional total variation (DTV)-based model [5] introduced one rotation matrix to rotate the gradient direction. However, when there exists several dominant directions, the DTV-based model will fail since one fixed angle parameter can not simultaneously describe them. This fact motivates us to find some schemes [57, 35] that the choosing angles adaptively depend on the local structure of the image. That is to say, the proposed model can simultaneously handle several dominant directions. To this end, this paper proposes a new image denoising model with the following novelties:

- We present a robust and efficient  $\ell^2 - TV^p$  ( $p \in (0, 1)$ ) denoising model, in which the regularization term uses a combination of the rotation matrix, the weighted matrix and the  $TV^p$ -quasinorm. In order to enhance the diffusion along with the tangential direction of the edge, the rotation matrix and the weighted matrix mainly depend on the edge information. With these settings, the proposed model has more robust adaptivity and then owns more strong abilities of suppressing strong noise, estimating and restoring complex structures.
- To deal with the nonsmoothing and the non-Lipschitz, one simple and effective numerical method is devised through using the ADMM to split the proposed model into several solvable subproblems. Especially, to keep the accuracy of the corresponding  $\ell^2 - \ell^p$  subproblem, we use the half-quadratic scheme to transform it into the multivariable optimization problem and then use the alternating direction method (ADM) to solve it. We also give the convergence in order to keep the feasibility and the stability of the proposed methods.

The rest of the paper is organized as follows. Section 2 mainly recalls some TV-based models related to our proposed model. Section 3 presents the proposed model and its numerical method. We also discuss the convergence of the proposed method in this section. We provide some details on how to choose the related parameters for the proposed model and method in Section 4. Section 5 gives some numerical comparisons to show the effectiveness of the proposed model. Concluding remarks are given in the last section.

**Notations:** Here we describe some notations used throughout this paper. To simplify them, we set  $X = \mathbb{R}^{m \times l}$  and  $Y = X \times X$ .  $\langle \cdot, \cdot \rangle$  denotes the usual scalar product as  $\langle \bar{u}, \bar{v} \rangle = \sum_{i=1}^m \sum_{j=1}^l \bar{u}_{i,j} \bar{v}_{i,j}$  and the  $\ell^2$  norm as  $\|\bar{u}\|_2^2 := \sum_{i=1}^m \sum_{j=1}^l \bar{u}_{i,j}^2$  for  $\bar{u}, \bar{v} \in X$ . For  $\bar{\mathbf{g}} \in Y$  and  $\bar{\mathbf{h}} \in Y$ , the scalar product is defined by  $\langle \bar{\mathbf{g}}, \bar{\mathbf{h}} \rangle = \sum_{i=1}^m \sum_{j=1}^l \sum_{s=1}^2 \bar{g}_{i,j,s} \bar{h}_{i,j,s}$ , and then the responding  $\ell^p$ -(quasi)norm is defined by  $\|\bar{\mathbf{g}}\|_{2,p}^p := \left( \sum_{i=1}^m \sum_{j=1}^l \left( \sqrt{\bar{g}_{i,j,1}^2 + \bar{g}_{i,j,2}^2} \right)^p \right)^{1/p}$  for  $p \in (0, \infty)$ . For the matrix operator  $A$ , its spectral norm is defined by  $\|A\|_2 := \max_{\mathbf{z}} \{ \|A\mathbf{z}\|_2 : \mathbf{z} \in \mathbb{R}^n \text{ with } \|\mathbf{z}\|_2 \leq 1 \}$ .

## 2. Related works

Because our work is closely related to the TV regularization, this section mainly reviews some TV-based models for the image denoising problem.

### 2.1. ROF model

The ROF model proposed by Rudin, Osher, and Fatemi [47] has been used in many image processing problems such as image restoration, image inpainting, medical image reconstruction, etc. This model can be written as

$$\min_u \left\{ \frac{\lambda}{2} \|u - f\|_2^2 + \|\nabla u\|_{2,1} \right\}, \quad (2)$$

where the second term (called the TV) discourages the solution to be oscillations or discontinuities. So this model can treat edges efficiently. However, this model usually supplies piecewise constant images (called the staircasing artifact) since the diffusion in the local normal direction is always zero [41], that is to say, smooth or slant regions in original image are recovered as piecewise smooth regions [30].

### 2.2. High-order TV-based model (HOTV)

In order to overcome the staircasing artifact for the ROF model, several models based on the high-order total variation (HOTV) such as the Laplacian penalty [48, 10], the anisotropic second order regularization [43], the Hessian Schatten-norm regularization [37] have been introduced. These models were reported to give better denoising performance than the standard TV-based regularization for preserving the smooth regions. In fact, the main reason is that piecewise-vanishing second-order derivatives lead to piecewise-linear solutions that better fit smooth intensity changes [36]. One example of the HOTV was proposed in [43, 10] as

$$\min_u \left\{ \frac{\lambda}{2} \|u - f\|_2^2 + \|\nabla^2 u\|_{F,1} \right\}, \quad (3)$$

where the inner norm of  $\|\cdot\|_{F,1}$  denotes the Frobenius norm and the outer norm is the  $\ell^1$  norm. Specially, it can be written by

$$\|\nabla^2 u\|_{F,1} = \sum_{i=1}^m \sum_{j=1}^l \sqrt{(u_{xx}(i, j))^2 + (u_{xy}(i, j))^2 + (u_{yx}(i, j))^2 + (u_{yy}(i, j))^2}.$$

However, this model often blurs image edges.

### 2.3. Total generalized variation (TGV)-based model

As the extension of the TV-based models, the total generalized variation (TGV)-based models have been used in many fields of the image processing. The original TGV model [8] was defined in the space of symmetric tensors of order  $k$ . However, there exist some numerical difficulties due to these tensors. In order to solve the TGV-based model efficiently, one work in [27] used the Legendre-Fenchel transform to consider the dual form as

$$\min_{u,w} \left\{ \frac{\lambda}{2} \|u - f\|_2^2 + \alpha_0 \|\nabla u - w\|_{2,1} + \alpha_1 \|\mathcal{E}(w)\|_{F,1} \right\}, \quad (4)$$

where  $\alpha_0$  and  $\alpha_1$  are the positive parameters,  $\mathcal{E}(w)$  is defined by

$$\mathcal{E}(w) = \frac{1}{2} \begin{bmatrix} 2w_1 & (w_1 + w_2) \\ (w_1 + w_2) & 2w_2 \end{bmatrix} \quad (5)$$

for  $w = (w_1, w_2)^T$ . The main advantage of the TGV model is the ability to reconstruct piecewise polynomial functions. So it outperforms the ROF model and the HOTV model for denoising piecewise affine signals.

#### 2.4. Directional total variation (DTV) model

In many applications related to **fibres**, the textures in images have very clear directionality. Examples include glass fibers in wind-turbine blades, optical **fibres** for communication, and ceramic **fibres** in fuel cells [34]. So it is very important to describe the geometric feature for the proposed model. However, the difference scheme of the gradient in above models (2)-(4) mainly depends on the horizontal and perpendicular directions in the numerical implementation. Then the proposed model based on this scheme can not couple with local structures of the image efficiently. One method to overcome this drawback is to rotate the gradient operator by using the rotation matrix. In addition, the weighted matrix can be added for enhancing the diffusion of the corresponding Euler-Lagrangian equation of the proposed model. One early work [5] based on this motivation was of considering

$$\min_u \left\{ \frac{\lambda}{2} \|u - f\|_2^2 + \|\Lambda_\beta \mathbf{R}_{-\theta} \nabla u\|_{2,1} \right\}, \quad (6)$$

where the weighted matrix  $\Lambda_\beta$  and the rotation matrix  $\mathbf{R}_{-\theta}$  are defined by

$$\Lambda_\beta = \begin{bmatrix} \beta & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \mathbf{R}_{-\theta} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}.$$

Here  $\theta$  denotes the affine angle and  $\beta > 1$  is a weighted parameter. Obviously, this choice can inherit the desirable properties of the classical gradient-based penalty (e.g., feature-preserving smoothing, invariance to translation, convexity, and simplicity). **However, this model is not suitable for restoring the image with complex structures since using fixed affine angle  $\theta$  lacks the local adaptivity.**

### 3. Adaptive weighted total variation $\ell^p$ -quasinorm (AWTVP) model

In general, the image have different structures in different regions. The tangential direction of the edge is one of main tools to describe them. **To this end, the proposed model needs to adaptively rotate the gradient to couple with the tangential direction and the weighted matrix to enhance the diffusion.** However, since the noise seriously affects the direction recognition, we here employ the method used in [29] to estimate the local edge orientation. **In addition, we use a  $\ell^p$ -quasinorm regularization based on TV-term to promote gradient sparsity by setting  $p \in (0, 1)$ , which allows for sharp discontinuities in the original signal and has received many attentions [31, 56, 54, 9].** Specially, we propose an adaptive weighted  $TV^p$ -quasinorm-based (AWTVP) model as

$$\min_u \left\{ \frac{\lambda}{2} \|u - f\|_2^2 + \|\mathbf{T}_\beta(f) \mathbf{R}_{-\theta(f)} \nabla u\|_{2,p}^p \right\}, \quad (7)$$

where

$$\mathbf{R}_{-\theta(f)} = \begin{bmatrix} \cos \theta(f) & \sin \theta(f) \\ -\sin \theta(f) & \cos \theta(f) \end{bmatrix} \text{ and } \mathbf{T}_\beta(f) = \begin{bmatrix} t_{1,\beta}(f) & 0 \\ 0 & t_2(f) \end{bmatrix} := \begin{bmatrix} \frac{\beta(f)}{1+\kappa|G_{\hat{\sigma}} * \nabla_x f|^2} & 0 \\ 0 & \frac{1}{1+\kappa|G_{\hat{\sigma}} * \nabla_y f|^2} \end{bmatrix}. \quad (8)$$

Here  $\theta(f)$  denotes that the angle  $\theta$  depends on the noisy image  $f$ ,  $G_{\hat{\sigma}}(\cdot)$  denotes the Gaussian convolution with the standard deviation  $\hat{\sigma}$  in order to reduce the effect of the noise,  $\kappa$  is a tuning parameter. **In the matrix  $\mathbf{T}_\beta(f)$ ,  $\beta(f)$  is used to indicate the smoothing and edge regions in order to describe the diffusion efficiently.** In particular, we set  $\beta(f) = \beta$  when  $\|G_{\hat{\sigma}} * \nabla f\|_2 > b$  and  $\beta(f) = 1$  for  $\|G_{\hat{\sigma}} * \nabla f\|_2 \leq b$ .

#### 3.1. ADMM to solve the model (7).

It can be seen that solving the proposed model (7) has many numerical difficulties since it is a large scale and non-convex optimization problem. One efficient numerical method is the alternating direction method of multipliers (ADMM), which was originally introduced in early 1970s [23, 20] and has been recently used to solve high-dimensional signal processing problems such as machine learning, computer vision, image and signal processing, networking. This method is closely related to many other methods such as dual decomposition, the method of multipliers, Douglas-Rachford splitting, Spingarns method of partial inverses, Dykstras alternating projections, see for instance of the works in [7, 24]. **The basic motivation of the ADMM is to decomposes a large global problem into a series of smaller local subproblems, and then uses the local solutions to compute the solution of the original problem.**

### 3.1.1. Alternating direction method

In order to use the ADMM to solve the model (7), we first need to consider how to solve the  $\ell^2 - \ell^p$  problem. This problem is nonsmooth and non-Lipschitz, so some schemes based on the smooth technique are employed to overcome these numerical difficulties, such as iteratively reweighted least squares (IRLS) [45, 14, 46], iteratively reweighted  $\ell^1$ -minimization (IR $\ell^1$ ) [13] and iteratively thresholding  $\ell^p$  method (ITM $\ell^p$ ) [50]. However, the fixed smooth parameter used in these schemes implies that we only obtain the approximated solution. In order to obtain the exact solution, this subsection proposes a new scheme through combining a half-quadratic scheme [9, 44, 22] with the continuation method [12]. In particular, the  $\ell^2 - \ell^p$  problem can be formed as

$$\min_{\mathbf{x}} \left\{ \frac{\tau}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 + \|\mathbf{x}\|_{2,p}^p \right\} \quad (9)$$

for  $p \in (0, 1)$  and  $\mathbf{y} \in Y$ . In order to solve it efficiently, we extend the half-quadratic scheme [9] to the vector space  $Y$ .

**Lemma 3.1.** *For any  $p \in (0, 1)$  and  $t \in Y \setminus \{(0, 0)^\top\}$ , then we have*

$$\|t\|_{2,p}^p = \min_{w>0} \left\{ w\|t\|_2^2 + \frac{1}{\delta w^\gamma} \right\},$$

where  $\gamma = \frac{p}{2-p}$  and  $\delta = \frac{2^{\frac{2}{2-p}}}{(2-p)p^{\frac{p}{2-p}}}$ . Furthermore, the minimizer is reached at

$$w^* = \frac{p}{2} \|t\|_2^{p-2} = \frac{p}{2} (t_1^2 + t_2^2)^{\frac{p-2}{2}}.$$

In the problem (9), the variable  $\mathbf{x}$  may be the zero vector. In order to use Lemma 3.1, we need to smooth it by introducing a small parameter  $\varepsilon > 0$  as

$$\min_{\mathbf{x}} \left\{ \mathcal{F}(\mathbf{x}) := \sum_{i=1}^m \sum_{j=1}^l \underbrace{\left( \sqrt{\|\mathbf{x}_{i,j}\|_2^2 + \varepsilon} \right)^p}_{:= \|\mathbf{x}_{i,j}\|_\varepsilon^p} + \frac{\tau}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 \right\}. \quad (10)$$

Then we have

$$\min_{\mathbf{x}, \mathbf{w} > 0} \left\{ \mathcal{G}(\mathbf{w}, \varepsilon, \mathbf{x}) = \sum_{i=1}^m \sum_{j=1}^l \left( \mathbf{w}_{i,j} \|\mathbf{x}_{i,j}\|_\varepsilon^2 + \frac{1}{\delta \mathbf{w}_{i,j}^\gamma} + \frac{\tau}{2} \|\mathbf{x}_{i,j} - \mathbf{y}_{i,j}\|_2^2 \right) := \sum_{i=1}^m \sum_{j=1}^l G(\mathbf{w}_{i,j}, \varepsilon, \mathbf{x}_{i,j}) \right\}, \quad (11)$$

where  $\gamma$  and  $\delta$  are defined in Lemma 3.1. The problem (11) is a multi-variable optimization problem, so the alternating direction method (ADM) can be employed to solve it as shown in Algorithm 1.

---

#### Algorithm 1: ADM to solve the problem (11).

---

1. Input: original values of  $\mathbf{x}^0, \varepsilon^0, \tau^0$ ;
2. Iterate (12a)-(12c) below until stopping criterion is met

$$\begin{cases} \mathbf{w}^{n+1} := \underset{\mathbf{w}>0}{\operatorname{argmin}} \mathcal{G}(\mathbf{w}, \varepsilon^n, \mathbf{x}^n), & (12a) \\ \varepsilon^{n+1} := \rho \varepsilon^n, & (12b) \\ \mathbf{x}^{n+1} := \underset{\mathbf{x}}{\operatorname{argmin}} \mathcal{G}(\mathbf{w}^{n+1}, \varepsilon^{n+1}, \mathbf{x}); & (12c) \end{cases}$$

where  $\rho \in (0, 1)$  as a contractive constant;

3. Output: Set  $\mathbf{x} := \mathbf{x}^{n+1}$ .
- 

**Remark 3.1.** *Algorithm 1 devolves into the fixed method if we set  $\rho = 1$ . In this case, the general iteration method can be used to solve it. However, our main motivation is of using the gradually approximated method to theoretically*

obtain an exact solution of the problem (11). To this end, we introduce the dynamically decreased technique as shown in (12b). In addition, using Lemma 3.1, the solutions of (12a) and (12c) can be explicitly obtained by

$$\begin{cases} \mathbf{w}^{n+1} = \frac{p}{2} \|\mathbf{x}^n\|_{\varepsilon^n}^{p-2}, \\ \mathbf{x}^{n+1} = \frac{\tau \mathbf{y}}{2\mathbf{w}^{n+1} + \tau}. \end{cases} \quad (13)$$

Furthermore, (12a) and (12c) also imply that

$$\begin{aligned} \mathcal{G}(\mathbf{w}^{n+1}, \varepsilon^{n+1}, \mathbf{x}^{n+1}) - \mathcal{G}(\mathbf{w}^n, \varepsilon^n, \mathbf{x}^n) &= [\mathcal{G}(\mathbf{w}^{n+1}, \varepsilon^{n+1}, \mathbf{x}^{n+1}) - \mathcal{G}(\mathbf{w}^{n+1}, \varepsilon^{n+1}, \mathbf{x}^n)] \\ &\quad + [\mathcal{G}(\mathbf{w}^{n+1}, \varepsilon^{n+1}, \mathbf{x}^n) - \mathcal{G}(\mathbf{w}^{n+1}, \varepsilon^n, \mathbf{x}^n)] + [\mathcal{G}(\mathbf{w}^{n+1}, \varepsilon^n, \mathbf{x}^n) - \mathcal{G}(\mathbf{w}^n, \varepsilon^n, \mathbf{x}^n)] < 0. \end{aligned}$$

Then we can deduce that the sequence  $\mathcal{G}(\mathbf{w}^{n+1}, \varepsilon^{n+1}, \mathbf{x}^{n+1})$  is convergent since  $\mathcal{G}(\mathbf{w}, \varepsilon, \mathbf{x})$  is bounded below by zero.

In the following, we give the convergence analysis of Algorithm 1.

**Theorem 3.1.** *In Algorithm 1, we assume that there exists a constant  $M$  such that  $\|\mathbf{w}^{n+1}\|_2 < M$ , then the sequence  $\{\mathbf{x}^n\}$  converges to one of minimizers of the problem (9).*

*Proof.* Since the problem (12a) is separable, we only consider the case at the  $(i, j)$ -position. For convenience, we set  $\bar{\mathbf{w}} := \mathbf{w}_{i,j}$  and  $\bar{\mathbf{x}} := \mathbf{x}_{i,j}$ . With this setting, the separable subproblem of (12a) is equivalent to consider

$$\bar{\mathbf{w}}^{n+1} := \underset{\bar{\mathbf{w}} > 0}{\operatorname{argmin}} \bar{\mathbf{w}} \|\bar{\mathbf{x}}^n\|_{\varepsilon^n}^2 + \frac{1}{\delta \bar{\mathbf{w}}^\gamma} := g(\bar{\mathbf{w}}),$$

which implies that  $\bar{\mathbf{w}}^{n+1} = \frac{p}{2} \|\bar{\mathbf{x}}^n\|_{\varepsilon^n}^{p-2}$ ,  $g'(\bar{\mathbf{w}}^{n+1}) = 0$  and  $g''(\bar{\mathbf{w}}) = \frac{\gamma(\gamma+1)}{\delta \bar{\mathbf{w}}^{\gamma+2}}$ . Using the Taylor expansion to  $G(\bar{\mathbf{w}}, \varepsilon^n, \bar{\mathbf{x}}^n)$  at  $\bar{\mathbf{w}}^{n+1}$ , we can deduce that

$$G(\bar{\mathbf{w}}^n, \varepsilon^n, \bar{\mathbf{x}}^n) - G(\bar{\mathbf{w}}^{n+1}, \varepsilon^n, \bar{\mathbf{x}}^n) = g'(\bar{\mathbf{w}}^{n+1})(\bar{\mathbf{w}}^n - \bar{\mathbf{w}}^{n+1}) + \frac{1}{2} g''(\bar{\mathbf{w}}_n) \|\bar{\mathbf{w}}^n - \bar{\mathbf{w}}^{n+1}\|_2^2$$

for  $\bar{\mathbf{w}}_n \in (\bar{\mathbf{w}}^n, \bar{\mathbf{w}}^{n+1})$ . Based on Remark 3.1, we then have

$$\lim_{n \rightarrow \infty} g''(\bar{\mathbf{w}}_n) \|\bar{\mathbf{w}}^n - \bar{\mathbf{w}}^{n+1}\|_2^2 = 0. \quad (14)$$

Now we discuss the limitation of (14) along with two folds. If  $\lim_{n \rightarrow \infty} \|\bar{\mathbf{w}}^n - \bar{\mathbf{w}}^{n+1}\|_2^2 \neq 0$ , it implies that  $g''(\bar{\mathbf{w}}_n) = \frac{\gamma(\gamma+1)}{\delta \bar{\mathbf{w}}_n^{\gamma+2}} \rightarrow 0$  as  $n \rightarrow \infty$ . Then we have  $\bar{\mathbf{w}}_n \rightarrow \infty$ . However, it does not hold for this choice. In fact, setting  $\widehat{\mathbf{w}}_n = \bar{\mathbf{w}}^n + \zeta(\bar{\mathbf{w}}^{n+1} - \bar{\mathbf{w}}^n)$  if  $\bar{\mathbf{w}}^{n+1} > \bar{\mathbf{w}}^n$  with  $\zeta \in (0, 1)$ , we deduce that

$$\widehat{\mathbf{w}}_n = \frac{p(1-\zeta)}{2} \|\bar{\mathbf{x}}^{n-1}\|_{\varepsilon^{n-1}}^{p-2} + \frac{p\zeta}{2} \|\bar{\mathbf{x}}^n\|_{\varepsilon^n}^{p-2} \rightarrow \infty \quad (15)$$

when  $n \rightarrow \infty$ . Then there is a contradiction between (15) and the assumption  $\|\mathbf{w}^{n+1}\|_2 < M$ . To the case that  $\widehat{\mathbf{w}}_n = \bar{\mathbf{w}}^{n+1} + \zeta(\bar{\mathbf{w}}^n - \bar{\mathbf{w}}^{n+1})$  if  $\bar{\mathbf{w}}^n > \bar{\mathbf{w}}^{n+1}$ , we can get the same contradiction. So we have

$$\lim_{n \rightarrow \infty} \|\bar{\mathbf{w}}^n - \bar{\mathbf{w}}^{n+1}\|_2^2 = 0. \quad (16)$$

In addition, if set  $\nabla \mathcal{F}(\bar{\mathbf{x}}) = \sum_{i=1}^m \sum_{j=1}^l F(\mathbf{x}_{i,j})$ , the gradient of the objective function in (10) at  $\bar{\mathbf{x}}^{n+1}$  can be expressed as

$$\begin{aligned} \nabla F(\bar{\mathbf{x}}^{n+1}(\varepsilon^{n+1})) &= 2\bar{\mathbf{w}}^{n+2} \bar{\mathbf{x}}^{n+1} + \tau(\bar{\mathbf{x}}^{n+1} - \bar{\mathbf{y}}) \\ &= 2(\bar{\mathbf{w}}^{n+2} - \bar{\mathbf{w}}^{n+1}) \bar{\mathbf{x}}^{n+1} + (2\bar{\mathbf{w}}^{n+1} + \tau) \bar{\mathbf{x}}^{n+1} - \tau \bar{\mathbf{y}} \\ &= 2(\bar{\mathbf{w}}^{n+2} - \bar{\mathbf{w}}^{n+1}) \bar{\mathbf{x}}^{n+1}, \end{aligned} \quad (17)$$

where  $\bar{\mathbf{x}}^{n+1}(\varepsilon^{n+1})$  denotes that  $\bar{\mathbf{x}}^{n+1}$  depends on  $\varepsilon^{n+1}$ . Using (16) and (13), we have

$$\lim_{n \rightarrow \infty} \nabla F(\bar{\mathbf{x}}^n(\varepsilon^n)) = 0.$$

This implies that the sequence  $\{\bar{\mathbf{x}}^n\}$  converges to one of minimizers of the problem (9) at the  $(i, j)$ -position since  $\varepsilon^n \rightarrow 0$  as  $n \rightarrow \infty$ . Then we can deduce the convergence of the sequence  $\{\mathbf{x}\}$  based on the separability of the problem (9).  $\square$

**Remark 3.2.** If  $\mathbf{x}^*$  is the (global/local) minimizer of the problem (9), then the optimization condition can be written as

$$0 = \begin{cases} p\|\mathbf{x}^*\|_{2,p}^{p-1}\mathbf{x}^* + \tau(\mathbf{x}^* - \mathbf{y}), & \text{if } p \neq 1, \\ \xi^* + \tau(\mathbf{x}^* - \mathbf{y}), & \text{if } p = 1, \end{cases}$$

where  $\xi^* \in \partial(\|\mathbf{x}^*\|_{2,1})$  denotes the subdifferential of  $\|\mathbf{x}^*\|_{2,1}$  at  $\mathbf{x}^*$ . In addition, we also need to explain the reason for using  $\mathbf{w}^{n+2}$  in the first equation of (17). This comes from the equivalency between (10) and (11) when using Lemma 3.1 and Algorithm 1. That is to say, when using  $\mathbf{w}^{n+1}$  for  $\mathcal{F}(\mathbf{x}^{n+1})$ , we need to set  $\mathbf{w} := \mathbf{w}^{n+2}$  corresponding to  $\mathbf{x}^{n+1}$ .

### 3.2. The alternating direction method of multipliers (ADMM).

This subsection presents the ADMM to solve the model (7) through iteratively updating the multiple primal variables and the corresponding dual variables of the related augmented Lagrangian function. Specifically, we introduce auxiliary variables  $\mathbf{v}$  and  $\mathbf{w}$ , and then reformulate the model (7) into the constraint optimization problem as

$$\begin{cases} \min_{\mathbf{w}, \mathbf{v}, u} \left\{ \frac{\lambda}{2} \|u - f\|_2^2 + \|\mathbf{w}\|_{2,p}^p \right\}, \\ \text{s.t. } \mathbf{v} := (v_1, v_2)^T = \nabla u \text{ and } \mathbf{w} := (w_1, w_2)^T = \mathbf{T}_\beta(f)\mathbf{R}_{-\theta(f)}\mathbf{v}. \end{cases} \quad (18)$$

In order to solve the problem (18), we introduce two Lagrange multipliers  $\alpha = (\alpha_1, \alpha_2)^T$  and  $\varrho = (\varrho_1, \varrho_2)^T$ , and then reformulate it as a saddle point problem

$$\begin{aligned} \min_{\mathbf{w}, \mathbf{v}, u} \max_{\alpha, \varrho} \mathcal{L}(\mathbf{w}, \mathbf{v}, u, \alpha, \varrho) = & \frac{\lambda}{2} \|u - f\|_2^2 + \langle \alpha, \mathbf{v} - \nabla u \rangle + \frac{\gamma_1}{2} \|\mathbf{v} - \nabla u\|_2^2 \\ & + \|\mathbf{w}\|_{2,p}^p + \langle \varrho, \mathbf{w} - \mathbf{T}_\beta(f)\mathbf{R}_{-\theta(f)}\mathbf{v} \rangle + \frac{\gamma_2}{2} \|\mathbf{w} - \mathbf{T}_\beta(f)\mathbf{R}_{-\theta(f)}\mathbf{v}\|_2^2, \end{aligned} \quad (19)$$

where  $\mathcal{L}(\mathbf{w}, \mathbf{v}, u, \alpha, \varrho)$  denotes the augmented Lagrangian function,  $\gamma_1$  and  $\gamma_2$  are the penalty parameters. Since the variables in the problem (19) are coupled together, it is difficult to solve them simultaneously. Therefore, we need to solve one of variables and simultaneously fix others when using the ADMM as shown in Algorithm 2.

---

#### Algorithm 2: ADMM to solve the problem (19)

---

1. Input: original values of  $\mathbf{v}^0, \alpha^0$  and  $\varrho^0, \lambda$ ;

2. Iterate (20a)-(20e) below until stopping criterion is met

$$\mathbf{w}^{k+1} := \underset{\mathbf{w}}{\operatorname{argmin}} \mathcal{L}(\mathbf{w}, \mathbf{v}^k, u^k, \alpha^k, \varrho^k), \quad (20a)$$

$$\mathbf{v}^{k+1} := \underset{\mathbf{v}}{\operatorname{argmin}} \mathcal{L}(\mathbf{w}^{k+1}, \mathbf{v}, u^k, \alpha^k, \varrho^k), \quad (20b)$$

$$u^{k+1} := \underset{u}{\operatorname{argmin}} \mathcal{L}(\mathbf{w}^{k+1}, \mathbf{v}^{k+1}, u, \alpha^k, \varrho^k), \quad (20c)$$

$$\alpha^{k+1} = \alpha^k + \gamma_1 (\mathbf{v}^{k+1} - \nabla u^{k+1}), \quad (20d)$$

$$\varrho^{k+1} = \varrho^k + \gamma_2 (\mathbf{w}^{k+1} - \mathbf{T}_\beta(f)\mathbf{R}_{-\theta(f)}\mathbf{v}^{k+1}); \quad (20e)$$

3. Output:  $u := u^{k+1}$  as the restored image.

---

In Algorithm 2, the efficiency relies on how to solve the subproblems with the high accuracy. Explicit solutions to the above sub-problems (20a)-(20c) are the following.

(1). **Subproblem (20a).** This subproblem can be expressed as

$$\mathbf{w}^{k+1} = \operatorname{argmin}_{\mathbf{w}} \|\mathbf{w}\|_{2,p}^p + \frac{\gamma_2}{2} \left\| \mathbf{w} - \left( \mathbf{T}_\beta(f) \mathbf{R}_{-\theta(f)} \mathbf{v}^k - \frac{\boldsymbol{\varrho}^k}{\gamma_2} \right) \right\|_2^2,$$

which is nonsmooth and non-Lipschitz. In order to use Algorithm 1, we need to smooth it as

$$\bar{\mathbf{w}}_n^{k+1} = \operatorname{argmin}_{\mathbf{w}} \|\mathbf{w}\|_\varepsilon^p + \frac{\gamma_2}{2} \left\| \mathbf{w} - \left( \mathbf{T}_\beta(f) \mathbf{R}_{-\theta(f)} \mathbf{v}^k - \frac{\boldsymbol{\varrho}^k}{\gamma_2} \right) \right\|_2^2 \quad (21)$$

by introducing a small positive parameter  $\varepsilon$ , where  $\|\mathbf{w}\|_\varepsilon^p := (\|\mathbf{w}\|_2^2 + \varepsilon)^{\frac{p}{2}}$ . In this case, the first term is nonzero, so Algorithm 1 can be employed. Furthermore, based on Theorem 3.1, we can obtain one of local solutions to the subproblem (20a) as  $\bar{\mathbf{w}}_n^{k+1} \rightarrow \mathbf{w}^{k+1}$  when  $k \rightarrow \infty$ .

(2). **Subproblem (20b).** This subproblem can be reformulated as

$$\mathbf{v}^{k+1} = \operatorname{argmin}_{\mathbf{v}} \langle \boldsymbol{\alpha}^k, \mathbf{v} \rangle + \frac{\gamma_1}{2} \|\mathbf{v} - \nabla u^k\|_2^2 - \langle \boldsymbol{\varrho}^k, \mathbf{T}_\beta(f) \mathbf{R}_{-\theta(f)} \mathbf{v} \rangle + \frac{\gamma_2}{2} \|\mathbf{w}^{k+1} - \mathbf{T}_\beta(f) \mathbf{R}_{-\theta(f)} \mathbf{v}\|_2^2,$$

which optimization condition can be written as

$$(\gamma_1 \mathcal{I} + \gamma_2 \mathbf{R}_{\theta(f)} \mathbf{T}_\beta(f) \mathbf{T}_\beta(f) \mathbf{R}_{-\theta(f)}) \mathbf{v}^{k+1} = (\gamma_1 \nabla u^k + \mathbf{R}_{\theta(f)} \mathbf{T}_\beta(f) \boldsymbol{\varrho}^k + \gamma_2 \mathbf{R}_{\theta(f)} \mathbf{T}_\beta(f) \mathbf{w}^{k+1} - \boldsymbol{\alpha}^k), \quad (22)$$

where  $\mathcal{I}$  is a unit operator. With the simple computation, above linear equation system can be rearranged as

$$\begin{cases} a_{11} v_1^{k+1} + a_{12} v_2^{k+1} = b_1^k, \\ a_{21} v_1^{k+1} + a_{22} v_2^{k+1} = b_2^k, \end{cases}$$

where

$$\begin{aligned} a_{11} &= \gamma_1 + \gamma_2 [t_{1,\beta}^2(f) \cos^2 \theta(f) + t_2^2(f) \sin^2 \theta(f)], & a_{12} &= \gamma_2 [t_{1,\beta}^2(f) - t_2^2(f)] \sin \theta(f) \cos \theta(f), \\ a_{22} &= \gamma_1 + \gamma_2 [t_{1,\beta}^2(f) \sin^2 \theta(f) + t_2^2(f) \cos^2 \theta(f)], & a_{21} &= \gamma_2 [t_{1,\beta}^2(f) - t_2^2(f)] \sin \theta(f) \cos \theta(f), \\ b_1^k &= \boldsymbol{\varrho}_1^k t_{1,\beta}(f) \cos \theta(f) - \boldsymbol{\varrho}_2^k t_2(f) \sin \theta(f) + \gamma_2 [w_1^{k+1} t_{1,\beta}(f) \cos \theta(f) - w_2^{k+1} t_2(f) \sin \theta(f)] - \boldsymbol{\alpha}_1^k + \gamma_1 \nabla_x u^k, \\ b_2^k &= \boldsymbol{\varrho}_1^k t_{1,\beta}(f) \sin \theta(f) + \boldsymbol{\varrho}_2^k t_2(f) \cos \theta(f) + \gamma_2 [w_1^{k+1} t_{1,\beta}(f) \sin \theta(f) + w_2^{k+1} t_2(f) \cos \theta(f)] - \boldsymbol{\alpha}_2^k + \gamma_1 \nabla_y u^k. \end{aligned}$$

Since the coefficient matrix of the equations (22) is nonsingular, that is to say  $\det(\gamma_1 \mathcal{I} + \gamma_2 \mathbf{R}_{\theta(f)} \mathbf{T}_\beta(f) \mathbf{T}_\beta(f) \mathbf{R}_{-\theta(f)}) \neq 0$ , so the solution  $\mathbf{v}^{k+1} = (v_1^{k+1}, v_2^{k+1})^T$  can be obtained by using the Cramer's rule as

$$\begin{cases} v_1^{k+1} = \frac{b_1^k a_{22} - b_2^k a_{12}}{a_{11} a_{22} - a_{12} a_{21}}, \\ v_2^{k+1} = \frac{b_2^k a_{11} - b_1^k a_{21}}{a_{11} a_{22} - a_{12} a_{21}}. \end{cases} \quad (23)$$

(3). **Subproblem (20c).** This subproblem is a smooth and convex optimization problem as

$$u^{k+1} = \operatorname{argmin}_u \frac{\lambda}{2} \|u - f\|_2^2 - \langle \boldsymbol{\alpha}^k, \nabla u \rangle + \frac{\gamma_1}{2} \|\mathbf{v}^{k+1} - \nabla u\|_2^2.$$

So the Euler-Lagrangian equation can be obtained by using the variational method as

$$(\lambda \mathcal{I} - \gamma_1 \Delta) u^{k+1} = \lambda f - \operatorname{div} \boldsymbol{\alpha}^k - \gamma_1 \operatorname{div} \mathbf{v}^{k+1}. \quad (24)$$

Choosing different boundary conditions, solving the linear equation system (24) is corresponded to different numerical methods. In fact, we shall notice that the Laplacian operator  $\Delta$  is negative semidefinite when using the

zero Neumann boundary condition or the zero Dirichlet boundary condition. In this case, the preconditioned conjugate gradient (PCG) method can be used. However, this paper assumes to use the periodic boundary condition, so the solution can be computed efficiently by the fast Fourier transform

$$u^{k+1} = \mathcal{F}^{-1} \left( \frac{\mathcal{F}(\lambda f - \operatorname{div} \alpha^k - \gamma_1 \operatorname{div} \mathbf{v}^{k+1})}{\mathcal{F}(\lambda I - \gamma_1 \Delta)} \right), \quad (25)$$

where  $\mathcal{F}(\cdot)$  and  $\mathcal{F}^{-1}(\cdot)$  are the fast Fourier transform (FFT) and its inverse.

**Remark 3.3.** To Algorithm 2, the equations (22) and (24) can be rearranged as

$$\alpha^k + \gamma_1 (\nabla u^k - \nabla u^{k-1}) - \mathbf{R}_{\theta(f)} \mathbf{T}_\beta(f) \varrho^k = 0 \quad (26)$$

$$\lambda (f - u^k) = \operatorname{div} \alpha^k \quad (27)$$

by using (20d) and (20e). Note that  $\mathbf{R}_{\theta(f)}$  denotes the adjoint transformation of  $\mathbf{R}_{-\theta(f)}$ . Since  $\mathbf{R}_{-\theta(f)}$  is a rotation matrix and  $\mathbf{T}_\beta(f)$  is a diagonal matrix, we have  $\|\mathbf{T}_\beta(f)\|_2 \leq \beta$  and  $\|\mathbf{R}_{-\theta(f)}\|_2 = 1$ . Furthermore, using the convexity of  $\|\cdot\|_2^2$ , the equations (26) and (27) follow that

$$\|\nabla u^k - \nabla u^{k-1}\|_2^2 \leq \frac{1}{2\gamma_1^2} \left( \beta^2 \|\varrho^k\|_2^2 + \|\alpha^k\|_2^2 \right), \quad (28)$$

$$\frac{\lambda^2}{\bar{s}^2} \|u^k - f\|_2^2 \leq \|\alpha^k\|_2^2 \leq \frac{\lambda^2}{\underline{s}^2} \|u^k - f\|_2^2, \quad (29)$$

where  $\bar{s}$  and  $\underline{s}$  respectively denote the biggest and smallest of singular values to the gradient operator  $\nabla$ .

In the following, we mainly discuss some properties of Algorithm 2.

**Theorem 3.2.** Assume that the sequence  $\{(\mathbf{w}^k, \mathbf{v}^k, u^k, \alpha^k, \varrho^k)\}$  generated by the iteration (20a)-(20e) is bounded and there exists a constant  $C$  such that  $\frac{1}{\gamma_2} \|\varrho^{k+1} - \varrho^k\|_2^2 \leq C \|\mathbf{v}^{k+1} - \mathbf{v}^k\|_2^2$ , then it holds that

$$\mathcal{L}(\mathbf{w}^{k+1}, \mathbf{v}^{k+1}, u^{k+1}, \alpha^{k+1}, \varrho^{k+1}) \leq \mathcal{L}(\mathbf{w}^k, \mathbf{v}^k, u^k, \alpha^k, \varrho^k) + h_1 \|u^{k+1} - u^k\|_2^2 + h_2 \|\mathbf{v}^{k+1} - \mathbf{v}^k\|_2^2, \quad (30)$$

where  $h_1 := \frac{\lambda^2}{\gamma_1 \bar{s}^2} - \frac{\lambda + \gamma_1 \underline{s}^2}{2}$  and  $h_2 := C - \frac{(\gamma_1 + \gamma_2)}{2}$ .

*Proof.* Following from the formulation (20e) and using above assumptions, we have

$$\begin{aligned} \mathcal{L}_1 &:= \mathcal{L}(\mathbf{w}^{k+1}, \mathbf{v}^{k+1}, u^{k+1}, \alpha^{k+1}, \varrho^{k+1}) - \mathcal{L}(\mathbf{w}^{k+1}, \mathbf{v}^{k+1}, u^{k+1}, \alpha^{k+1}, \varrho^k) \\ &= \langle \varrho^{k+1} - \varrho^k, \mathbf{w}^{k+1} - \mathbf{T}_\beta(f) \mathbf{R}_{-\theta(f)} \mathbf{v}^{k+1} \rangle \\ &= \frac{1}{\gamma_2} \|\varrho^{k+1} - \varrho^k\|_2^2 \leq C \|\mathbf{v}^{k+1} - \mathbf{v}^k\|_2^2. \end{aligned}$$

By using the formulation (20d), we also deduce that

$$\begin{aligned} \mathcal{L}_2 &:= \mathcal{L}(\mathbf{w}^{k+1}, \mathbf{v}^{k+1}, u^{k+1}, \alpha^{k+1}, \varrho^k) - \mathcal{L}(\mathbf{w}^{k+1}, \mathbf{v}^{k+1}, u^{k+1}, \alpha^k, \varrho^k) \\ &= \langle \alpha^{k+1} - \alpha^k, \mathbf{v}^{k+1} - \nabla u^{k+1} \rangle \\ &= \frac{1}{\gamma_1} \|\alpha^{k+1} - \alpha^k\|_2^2 \leq \frac{\lambda^2}{\gamma_1 \bar{s}^2} \|u^{k+1} - u^k\|_2^2. \end{aligned}$$

With the help of formulations (20c) and (20d), the following fact is held as

$$\begin{aligned}
\mathcal{L}_3 &:= \mathcal{L}(\mathbf{w}^{k+1}, \mathbf{v}^{k+1}, u^{k+1}, \boldsymbol{\alpha}^k, \boldsymbol{\varrho}^k) - \mathcal{L}(\mathbf{w}^{k+1}, \mathbf{v}^{k+1}, u^k, \boldsymbol{\alpha}^k, \boldsymbol{\varrho}^k) \\
&= \frac{\lambda}{2} \|u^{k+1} - f\|_2^2 - \frac{\lambda}{2} \|u^k - f\|_2^2 + \langle \boldsymbol{\alpha}^k, \nabla u^k - \nabla u^{k+1} \rangle \\
&\quad + \frac{\gamma_1}{2} \|\mathbf{v}^{k+1} - \nabla u^{k+1}\|_2^2 - \frac{\gamma_1}{2} \|\mathbf{v}^{k+1} - \nabla u^k\|_2^2 \\
&= \frac{\lambda}{2} \langle (u^{k+1} + u^k) - 2f, (u^{k+1} - u^k) \rangle + \langle \operatorname{div} \boldsymbol{\alpha}^k, u^{k+1} - u^k \rangle \\
&\quad + \frac{\gamma_1}{2} \langle 2\mathbf{v}^{k+1} - \nabla u^{k+1} - \nabla u^k, \nabla u^k - \nabla u^{k+1} \rangle \\
&= -\frac{\lambda}{2} \|u^{k+1} - u^k\|_2^2 - \frac{\gamma_1}{2} \|\nabla u^{k+1} - \nabla u^k\|_2^2 \\
&\leq -\frac{\lambda + \gamma_1 \underline{s}^2}{2} \|u^{k+1} - u^k\|_2^2.
\end{aligned}$$

Based on the Taylor expansion at the point  $u^{k+1}$  of  $\mathcal{L}(\mathbf{w}^{k+1}, \mathbf{v}^{k+1}, u^k, \boldsymbol{\alpha}^k, \boldsymbol{\varrho}^k)$  and by using the formulations (20b)-(20e), we have

$$\begin{aligned}
\mathcal{L}_4 &:= \mathcal{L}(\mathbf{w}^{k+1}, \mathbf{v}^{k+1}, u^k, \boldsymbol{\alpha}^k, \boldsymbol{\varrho}^k) - \mathcal{L}(\mathbf{w}^{k+1}, \mathbf{v}^k, u^k, \boldsymbol{\alpha}^k, \boldsymbol{\varrho}^k) \\
&= \langle \boldsymbol{\alpha}^k, \mathbf{v}^{k+1} - \nabla u^k \rangle - \langle \boldsymbol{\alpha}^k, \mathbf{v}^k - \nabla u^k \rangle + \frac{\gamma_1}{2} \|\mathbf{v}^{k+1} - \nabla u^k\|_2^2 - \frac{\gamma_1}{2} \|\mathbf{v}^k - \nabla u^k\|_2^2 \\
&\quad + \langle \boldsymbol{\varrho}^k, \mathbf{w}^{k+1} - \mathbf{T}_\beta(f)\mathbf{R}_{-\theta(f)}\mathbf{v}^{k+1} \rangle - \langle \boldsymbol{\varrho}^k, \mathbf{w}^{k+1} - \mathbf{T}_\beta(f)\mathbf{R}_{-\theta(f)}\mathbf{v}^k \rangle + \frac{\gamma_2}{2} \|\mathbf{w}^{k+1} - \mathbf{T}_\beta(f)\mathbf{R}_{-\theta(f)}\mathbf{v}^{k+1}\|_2^2 \\
&\quad - \frac{\gamma_2}{2} \|\mathbf{w}^{k+1} - \mathbf{T}_\beta(f)\mathbf{R}_{-\theta(f)}\mathbf{v}^k\|_2^2 \\
&= \langle \boldsymbol{\alpha}^k, \mathbf{v}^{k+1} - \mathbf{v}^k \rangle + \frac{\gamma_1}{2} \langle \mathbf{v}^{k+1} + \mathbf{v}^k - 2\nabla u^k, \mathbf{v}^{k+1} - \mathbf{v}^k \rangle + \langle \boldsymbol{\varrho}^k, \mathbf{T}_\beta(f)\mathbf{R}_{-\theta(f)}\mathbf{v}^k - \mathbf{T}_\beta(f)\mathbf{R}_{-\theta(f)}\mathbf{v}^{k+1} \rangle \\
&\quad + \frac{\gamma_2}{2} \langle 2\mathbf{w}^{k+1} - \mathbf{T}_\beta(f)\mathbf{R}_{-\theta(f)}\mathbf{v}^{k+1} - \mathbf{T}_\beta(f)\mathbf{R}_{-\theta(f)}\mathbf{v}^k, \mathbf{T}_\beta(f)\mathbf{R}_{-\theta(f)}\mathbf{v}^k - \mathbf{T}_\beta(f)\mathbf{R}_{-\theta(f)}\mathbf{v}^{k+1} \rangle \\
&= \frac{\gamma_1}{2} \|\mathbf{v}^{k+1} - \mathbf{v}^k\|_2^2 - \frac{\gamma_2}{2} \|\mathbf{T}_\beta(f)\mathbf{R}_{-\theta(f)}(\mathbf{v}^{k+1} - \mathbf{v}^k)\|_2^2 + \langle \boldsymbol{\alpha}^k + \gamma_1(\mathbf{v}^k - \nabla u^k) - \mathbf{R}_{\theta(f)}\mathbf{T}_\beta(f)\boldsymbol{\varrho}^{k+1}, \mathbf{v}^{k+1} - \mathbf{v}^k \rangle \\
&= -\frac{\gamma_1}{2} \|\mathbf{v}^{k+1} - \mathbf{v}^k\|_2^2 - \frac{\gamma_2}{2} \|\mathbf{T}_\beta(f)\mathbf{R}_{-\theta(f)}(\mathbf{v}^{k+1} - \mathbf{v}^k)\|_2^2 \\
&\leq -\frac{\gamma_1 + \gamma_2}{2} \|\mathbf{v}^{k+1} - \mathbf{v}^k\|_2^2.
\end{aligned}$$

On the other hand, the subproblem (20a) implies that

$$\mathcal{L}_5 := \mathcal{L}(\mathbf{w}^{k+1}, \mathbf{v}^k, u^k, \boldsymbol{\alpha}^k, \boldsymbol{\varrho}^k) - \mathcal{L}(\mathbf{w}^k, \mathbf{v}^k, u^k, \boldsymbol{\alpha}^k, \boldsymbol{\varrho}^k) < 0.$$

Combining with  $\mathcal{L}_i$  for  $i = 1, \dots, 5$ , we deduce that

$$\mathcal{L}(\mathbf{w}^{k+1}, \mathbf{v}^{k+1}, u^{k+1}, \boldsymbol{\alpha}^{k+1}, \boldsymbol{\varrho}^{k+1}) \leq \mathcal{L}(\mathbf{w}^k, \mathbf{v}^k, u^k, \boldsymbol{\alpha}^k, \boldsymbol{\varrho}^k) + h_1 \|u^{k+1} - u^k\|_2^2 + h_2 \|\mathbf{v}^{k+1} - \mathbf{v}^k\|_2^2.$$

Then above assertion is held.  $\square$

**Remark 3.4.** In order to improve the effectiveness of the ADMM, we usually choose bigger penalty parameters  $\gamma_1$  and  $\gamma_2$ . With these suitable choices, we can simultaneously deduce that  $h_1 < 0$  and  $h_2 < 0$  and then get the monotonicity as

$$\mathcal{L}(\mathbf{w}^{k+1}, \mathbf{v}^{k+1}, u^{k+1}, \boldsymbol{\alpha}^{k+1}, \boldsymbol{\varrho}^{k+1}) \leq \mathcal{L}(\mathbf{w}^k, \mathbf{v}^k, u^k, \boldsymbol{\alpha}^k, \boldsymbol{\varrho}^k). \quad (31)$$

**Theorem 3.3.** If the assertions used in Theorem 3.2 are held, then the sequence  $\{(\mathbf{w}^k, \mathbf{v}^k, u^k, \boldsymbol{\alpha}^k, \boldsymbol{\varrho}^k)\}$  generated by Algorithm 2 converges to a critical point  $(\mathbf{w}^*, \mathbf{v}^*, u^*, \boldsymbol{\alpha}^*, \boldsymbol{\varrho}^*)$  of the augmented Lagrangian function  $\mathcal{L}(\mathbf{w}, \mathbf{v}, u, \boldsymbol{\alpha}, \boldsymbol{\varrho})$  defined in (19).

*Proof.* Since  $\{(\mathbf{w}^k, \mathbf{v}^k, u^k, \alpha^k, \varrho^k)\}$  is assumed to be bounded, then there exists a subsequence (we still denote it as  $\{(\mathbf{w}^k, \mathbf{v}^k, u^k, \alpha^k, \varrho^k)\}$  without any confusions in the following) such that

$$\lim_{j \rightarrow \infty} (\mathbf{w}^k, \mathbf{v}^k, u^k, \alpha^k, \varrho^k) = (\mathbf{w}^*, \mathbf{v}^*, u^*, \alpha^*, \varrho^*). \quad (32)$$

Due to the fact that  $\mathcal{F}_1(u) := \frac{\lambda}{2} \|u - f\|_2^2$  is convex and  $\mathcal{F}_2(\mathbf{w}) := \|\mathbf{w}\|_{2,p}^p$  is low semicontinuous [52], we have

$$\mathcal{L}(\mathbf{w}^*, \mathbf{v}^*, u^*, \alpha^*, \varrho^*) \leq \liminf_{k \rightarrow \infty} \mathcal{L}(\mathbf{w}^k, \mathbf{v}^k, u^k, \alpha^k, \varrho^k)$$

Combining with the monotonicity as shown in Theorem 3.2, we deduce that the sequence

$$\{(\mathbf{w}^k, \mathbf{v}^k, u^k, \alpha^k, \varrho^k)\}$$

is convergent. Then by summing up two sides of the inequality (30), we get

$$\begin{aligned} & \sum_{k=0}^{\infty} \left( -h_1 \|u^{k+1} - u^k\|_2^2 - h_2 \|\mathbf{v}^{k+1} - \mathbf{v}^k\|_2^2 \right) \\ & \leq \sum_{k=0}^{\infty} \left( \mathcal{L}(\mathbf{w}^k, \mathbf{v}^k, u^k, \alpha^k, \varrho^k) - \mathcal{L}(\mathbf{w}^{k+1}, \mathbf{v}^{k+1}, u^{k+1}, \alpha^{k+1}, \varrho^{k+1}) \right) \\ & \leq \mathcal{L}(\mathbf{w}^0, \mathbf{v}^0, u^0, \alpha^0, \varrho^0) - \mathcal{L}(\mathbf{w}^*, \mathbf{v}^*, u^*, \alpha^*, \varrho^*). \end{aligned}$$

Here  $(\mathbf{w}^0, \mathbf{v}^0, u^0, \alpha^0, \varrho^0)$  denotes the original value of Algorithm 2. Since  $h_1 < 0$  and  $h_2 < 0$ , we then obtain that

$$\lim_{k \rightarrow \infty} \|u^{k+1} - u^k\|_2^2 = 0 \quad \text{and} \quad \lim_{k \rightarrow \infty} \|\mathbf{v}^{k+1} - \mathbf{v}^k\|_2^2 = 0.$$

These imply that there exist  $\bar{u}^*$  and  $\bar{\mathbf{v}}^*$  such that  $u^k \rightarrow \bar{u}^*$  and  $\mathbf{v}^k \rightarrow \bar{\mathbf{v}}^*$  when  $k \rightarrow \infty$ . Using facts (26)-(27) and (20e), we get that there exist  $\bar{\alpha}^*$ ,  $\bar{\varrho}^*$  and  $\bar{\mathbf{w}}^*$  such that

$$\lim_{k \rightarrow \infty} \alpha^k = \bar{\alpha}^*, \quad \lim_{k \rightarrow \infty} \varrho^k = \bar{\varrho}^* \quad \text{and} \quad \lim_{k \rightarrow \infty} \mathbf{w}^k = \bar{\mathbf{w}}^*.$$

Furthermore, we also find that

$$\bar{\mathbf{v}}^* = \nabla \bar{u}^* \quad \text{and} \quad \bar{\mathbf{w}}^* = \mathbf{T}_\beta(f) \mathbf{R}_{-\theta(f)} \bar{\mathbf{v}}^*.$$

Thus  $(\bar{\mathbf{w}}^*, \bar{\mathbf{v}}^*, \bar{u}^*, \bar{\alpha}^*, \bar{\varrho}^*)$  is a critical point of the augmentation Lagrangian function  $\mathcal{L}(\mathbf{w}, \mathbf{v}, u, \alpha, \varrho)$  defined in (19). Furthermore, since  $(\bar{\mathbf{w}}^*, \bar{\mathbf{v}}^*, \bar{u}^*, \bar{\alpha}^*, \bar{\varrho}^*)$  can be regarded as one of cases of limitation points in (32), we can choose  $(\mathbf{w}^*, \mathbf{v}^*, u^*, \alpha^*, \varrho^*) := (\bar{\mathbf{w}}^*, \bar{\mathbf{v}}^*, \bar{u}^*, \bar{\alpha}^*, \bar{\varrho}^*)$  without any confusions. Then the above assertion is held.  $\square$

**Remark 3.5.** For the problem (18), we can rearrange it as three blocks nonconvex separable optimization problem

$$\begin{cases} \min_{u, v, w} \frac{\lambda}{2} \|u - f\|_2^2 + \|\mathbf{w}\|_{2,p}^p \\ \text{s.t. } \begin{bmatrix} -\nabla_x & & \\ & -\nabla_y & \\ & & 0 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} + \begin{bmatrix} 0 & 0 & w_1 \\ 0 & I & w_2 \\ I & & \mathbf{w} \end{bmatrix} + \begin{bmatrix} I & & \\ & I & \\ & & -\mathbf{T}_\beta(f) \mathbf{R}_{-\theta(f)} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \mathbf{0} \end{bmatrix}. \end{cases}$$

Then Algorithm 2 is of using the ADMM to solve above three-block separable optimization problem. There are many convergence analysis to multi-block constraint optimization problem [52, 26, 15], where they usually assumed that the coefficient matrices are nonsingular. It is obvious that our model is not summarized into this case. So the sequence generated by Algorithm 2 converges to the critical point. In fact, we also do not expect to obtain the global convergence for the model (18). In other words, the model (18) does not satisfy the Kurdyka-Łojasiewicz (KL) property when  $p$  is a irrational number [26, 1, 2].

#### 4. Parameters estimations

The restored ability of the model (7) depends on how to choose the rotation angle  $\theta(f)$ , the weighted parameter  $\beta(f)$  and the regularization parameter  $\lambda$ . Among these parameters, the main role of  $\theta(f)$  is of rotating the diffusion along with the tangential direction of edges and  $\beta(f)$  can make the diffusion more robust. The parameter  $\lambda$  balances the effect of the data fitting term and the regularization term. In addition, there are some penalty parameters  $\gamma_i$  for the ADMM. Here we can obtain an efficient denoising image when  $\gamma_i \in (0, 2]$ . So we mainly focus on how to choose  $\theta(f)$ ,  $\beta(f)$  and  $\lambda$  as follows.

##### 4.1. Orientation field estimation (OFE) of $\theta(f)$

The OFE has a critical impact on almost all subsequent processes in automatic fingerprint recognition systems [29, 61, 55, 42]. Thus, an incorrect orientation estimation can generate faulty recognition results. The most popular method is of computing the orientation field by using the least squares gradient-based method [32]. One important advantage of this scheme uses the gradient relationship of neighboring pixels to estimate the value of the ridge orientation. Since the edge is similar to the ridge in the image except for the steep jump, this paper employs the OFE to confirm the value of the spatially varying  $\theta(f)$ . Formally, the coarse orientation field at each pixel can be computed by

$$o(f) = \frac{1}{2} \tan^{-1} \frac{\sum_W 2\nabla_x f \cdot \nabla_y f}{\sum_W ((\nabla_x f)^2 - (\nabla_y f)^2)} + \frac{\pi}{2} \quad (33)$$

where  $W$  is a neighborhood window at the pixel point  $(i, j)$  for  $i = 1, \dots, m$  and  $j = 1, \dots, l$  (Note: here we set the  $W$  with the size  $16 \times 16$ ).  $\tan^{-1}$  is the arctangent function with the output range of  $(-\pi/2, \pi/2)$ . Since the formula (33) becomes less reliable when restoring the noisy image, we need a Gaussian smoothing function  $G_\sigma(f)$  to smooth the noisy image  $f$  via following scheme

$$\theta(f) = \frac{1}{2} \tan^{-1} \left\{ \frac{G_\sigma(f) * \sin(2 \cdot o(f))}{G_\sigma(f) * \cos(2 \cdot o(f))} \right\}. \quad (34)$$

##### 4.2. Weighted parameter $\beta(f)$ and regularization parameter $\lambda$

We use the binary function  $\beta(f)$  to describe the smooth and edge region for the matrix (8), where how to choose the shrink parameter  $b$  is the key. Here we set  $b = 0.001$  based on the histogram of  $\|G_{\hat{f}} * \nabla f\|_2$  and also find that this setting can distinguish the edge and other regions. Once  $b$  is set, we can choose  $\beta \in (1.3, 3.0)$  based on the empirical selection through the visual inspection. Specially, we choose a bigger value for the image with the simple structure and a small value for the image with the complicated structure.

To regularization parameter  $\lambda$ , if  $\lambda$  is too small, the restored image is oversmoothed. On the contrary, the restored image contains more noise. Some methods based on the noise prior have been proposed such as the L-curve method [28], the discrepancy principle [53], the variational Bayes approach [4], and the generalized cross validation (GCV) method [21, 12]. However, this paper mainly focuses on choosing an adaptive rotation angle  $\theta(f)$  to improve the TV-based models [47, 5]. So this paper manually choose  $\lambda$  by trials and errors. Specifically, we set the original value of  $\lambda$  into a bigger range as  $[a, b]$ , and then find a suitable subset as  $[a_1, b_1] \subset [a, b]$  by the bisection method. In the next we find a more suitable parameter in  $[a_i, b_i]$ . When the difference between the successive SNR is below 0.001 after  $\kappa$  iteration, we set  $\lambda = (a_\kappa + b_\kappa)/2$  as the value of the regularization parameter.

#### 5. Numerical experiments

This section mainly evaluates the AWTVP (7) compared to other related TV-based models through a series of numerical experiments. To the general gradient-based model, we can get the different regularization term  $\mathcal{R}(u)$ , which corresponds different TV-based restored models such as the ROF model (2), the HOTV model (3), the TGV model (4) and the DTV model (6)<sup>2</sup>. These models have been used in many fields of image processing as the standard

<sup>2</sup>Numerical comparisons use same symbol  $\lambda$  for different gradient-based models. To the TGV, we also carefully choose the suitable values  $\alpha_0$  and  $\alpha_1$  when obtaining the higher SNR.

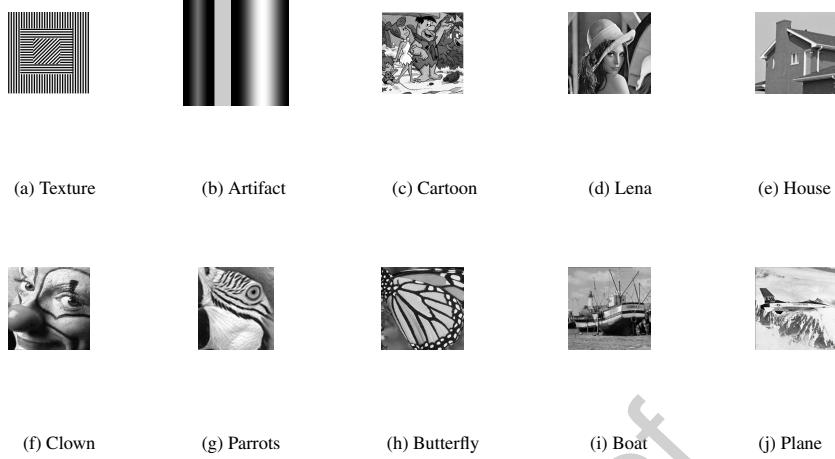


Figure 1. Original clean images.

compared models. Since they are nonsmooth and separable, we still employ the ADMM to solve them. In addition, the discriminative learning-based models have been used for the image denoising problem [33, 25, 58, 59, 18]. This scheme directly maps the input data to the desired solution through an optimization of a loss function on a training set containing degraded-clean image pairs. The CNN-based denoising methods belong to this category. Experimental results show that the CNN-based methods can consistently improve denoising performance for deterministic denoisers in some cases. A drawback of most learning-based approaches is that they are trained on a specific type of data, as well as a specific type and strength of noise. Since this paper mainly focuses on improving the TV-based models, we only choose one CNN-based model (called DnCNN) proposed in [58] for numerical comparisons.

For numerical comparisons, we choose ten classical testing images as shown in Figure 1, where degraded images are added to the white Gaussian noise with different standard deviations as  $\sigma = 0.01, 0.05, 0.1, 0.15$ . In addition, in order to evolve the restored ability, we also use the Matlab functions “*snr*(·)” and “*ssim*(·)” as the evaluation standard. In the numerical implementations, choosing parameter depends on to obtain higher SNR. It is obvious that the high SNR does not imply the high SSIM which can be seen from numerical comparisons. In addition ,Algorithm 2 includes an inner iteration and an outer iteration. For the inner iteration, it is of solving the subproblem (20a) by using Algorithm 1, which can theoretically give the exact solution when  $k \rightarrow \infty$ . Here we set  $\varepsilon = 0.01$  and  $k = 1$  since this setting still obtains a satisfied numerical result. To the outer iteration, the algorithm will be stopped if the iteration exceeds 500 or  $\|u^{k+1} - u^k\|_2 / \|u^{k+1}\|_2 \leq 10^{-5}$  is satisfied.

### 5.1. Affine images

Since the AWTVP is of introducing an adaptive angle  $\theta(f)$  to improve the DTV, this subsection first compares them by restoring the degraded image as shown in Figure 2 (a). It is obvious that Figure 2 (a) have three different affine structures with angles  $0, \pi/4, \pi/2$ . Since the angle of the DTV is fixed, we respectively set it to be  $0, \pi/4, \pi/2$ . As we can see from Figure 2 (b)-(c) that the AWTVP is more efficient due to the high SNR and SSIM. In order to enhance comparability, we plot the difference image between the clean image and the restored image as shown in Figure 2 (d)-(h). Formally, the difference image can be used to be a measure on how the models preserve image details, i.e., if the degraded image is restored efficiently, then the difference image clearly contains the information of noise. It is easy to find that there are fewer residual structures for the AWTVP. Furthermore, we can find that residual structures depend on the choice of the angle for the DTV, i.e., Figure 2 (f) gives a better restoration since its main affine direction is  $\pi/2$ .

In order to verify the availability of the AWTVP by using the threshold function  $\beta(f)$ , we artificially generate an affine image including the piecewise constant, slant and smooth regions as shown in Figure 1 (b). Here the degraded image is added to the white Gaussian noise with the standard deviation  $\sigma = 0.01$ . As we can see from the caption in

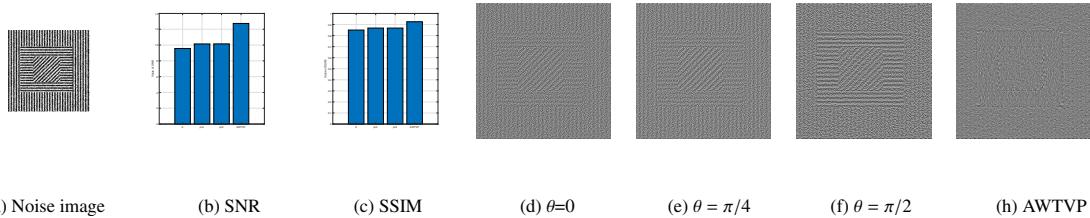


Figure 2. (a) Noisy image with  $\sigma = 0.1$ . (b) and (c) SNR and SSIM based on the DTV with different affine angles and AWTVP. (d)-(e) show the differences between the clean image and the restored images by using the DTV with choosing different angles and the AWTVP.

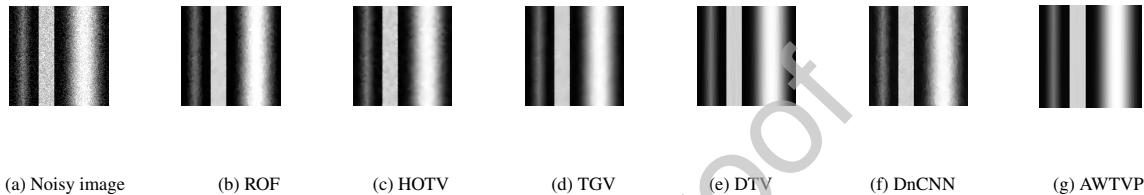


Figure 3. Denoising of a synthetical image that has been corrupted with the white Gaussian noise of variance 0.01. Related data: (a)  $SNR = 15.8685$ ;  $SSIM = 0.3882$ ; (b)  $\lambda = 13$ ,  $SNR = 22.8835$ ;  $SSIM = 0.7750$ ; (c)  $\lambda = 25$ ,  $SNR = 23.8509$ ;  $SSIM = 0.8194$ ; (d)  $\lambda = 2.5$ ,  $SNR = 25.4389$ ;  $SSIM = 0.8469$ ; (e)  $\lambda = 69$ ,  $SNR = 27.2060$ ;  $SSIM = 0.8444$ ; (f)  $SNR = 24.2178$ ;  $SSIM = 0.7925$ ; (g)  $\lambda = 1.6$ ,  $SNR = 27.6470$ ;  $SSIM = 0.8546$ .

Figure 3 that our proposed model gives the highest SSIM and SNR. Figure 4 plots the 30th row slice of the restored images for better visualisation. We can observe the staircasing effect for the ROF and the blurry edge for the HOTV. These two drawbacks can be suppressed efficiently for other four models. However, the DnCNN is inferior to the TGV, the DTV and the AWTVP. In general, the TGV is based on the couple of the total variation and the high-order total variation. The DTV and the AWTVP choose the affine angle  $\theta = \pi/2$  and make the diffusion of the corresponding Euler-Lagrange equation along with the tangential direction of the edge. Furthermore, the AWTVP preserves the constant region efficiently since the  $\ell^p$ -quasinorm has the property to keep the sparsity based on the total variation regularization.

## 5.2. Benchmark images

This subsection chooses Figure 1 (c)-(f) as testing images. Among these images, Cartoon is comprised by piecewise constant regions. Lena is widely used in the literatures since it contains both sharp edges and smooth regions. House includes many approximated cartoon regions and Clown image has many rough regions. From the SNR and SSIM results of the recovery images as shown in Figure 8, the AWTVP is still competitive with other TV-based methods. The related parameters to different models can be found in Table 1. To show the effectiveness via the visual quality of restored images, we show a part of restored images and colorbars based on the difference between the clean

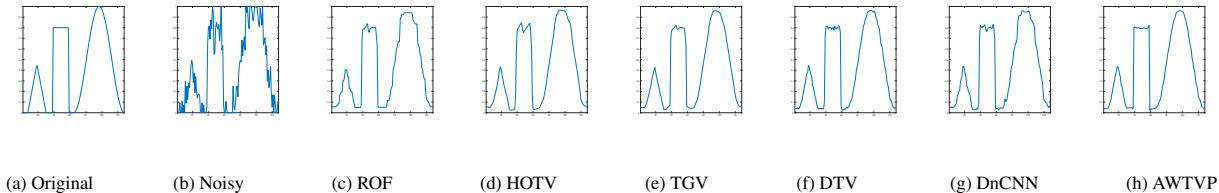


Figure 4. The 20th row slice of the restored image for understanding different methods to keep image details.

images and restored images. Here we only choose the noisy level  $\sigma = 0.05$  as an example to indicate the visual comparisons. It is easy to observe from Figure 5-7 that the AWTVP performs quite well and manages to completely preserve image details such as more obvious edges and smoother regions. Furthermore, the colorbars of the AWTVP are more dark, especially in the region of shock structures. It implies that using the AWTVP model includes more features than other TV-based models. In addition, we also give some comments between the DnCNN and the AWTVP. From SSIM and SNR in Figure 8, we can find that the DnCNN is superior to the AWTVP for lower noise level. To higher noise level as  $\sigma = 0.15$ , the AWTVP is superior to the DnCNN.

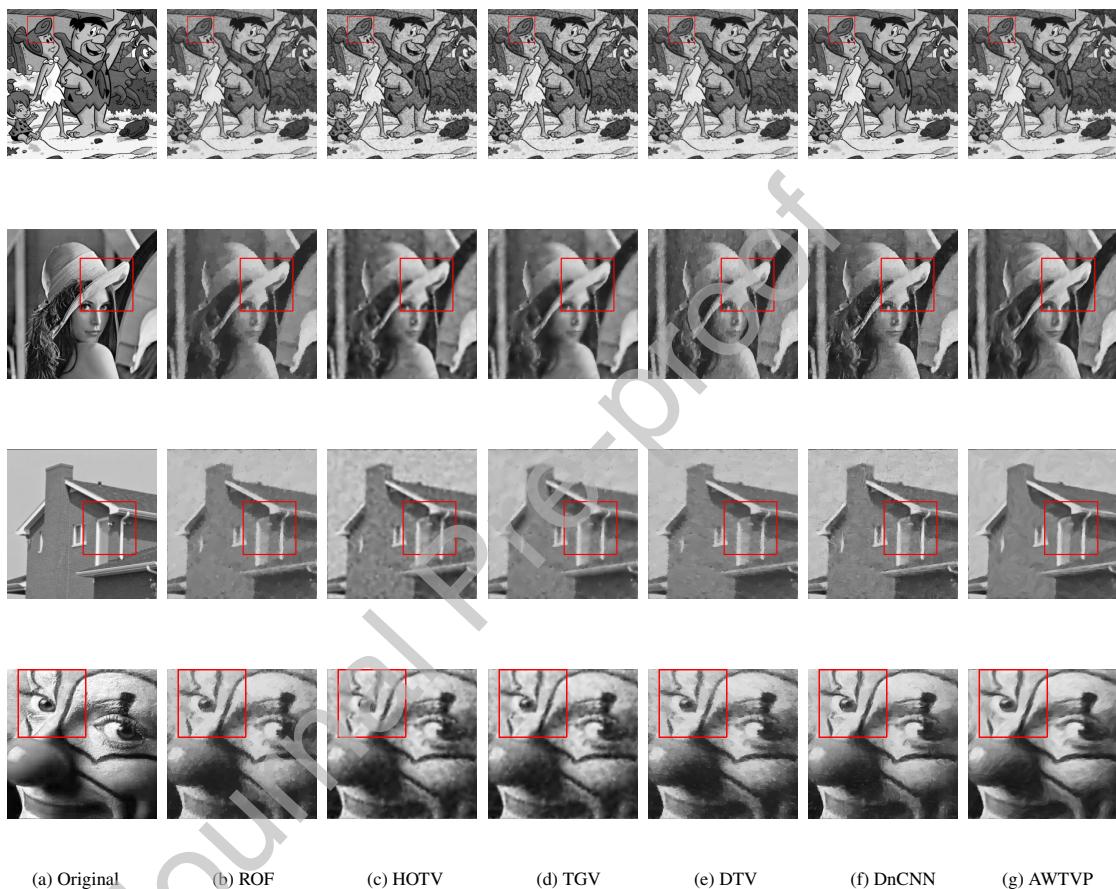


Figure 5. Denoising results are shown by different models (noise level  $\sigma = 0.05$ ).

### 5.3. Other images

We restore other degraded images based on Figure 1 (g)-(j). The SNR/SSIM of restored images are shown in Figure 10. We similarly observe that the AWTVP almost outperforms other TV-based models. More specifically, the AWTVP has better performance in preserving image structures and reduces the heavy noise effectively. In addition, the contours are plotted in Figure 11 by using the *imcontour* function in MATLAB. This function can evaluate the restored results through revealing the level curves. We can observe that the level curves generated by the AWTVP have more less oscillating. In addition, the same conclusion between the AWTVP and DnCNN can be obtained as commented in Section 5.2. Furthermore, since the AWTVP is nonconvex, we also verify the convergence of Algorithm 2 via the relative error (RErr)  $RErr = \frac{\|u^{k+1} - u^k\|_2}{\|u^{k+1}\|_2}$  and the mean squared error (MSE) as shown in Figure 11. It is obvious that Algorithm 2 is convergent in respect to efficiency and accuracy.

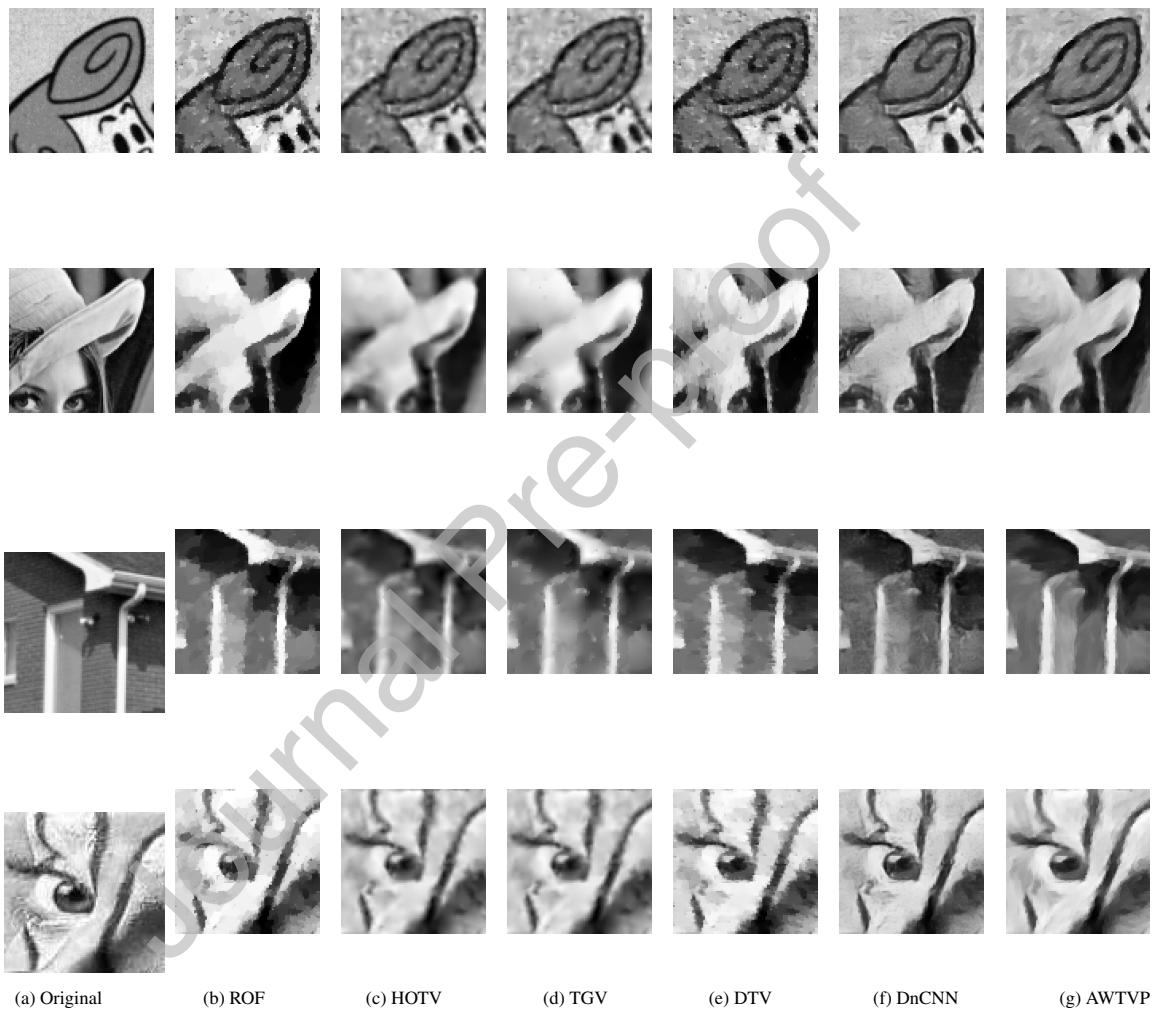


Figure 6. Zooming a part of denoising results are shown by different models (noise level  $\sigma = 0.05$ ), where the corresponding regions are marked red in Figure 5.

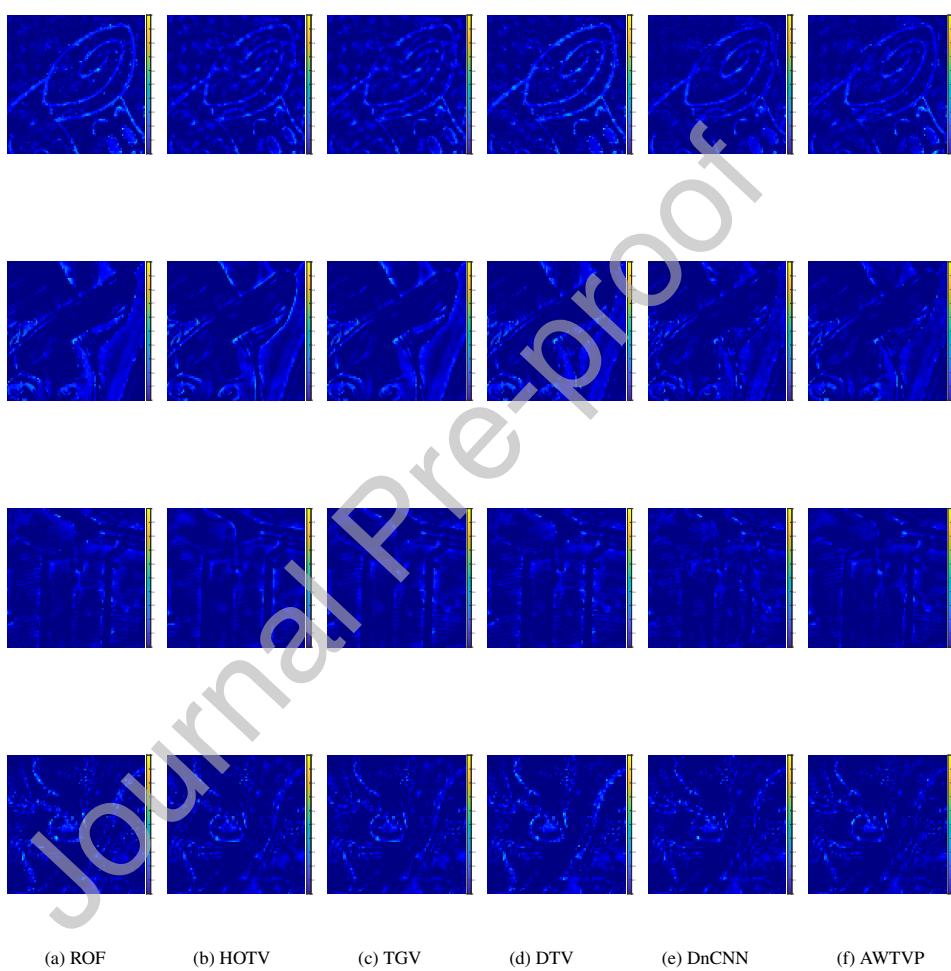


Figure 7. Visual comparison of colorbar to the difference between the clean images and restored images based on Figure 1 and 5.

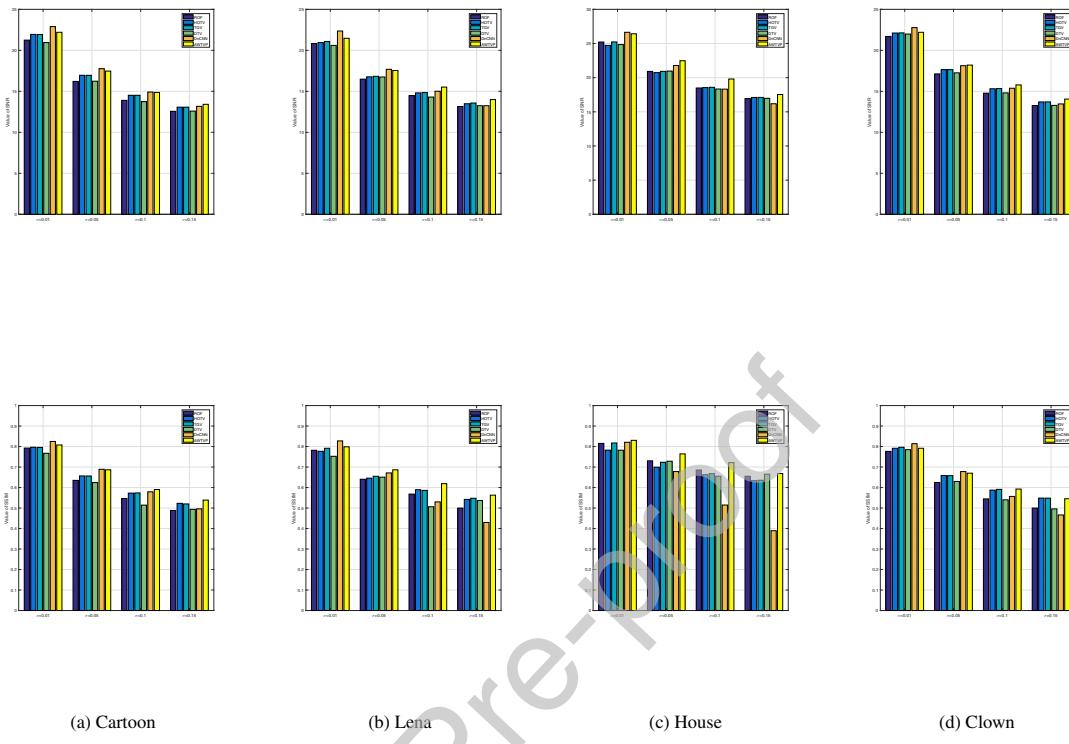
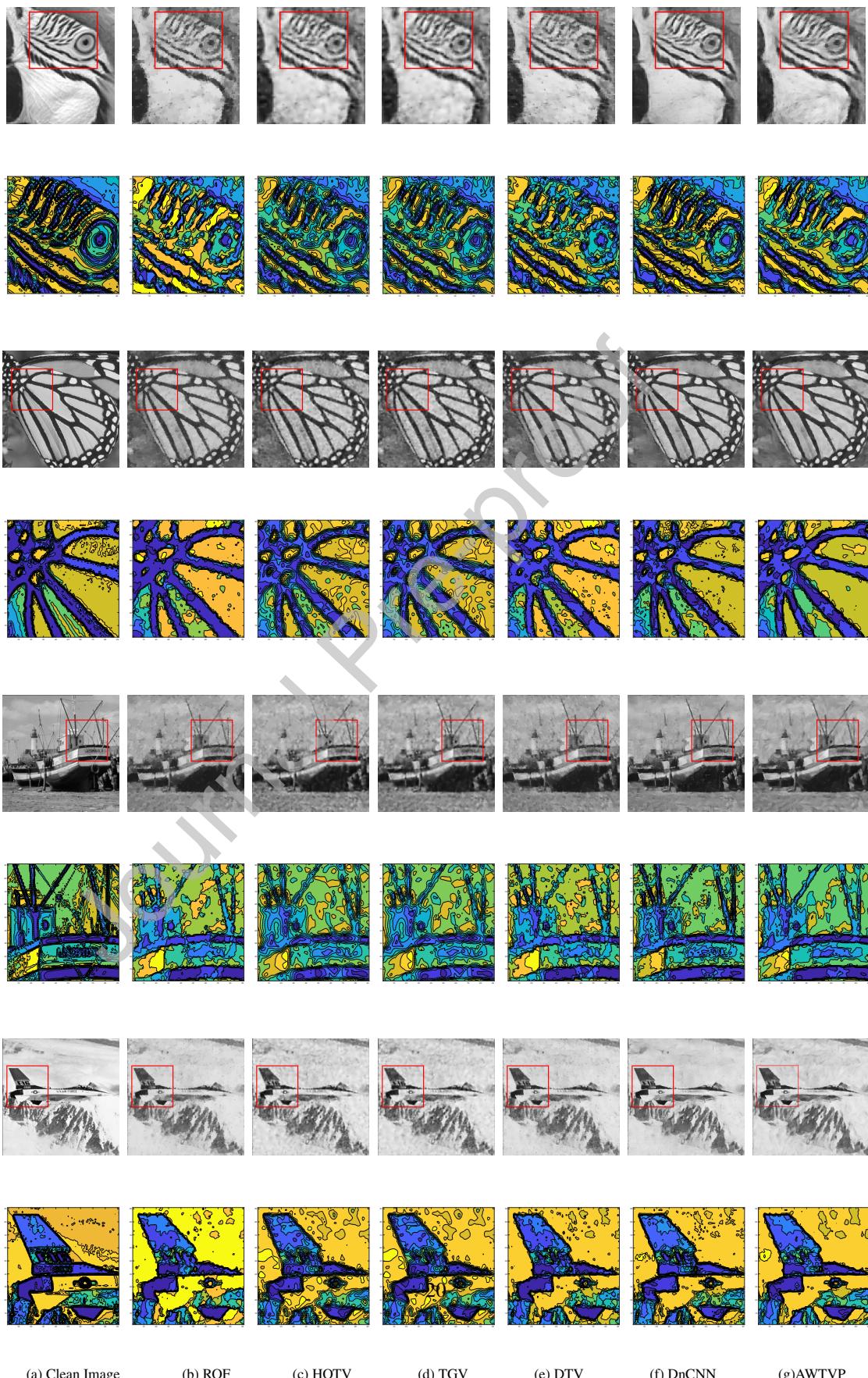


Figure 8. Bar graphs of SNR and SSIM to restored images.

Table 1. Related parameters are used in different gradient-based models.

Image	Noise	ROF	HOTV	TGV	DTV	AWTVP
Cartoon(512×512)	0.01	16.7	31	6.9	(116,8.3)	(1,1.38)
	0.05	7.4	13	2.86	(15.7,2.5)	(0.36,1.72)
	0.1	5.6	9.6	2.1	(23,5.4)	(0.34,1.5)
	0.15	5	8	1.8	(5.8,1.3)	(0.24,2)
Lena(256×256)	0.01	14.6	28	5.4	(76,9.1)	(0.95,2)
	0.05	6.3	10	2.36	(8.8,1.82)	(0.142,2)
	0.1	4.82	6.7	1.5	(19.6,5.6)	(0.22,2)
	0.15	4.3	5.7	1.3	(4.8,1.3)	(0.31,2.1)
House(256×256)	0.01	12	21.4	4.4	(42,5.7)	(0.49,1.9)
	0.05	5.5	7.6	1.67	(8.3,2)	0.18,2.5)
	0.1	4.2	5.1	1.06	(10.2,3.6)	0.16,3)
	0.15	3.7	4.4	0.86	(4.72,1.51)	(0.062,2.1)
Clown(200×200)	0.01	15.6	30	5.92	(23.1,1.36)	(1.08,1.72)
	0.05	6.8	10.8	1.96	(12.7,2.3)	(0.31,1.9)
	0.1	5.2	7.5	1.37	(11.4,2.82)	(0.36,2.18)
	0.15	4.6	6.3	1.16	(9.6,2.7)	(0.38,2.3)

Figure 9. Visual comparisons of contours are only shown for the standard deviation  $\sigma = 0.05$ .

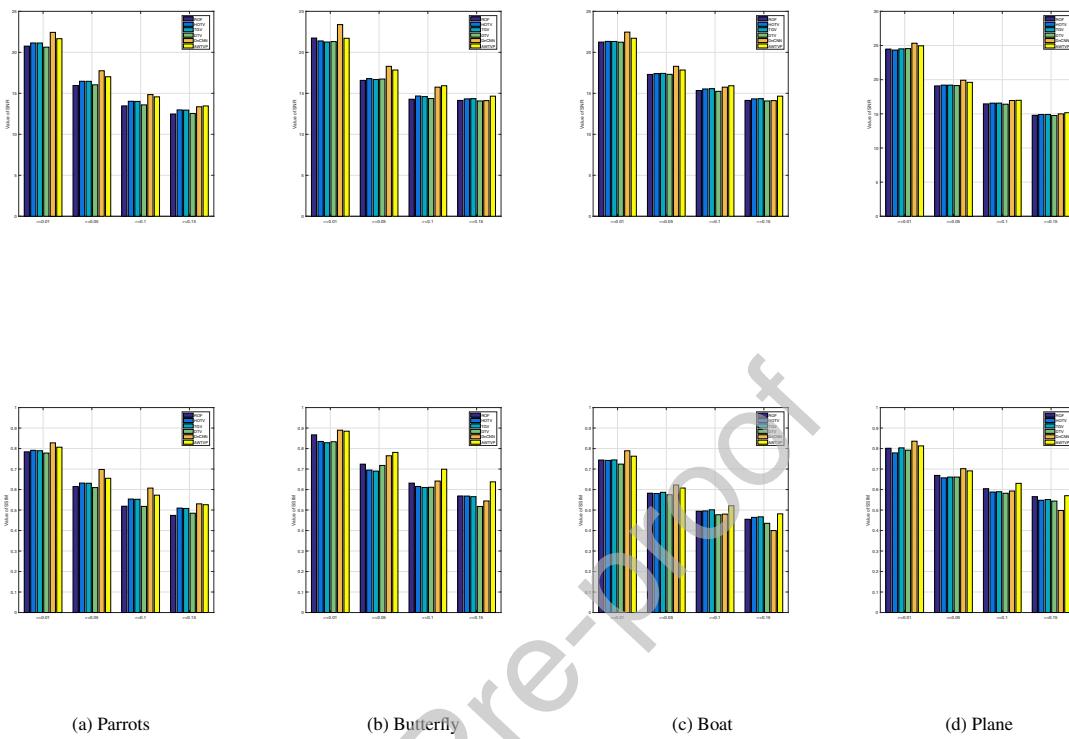


Figure 10. The bar graphs of SNR and SSIM to restored images.

Table 2. Related parameters are used in different gradient-based models.

Image	Noise	ROF	HOTV	TGV	DTV	AWTVP
Parrots(104×96)	$\sigma$	$\lambda$	$\lambda$	$\lambda$	( $\lambda, \beta$ )	( $\lambda, \beta$ )
	0.01	19.7	40	7.2	(102,5.7)	(1.3,2.5)
	0.05	8.4	15.5	3.3	(23,3.04)	(0.46,1.71)
	0.1	6.4	11.4	3.4	(42,8.2)	(0.39,2)
Butterfly(256×256)	0.15	5.6	9.6	2.9	(6.2,1.3)	(0.4,2.1)
	0.01	14.2	29.8	6.7	(87,7.5)	(0.63,1.62)
	0.05	6.8	12.4	2.74	(14.8,2.8)	(0.35,2.1)
	0.1	5.4	9.2	1.29	(17.1,4.36)	(0.21,1.7)
Boat(256×256)	0.15	4.8	7.9	1.1	(15,4.2)	(0.23,2)
	0.01	15.7	30.8	6.3	(30,1.4)	(1.5,1.9)
	0.05	6.5	10.6	1.88	(11.2,2)	(0.29,1.9)
	0.1	4.9	7.1	1.35	(11.3,3.1)	(0.32,2.4)
Plane(256×256)	0.15	4.3	5.7	1.1	(10,3)	(0.34,2.6)
	0.01	17.2	35.3	6.9	(24.3,1.34)	(1.56,2.4)
	0.05	7.5	13.6	2.4	(12.8,1.95)	(0.5,1.9)
	0.1	5.4	9.1	1.58	(12.6,3)	(0.45,2.8)
	0.15	4.6	7.3	1.34	(10.2,3)	(0.47,2.7)

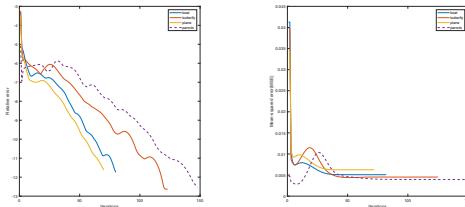


Figure 11. Relative error  $\log(ReErr)$  and MSE vs iteration of Algorithm 2 to solve our proposed AWTVP (7).

## 6. Conclusions

This paper proposed a new image denoising model via combining the rotation matrix and the weighted matrix into the  $\ell^p$ -quasinorm total variation regularization. For the rotation matrix, we used orientation field estimation (OFE) to obtain the adaptive angle  $\theta(f)$ . For the weighted matrix, we used the edge detection function to indicate the smoothing and edge regions. With these setting, the AWTVP can describe the local structure of the image more robust. Since the AWTVP is nonconvex and non-Lipschitz, we used the ADMM with the help of the half-quadratic scheme to solve it. In the theoretical part, the convergences of numerical algorithms were discussed. Experimental results reported the effectiveness of our proposed model. As a conclusion, we also noticed some slight disturbance in the approximated smooth regions for the AWTVP. Actually, this phenomena can be improved by choosing an adaptive parameter  $\beta(f)$  in the matrix  $\mathbf{T}_\beta(f)$ . So our next work will consider how to choose more robust parameter  $\beta(f)$  to improve our proposed model (7).

## Acknowledgment

The authors would like to thank the editor and the reviewers for their valuable comments, suggestions, and to make this paper much improved. This work was partially supported by the National Basic Research Program of China (973 Program) (No. 2015CB 856003), Hunan Provincial Key Laboratory of Mathematical Modeling and Analysis in Engineering (Changsha University of Science & Technology) and Programs for Science and Technology Development of HeNan Province (192102310181).

## References

### References

- [1] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. *Mathematics of Operations Research*, 35(2):438-457, 2010.
- [2] H. Attouch, J. Bolte, and B. Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized gauss Gseidel methods. *Mathematical Programming*, 137(1-2):91-129, 2013.
- [3] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*. Springer-Verlag, New York, 2002.
- [4] S. Babacan, R. Molina, and A. Katsaggelos. Parameter estimation in TV image restoration using variational distribution approximation. *IEEE Transactions on image processing*, 17(3):326-339, 2008.
- [5] I. Bayram and M. Kamasak. Directional total variation. *IEEE Signal Processing Letters*, 19(12):781-784, 2012.
- [6] M. Benning and M. Burger. Modern regularization methods for inverse problems. *Acta Numerica*, 27:1-111, 2018.
- [7] 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):1-122, 2011.
- [8] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492-526, 2010.
- [9] R. Chan and H. Liang. Half-Quadratic algorithm for  $\ell_q - \ell_p$  problems with application to TV- $\ell_1$  image restoration and compressive sensing. *Global Optimization Methods*, 78-103, 2014.

- [10] T. Chan, A. Marquina, and P. Mulet. High-order total variation-based image restoration. *SIAM Journal on Scientific Computing*, 22(2):503-516, 2000.
- [11] T. Chan and J. Shen. *Image Processing and Analysis Variational, PDE, Wavelet and Stochastic Methods*. SIAM, Pennsylvania, 2005.
- [12] T. Chan , H. Zhou, and R. Chan. Continuation method for total variation denoising problem. *International Symposium on Optical Science, Engineering, and Instrumentation*, 1995.
- [13] R. Chartrand. Exact reconstruction of sparse signals via nonconvex minimization. *IEEE Signal Processing Letters*, 14(10):707-710, 2007.
- [14] R. Chartrand and W. Yin. Iteratively reweighted algorithms for compressive sensing. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 3869-3872, 2008.
- [15] X. Chen, L. Guo, Z. Lu, and J. Ye. An augmented Lagrangian method for non-Lipschitz nonconvex programming. *SIAM Journal on Numerical Analysis*, 55(1):168-193, 2017.
- [16] B. Dong. Image restoration: wavelet frame shrinkage, nonlinear evolution PDEs, and beyond. *Multiscale Modeling & Simulation*, 15(1):606-660, 2017.
- [17] B. Dong. Sparse representation on graphs by tight wavelet frames and applications. *Applied and Computational Harmonic Analysis*, 42(3):452-479, 2017.
- [18] **W. Dong, P. Wang, W. Yin, and G. Shi.** Denoising prior driven deep neural network for image restoration. *IEEE transactions on pattern analysis and machine intelligence*, DOI: 10.1109/TPAMI.2018.2873610, 2018.
- [19] **W. Feng, P. Qiao, X. Xi, and Y. Chen.** Image denoising via multiscale nonlinear diffusion models. *SIAM Journal on Imaging Sciences*, 10(3):1234-1257, 2017.
- [20] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers and Mathematics with Applications*, 2(1):17-40, 1976.
- [21] N. Galatsanos and A. Katsaggelos. Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation. *IEEE Transactions on Image Processing*, 1(3):322-336, 1992.
- [22] D. Geman and C. Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE Transactions on Image Processing*, 47(2):932-946, 1995.
- [23] R. Glowinski and A. Marroco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par équilibration-dualité, d'une classe de problèmes de Dirichlet non linéaires. *Revue Française d'Automatique, Informatique, Recherche Opérationnelle*, 9(R-2):41-76, 1975.
- [24] R. Glowinski, S. Osher, and W. Yin. *Splitting Methods in Communication, Imaging, Science, and Engineering*. Springer, 2016.
- [25] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang. Weighted nuclear norm minimization and its applications to low level vision. *International Journal of Computer Vision*, 121(2):183-208, 2017.
- [26] K. Guo, D. Han, and T. Wu. Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints. *International Journal of Computer Mathematics*, 94(8):1653-1669, 2017.
- [27] W. Guo, J. Qin, and W. Yin. A new detail-preserving regularization scheme. *SIAM Journal on Imaging Sciences*, 7(2):1309-1334, 2014.
- [28] P. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, 34(4):561-580, 1992.
- [29] L. Hong, Y. Wan, and A. Jain. Fingerprint image enhancement: Algorithm and performance evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):777-789, 1998.
- [30] M. Jalalzai. Some remarks on the staircasing phenomenon in total variation-based image denoising. *Journal of Mathematical Imaging and Vision*, 52(2):256-268, 2016.
- [31] M. Jung and M. Kang. Efficient nonsmooth nonconvex optimization for image restoration and segmentation. *Journal of Scientific Computing*, 62(2):336-370, 2015.
- [32] M. Kass and A. Witkin. Analyzing oriented patterns. *Computer Vision, Graphics, and Image Processing*, 37(3):362-385, 1987.
- [33] D. Kim and Z. Shamsi. Enhanced residual noise estimation of low rank approximation for image denoising. *Neurocomputing*, 293(7):1-11, 2018.
- [34] R. Kongskov and Y. Dong. Directional total generalized variation regularization for impulse noise removal. *Scale Space and Variational Methods in Computer Vision*, 221-231, 2017.
- [35] R. Kongskov, Y. Dong, and K. Knudsen. Directional total generalized variation regularization. arXiv:1701.02675, 2017.
- [36] S. Lefkimmatis, A. Bourquard, and M. Unser. Hessian-based norm regularization for image restoration with biomedical applications. *IEEE Transactions on Image Processing*, 21(3):983-995, 2012.
- [37] S. Lefkimmatis, J. Ward, and M. Unser. Hessian Schatten-norm regularization for linear inverse problems. *IEEE Transactions on image processing*, 22(5):1873-1888, 2013.
- [38] F. Li, L. Pi, and T. Zeng. Explicit coherence enhancing filter with spatial adaptive elliptical kernel. *IEEE Signal Processing Letters*, 19(9):555-558, 2012.
- [39] **M. Liu, L. Nie, X. Wang, Q. Tian, and B. Chen.** Online data organizer: micro-video categorization by structure-guided multimodal dictionary learning. *IEEE Transactions on Image Processing*, 28(3):1235-1247, 2019.
- [40] R. Chan, S Setzer, and G Steidl. Inpainting by flexible Haar-wavelet shrinkage. *SIAM Journal on Imaging Sciences*, 1(3):273-293, 2008.
- [41] P. Hsieh, P. Shao, and S. Yang. A regularization model with adaptive diffusivity for variational image denoising. *Signal Processing*, 149:214-228, 2018.
- [42] S. Liu, M. Liu, and Z. Yang. Sparse coding based orientation estimation for latent fingerprints. *Pattern Recognition*, 67:164-176, 2017.
- [43] M. Lysaker, A. Lundervold, and X. Tai. Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. *IEEE Transactions on image processing*, 12(12):1579-1590, 2003.
- [44] M. Nikolova and T. Chan. The equivalence of the half-quadratic minimization and the gradient linearization iteration. *IEEE Transactions on Image Processing*, 16(6):1623-1627, 2007.
- [45] B. Rao and K. Delgado. An affine scaling methodology for best basis selection. *IEEE Transactions on Signal Processing*, 47(1):187-200, 1999.
- [46] P. Rodriguez and B. Wohlberg. Efficient minimization method for a generalized total variation functional. *IEEE Transactions on Image Processing*, 18(2):322-332, 2009.

- [47] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1-4):259-268, 1992.
- [48] O. Scherzer. Denoising with higher order derivatives of bounded variation and an application to parameter estimation. *Computing*, 60(1):1-27, 1998.
- [49] O. Scherzer. *Handbook of Mathematical Methods in Imaging*. Springer, New York, 2015.
- [50] Y. She. Thresholding-based iterative selection procedures for model selection and shrinkage. *Electronic Journal of Statistics*, 3:384-415, 2009.
- [51] Y. Shen, Q. Liu, S. Lou, and Y. Hou. Wavelet-based total variation and nonlocal similarity model for image denoising. *IEEE Signal Processing Letters*, 24(6):877-881, 2017.
- [52] Y. Wang, W. Yin, and J. Zeng. Global convergence of ADMM in nonconvex nonsmooth optimization. *Journal of Scientific Computing*, doi.org/10.1007/s10915-018-0757-z.
- [53] Y. Wen and R. Chan. Parameter selection for total-variation-based image restoration using discrepancy principle. *IEEE Transactions on image processing*, 21(4):1770-1781, 2012.
- [54] L. Yang, T. Pong, and X. Chen. Alternating direction method of multipliers for a class of nonconvex and nonsmooth problems with applications to background/foreground extraction. *SIAM Journal on Imaging Sciences*, 10(1):74-110, 2017.
- [55] J. Zhang, R. Lai, and C. Jay. Adaptive directional total-variation model for latent fingerprint segmentation. *IEEE Transactions on Information Forensics and Security*, 8(8):1261-1273, 2013.
- [56] H. Zhang, L. Tang, Z. Fang, C. Xiang, and C. Li. Nonconvex and nonsmooth total generalized variation model for image restoration. *Signal Processing*, 143:69-85, 2018.
- [57] H. Zhang and Y. Wang. Edge adaptive directional total variation. *The Journal of Engineering*, 2013(11):61-62, 2013.
- [58] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142-3155, 2017.
- [59] K. Zhang, W. Zuo, and L. Zhang. FFDNet: toward a fast and flexible solution for CNN-based image denoising. *IEEE Transactions on Image Processing* 27 (9), 4608-4622, 2018.
- [60] L. Zhang, P. Bao, and X. Wu. Hybrid inter-and intra-wavelet scale image restoration. *Pattern Recognition*, 36(8):1737-1746, 2003.
- [61] E. Zhu, J. Yin, C. Hu, and G. Zhang. A systematic method for fingerprint ridge orientation estimation and image segmentation. *Pattern Recognition*, 39(8):1452-1472, 2006.

**Conflict of Interest**

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.