ADAPTIVELY WEIGHTED DIFFERENCE MODEL OF ANISOTROPIC AND ISOTROPIC TOTAL VARIATION FOR IMAGE DENOISING

BAOLI SHI[1], MENGXIA LI[2], YIFEI LOU[3,*]

[1]*Department of Mathematics, Henan University, 475004, Kaifeng, China*
[2]*School of Science, Anyang Univesity, Anyang, 455000, China*
[3]*Department of Mathematical Sciences, The University of Texas Dallas, Richardson, TX 75080, USA*

Abstract. This paper proposes a novel nonconvex regularization functional by using an adaptively weighted difference model of anisotropic and isotropic total variation. By choosing the weights adaptively at each pixel, our model can enhance the anisotropic diffusion so as to achieve robust image recovery. Regarding to numerical implementations, we express the proposed model into a saddle point problem with the help of a dual formulation of the total variation, followed by a primal dual method to find a model solution. Numerical experiments demonstrate that the proposed approach is superior over several gradient-based methods for image denoising in terms of both visual appearance and quantitative metrics of signal noise ratio (SNR) and structural similarity index measure (SSIM).

Keywords. Anisotropic and isotropic total variation model; Difference of convex function; Image denoising; Noconvex optimization; Primal dual method.

1. INTRODUCTION

Digital images commonly suffer from degradations caused by imaging systems during formation, storage, transmission, *etc*. Image denoising plays a critical role in various applications such as medical and astronomical imaging, video coding, and computer vision [1, 2]. Without loss of generality, we assume that an underlying image is of size $n \times n$, which can be represented by an n^2-dimension vector using a column lexicographic order. Our approach can be easily extended to an arbitrary dimension. We consider a single-channel image degradation model as follows

$$\mathbf{f} = \mathbf{u} + \mathbf{n},$$

where $\mathbf{f} \in \mathbb{R}^{n^2}$ is a noisy observation, $\mathbf{u} \in \mathbb{R}^{n^2}$ is a clean image, and $\mathbf{n} \in \mathbb{R}^{n^2}$ denotes the additive white Gaussian noise with the zero mean and the variance σ^2. Restoring an underlying image $\mathbf{u}$ from the noisy input $\mathbf{f}$ is known as an ill-posed problem. It is specifically challenging to preserve image details such as edges and textures due to the lack of prior information. Among various

---

traditional denoising techniques, a large number of studies have demonstrated that a variational-based model is particularly useful in solving many ill-posed inverse problems [3, 4, 5, 6]. A general mathematical model can be expressed as follows,

$$\min_{\mathbf{u} \in \mathbb{R}^{n^2}} \frac{\lambda}{2} \|\mathbf{u} - \mathbf{f}\|_2^2 + \mathrm{R}(\mathbf{u}), \tag{1.1}$$

where the quadratic fidelity term models the presence of the additive white Gaussian noise, the regularization term $\mathrm{R}(\mathbf{u})$ encodes prior information on the target image $\mathbf{u}$, and the positive parameter λ balances these two terms. Total variation (TV) [7] is widely used as a regularization functional due to its edge-preserving ability. However, it tends to favor a piecewise constant output that contains undesirable staircasing artifacts. There are three major categories of improvements over TV. First, nonlocal-based regularizations [8, 9, 10, 11, 12, 13, 14] take advantages of image self-similarities that can significantly increase image recovery quality, but at the cost of high computational complexity. Second, higher-order derivative models that rely on (local) gradient information [4, 15, 16, 17] are proposed to preserve piecewise smoothness of the reconstructed solution. Third, nonconvex regularizations [18, 19, 20, 21, 22] have recently gained popularity to promote sparsity after taking the gradient. One naive way of enforcing the sparsity involves the minimization of the ℓ_0 quasinorm [23]. As it is NP hard to optimize, some continuous and nonconvex alternatives are sought such as the capped-ℓ^1 regularization [24], the ℓ^p ($0 < p < 1$) quasinorm regularization [25], the minimax concave penalty (MCP) [26], etc.

This paper proposes an adaptively weighted difference of anisotropic and isotropic TV (AW-DAITV) regularization. Specifically, we define the image gradient by

$$\nabla \mathbf{u} = (\nabla_x \mathbf{u}, \nabla_y \mathbf{u}) = ((I \otimes D)\mathbf{u}, (D \otimes I)\mathbf{u}) \in \mathbb{R}^{n^2 \times 2},$$

where $\otimes$ denotes the Kronecker product, I denotes the $n \times n$ identity matrix, and D denotes the $n \times n$ difference matrix (please refer to Section 2 for more details on the notations.) Then the proposed AWAITV model for image denoising is given by

$$\min_{\mathbf{u} \in \mathbb{R}^{n^2}} \frac{\lambda}{2} \|\mathbf{u} - \mathbf{f}\|_2^2 + \|T(\nabla \mathbf{u})\|_{1,1} - \alpha \|\nabla \mathbf{u}\|_{2,1}, \tag{1.2}$$

where $\lambda > 0$, $\alpha \in [0, 1]$, and

$$T(\nabla \mathbf{u}) = (\mathbf{t}_1 \nabla_x \mathbf{u}, \mathbf{t}_2 \nabla_y \mathbf{u}) := (\mathbf{t}_1 \circ (I \otimes D)\mathbf{u}, \mathbf{t}_2 \circ (D \otimes I)\mathbf{u}),$$

with pointwise multiplication $\circ$ and two vectors $\mathbf{t}_1, \mathbf{t}_2 \in \mathbb{R}^{n^2}$. When setting $\mathbf{t}_1 = \mathbf{t}_2 = (1, 1, \cdots, 1) \in \mathbb{R}^{n^2}$, the proposed model (1.2) reduces to the weighted difference of anisotropic and isotropic TV (WDAITV) that was originally proposed by Lou et al. [27]. WDAITV was later extended to image segmentation [28, 29] and impulsive noise removal [30]. By adding the weighted operator T to the anisotropic term, i.e., $\|\nabla \mathbf{u}\|_{1,1}$, our main motivation is to increase the anisotropic diffusion along with the tangential direction of the edge in order to adaptively align with (true) image gradients.

Due to the non-convexity and non-differentiability of the proposed AWDAITV model (1.2), it is challenging to design an efficient algorithm that achieves a compromise between quality and scalability. One way to solve for (1.2) is via the difference of convex algorithm (DCA) proposed by Pham-Dihn and Le-Thi [31, 32, 33]. In general, DCA minimizes the difference of convex functions in the form of $f(x) - g(x)$, so-called the DC function. The idea of DCA is to

use the subgradient of $f(x)$ and the subgradient of the Fenchel conjugate of $g(x)$ to obtain an iterative sequence. Under some conditions, the generated sequence satisfies the monotonicity conditions, thus converging to a critical point of the original problem. DCA was used in [27, 28] for the WDAITV model, in which additional operator splitting techniques were involved to decouple the gradient operators for the ℓ^1 and the ℓ^2 norms. For example, the alternating direction method of multipliers (ADMM) was applied to solve a subproblem in [27], while primal-dual hybrid gradient (PDHG) with a linesearch (PDHGLS) [34] was adopted for image segmentation [28]. Due to an extra call of ADMM and PDHGLS, the DCA framework is computationally expensive.

To enhance computational efficiency, we propose a new numerical method that combines the dual formulation of TV and primal-dual optimization framework. In particular, we rewrite the model (1.2) as a saddle point problem by using the equivalent representations of the ℓ^1 and the ℓ^2 norms with conjugate functions. As a result, the primal dual method (PDM) [35] leads to a single-loop scheme. The main contributions of this work are twofold:

- We propose a novel AWDAITV model (1.2) to increase the strength of the anisotropic diffusion.
- We propose a single-loop PDM that significantly improves the computational efficiency over the previous DCA approach for minimizing a DC function.

The remainder of this manuscript is organized as follows. Section 2 introduces the notations, establishes the solution's existence of the proposed model, and presents a toy example to illustrate the advantages of the weighting operator T. A numerical algorithm based on PDM is described in Section 3. Section 4 presents experimental results of the proposed approach in comparison to several related models. Finally, conclusions are given in Section 5.

2. MODEL ANALYSIS

We start with necessary notations and definitions that will be used throughout the paper. We denote $\mathbf{X} \subset \mathbb{R}^{n^2}$ as the image domain and $\mathbf{Y} := \mathbf{X} \times \mathbf{X}$ as the domain of image gradient. The inner products on the spaces $\mathbf{X}$ and $\mathbf{Y}$ are defined by

$$\langle \mathbf{a}, \mathbf{b} \rangle_{\mathbf{X}} = \sum_{i=1}^{n^2} a_i b_i \ \text{ and } \ \left\langle \widehat{\mathbf{c}}, \widehat{\mathbf{d}} \right\rangle_{\mathbf{Y}} = \sum_{i=1}^{n^2} \left(c_{1,i} d_{1,i} + c_{2,i} d_{2,i} \right),$$

for $\mathbf{a}, \mathbf{b} \in \mathbf{X}$, $\widehat{\mathbf{c}} = (\mathbf{c}_1, \mathbf{c}_2) \in \mathbf{Y}$, and $\widehat{\mathbf{d}} = (\mathbf{d}_1, \mathbf{d}_2) \in \mathbf{Y}$. We define the related norms on the spaces $\mathbf{X}$ and $\mathbf{Y}$ as follows

$$\|\mathbf{a}\|_2 = \sqrt{\sum_{i=1}^{n^2} a_i^2}, \ \|\widehat{\mathbf{c}}\|_{1,1} = \sum_{i=1}^{n^2} (|c_{i,1}| + |c_{i,2}|), \ \|\widehat{\mathbf{c}}\|_{2,1} = \sum_{i=1}^{n^2} \sqrt{|c_{i,1}|^2 + |c_{i,2}|^2},$$

$$\text{and } \|\widehat{\mathbf{c}}\|_{2,\infty} = \max_{1 \le i \le n^2} \left\{ \sqrt{|c_{i,1}|^2 + |c_{i,2}|^2} \right\}.$$

For convenience, we introduce the n-order difference matrix with the von Neumann boundary condition used in the model (1.2) as

$$D = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix}.$$

The gradient operator can be written as $\nabla = (\nabla_x, \nabla_y) := (I \otimes D, D \otimes I)$ based on the formulation used in [36]. The Green Theorem indicates that the divergence operator is given by

$$\text{div}\widehat{\mathbf{d}} = -(I \otimes D^T)\mathbf{d}_1 - (D^T \otimes I)\mathbf{d}_2,$$

for $\widehat{\mathbf{d}} = (\mathbf{d}_1, \mathbf{d}_2) \in \mathbf{Y}$. Then we establish in Theorems 2.1 the existence of solutions of model (1.2).

Theorem 2.1. *If* $\min\{t_1, t_2\} \geq \alpha$, *then there exists a solution to problem (1.2).*

Proof. Denote the objective function in model (1.2) as

$$\mathscr{P}(\mathbf{u}) := \frac{\lambda}{2}\|\mathbf{u} - \mathbf{f}\|_2^2 + \|T(\nabla\mathbf{u})\|_{1,1} - \alpha\|\nabla\mathbf{u}\|_{2,1}, \tag{2.1}$$

referred to as the *primal* problem. Letting $t = \min\{\mathbf{t}_1, \mathbf{t}_2\}$, we have

$$\mathscr{P}(\mathbf{u}) \geq \frac{\lambda}{2}(\|\mathbf{u}\|_2 - \|\mathbf{f}\|_2)^2 + (t - \alpha)\|\nabla\mathbf{u}\|_{2,1},$$

which implies that $\mathscr{P}(\mathbf{u})$ is coercive and nonnegative. Consequently, there exists a minimizing sequence $\{\mathbf{u}^l\}$ such that $\lim_{l\to\infty}\mathscr{P}(\mathbf{u}^l) = \min_{\mathbf{u}}\mathscr{P}(\mathbf{u})$. Due to the coercivity of $\mathscr{P}(\mathbf{u})$, the sequences $\{\|\mathbf{u}^l\|_2\}$, $\{\|\nabla\mathbf{u}^l\|_{1,1}\}$ and $\{\|\nabla\mathbf{u}^l\|_{2,1}\}$ shall be bounded, and hence there exists a convergent subsequence $\{\mathbf{u}^{l_s}\}$ such that $\mathbf{u}^{l_s} \to \mathbf{u}^*$. By using the weak lower semicontinuity of $\mathscr{P}(\mathbf{u})$, we obtain

$$\mathscr{P}(\mathbf{u}^*) \leq \lim_{s\to\infty}\min\mathscr{P}(\mathbf{u}^{l_s}) = \lim_{l\to\infty}\min\mathscr{P}(\mathbf{u}^l) = \min_{\mathbf{u}}\mathscr{P}(\mathbf{u}).$$

Therefore, $\mathbf{u}^*$ is the minimizer to problem (1.2). □

We take a close look at the AWAITV regularization. At every pixel i, we denote $x = (I \otimes D)_i, y = (D \otimes I)_i$ as its derivatives along horizontal and vertical directions, respectively, and $\tau_1 := (\mathbf{t}_1)_i, \tau_2 := (\mathbf{t}_2)_i$. Then the contribution of the pixel i to the AWAITV regularization can be expressed as

$$g(x, y) = \tau_1|x| + \tau_2|y| - \alpha\sqrt{x^2 + y^2}.$$

Without loss of generality, we assume $x \geq 0, y \geq 0$ and hence $g(x, y)$ is differentiable with respect to x and y. Taking derivatives of g gives the optimality condition, i.e.,

$$\frac{\partial g}{\partial x} = \tau_1 - \frac{\alpha x}{\sqrt{x^2 + y^2}} = 0 \quad \text{and} \quad \frac{\partial g}{\partial y} = \tau_2 - \frac{\alpha y}{\sqrt{x^2 + y^2}} = 0,$$

which implies that (x, y) should be parallel to (τ_1, τ_2). By adaptively choosing (τ_1, τ_2) at each pixel, we enforce the reconstructed (x, y) to align with the ideal image gradient. As the ground-truth image is unknown, we choose the weighting vectors in the same way as in [37],

$$\mathbf{t}_1 = \frac{1}{1 + \kappa |G_s * [(I \otimes D) \circ \mathbf{f}]|} \text{ and } \mathbf{t}_2 = \frac{1}{1 + \kappa |G_s * [(D \otimes I) \circ \mathbf{f}]|}, \tag{2.2}$$

where κ is a positive parameter and G_s denotes the Gaussian convolution function with the variance s^2 for smoothing.

To illustrate the advantages of adaptive weights by $\mathbf{t}_1$ and $\mathbf{t}_2$, we present denoising results on two simple images in Figure 1. These two artificial images contain smooth structures as well as piecewise constant regions. We add the Gaussian noise with the variance $\sigma^2 = 0.01$ and recover the images by WDAITV (i.e., by setting $\mathbf{t}_1$ and $\mathbf{t}_2$ as all-one vectors) and AWDAITV with $\mathbf{t}_1$ and $\mathbf{t}_2$ defined in (2.2). Figure 1 presents horizontal profiles and surface plots of the denoised images, showing that the AWDAITV model can efficiently suppress the staircase and preserve smooth regions.

3. NUMERICAL METHOD

Model (1.2) is a nonconvex optimization problem due to the DC term $\|T(\nabla\mathbf{u})\|_{1,1} - \alpha\|\nabla\mathbf{u}\|_{2,1}$. It can be solved via DCA by interweaving two subgradients of $\|T(\nabla\mathbf{u})\|_{1,1}$ and $\alpha\|\nabla\mathbf{u}\|_{2,1}$. But this approach involves a double loop, thus computationally expensive. Instead, we consider the primal dual method (PDM) [35, 39, 40, 41], which is widely used in many fields, especially in image processing.

To apply PDM to minimize the proposed model (1.2), we start by the dual forms of the ℓ_1 norm and anisotropic total variations [42], i.e.,

$$\|T(\nabla\mathbf{u})\|_{1,1} = \max_{\hat{\mathbf{q}} \in \mathscr{C}_1} \langle \mathbf{t}_1 \nabla_x \mathbf{u}, \mathbf{q}_1 \rangle_{\mathbf{X}} + \langle \mathbf{t}_2 \nabla_y \mathbf{u}, \mathbf{q}_2 \rangle_{\mathbf{X}} \tag{3.1}$$

$$\|\nabla\mathbf{u}\|_{2,1} = \max_{\hat{\mathbf{p}} \in \mathscr{C}_2} \langle \nabla\mathbf{u}, \hat{\mathbf{p}} \rangle_{\mathbf{Y}}, \tag{3.2}$$

where $\mathscr{C}_1 := \{\hat{\mathbf{q}} = (\mathbf{q}_1, \mathbf{q}_2) \in \mathbf{Y} \text{ s.t. } \|\mathbf{q}_1\|_\infty \leq 1 \text{ and } \|\mathbf{q}_2\|_\infty \leq 1\}$ and $\mathscr{C}_2 := \{\hat{\mathbf{p}} = (\mathbf{p}_1, \mathbf{p}_2) \in \mathbf{Y} \text{ s.t. } \|\hat{\mathbf{p}}\|_{2,\infty} \leq 1\}$. We further define two indicator functions $\delta_{\mathscr{C}_1}$ and $\delta_{\mathscr{C}_2}$ as

$$\delta_{\mathscr{C}_1}(\hat{\mathbf{q}}) = \begin{cases} 0 & \text{if } \hat{\mathbf{q}} \in \mathscr{C}_1 \\ +\infty & \text{if } \hat{\mathbf{q}} \notin \mathscr{C}_1 \end{cases} \text{ and } \delta_{\mathscr{C}_2}(\hat{\mathbf{p}}) = \begin{cases} 0 & \text{if } \hat{\mathbf{p}} \in \mathscr{C}_2 \\ +\infty & \text{if } \hat{\mathbf{p}} \notin \mathscr{C}_2. \end{cases}$$

Using the dual formulations, the problem (1.2) can be written as a saddle point problem

$$\min_{\mathbf{u}, \hat{\mathbf{p}}} \max_{\hat{\mathbf{q}}} \mathscr{L}(\mathbf{u}, \hat{\mathbf{p}}, \hat{\mathbf{q}}) := \frac{\lambda}{2} \|\mathbf{u} - \mathbf{f}\|_2^2 + \langle \mathbf{t}_1 \circ (I \otimes D)\mathbf{u}, \mathbf{q}_1 \rangle_{\mathbf{X}}$$
$$+ \langle \mathbf{t}_2 \circ (D \otimes I)\mathbf{u}, \mathbf{q}_2 \rangle_{\mathbf{X}} - \delta_{\mathscr{C}_1}(\hat{\mathbf{q}}) - \langle \alpha\nabla\mathbf{u}, \hat{\mathbf{p}} \rangle_{\mathbf{Y}} + \alpha\delta_{\mathscr{C}_2}(\hat{\mathbf{p}}), \tag{3.3}$$

where $\hat{\mathbf{p}} = (\mathbf{p}_1, \mathbf{p}_2)$ and $\hat{\mathbf{q}} = (\mathbf{q}_1, \mathbf{q}_2)$. We can rewrite (3.3) equivalently into a standard saddle point formulation, i.e.,

$$\min_{\mathbf{u}, \hat{\mathbf{p}}} \max_{\hat{\mathbf{q}}} F(\mathbf{u}, \hat{\mathbf{p}}) + \langle K(\mathbf{u}), \hat{\mathbf{q}} \rangle_{\mathbf{Y}} - G(\hat{\mathbf{q}}), \tag{3.4}$$

where $F(\mathbf{u}, \hat{\mathbf{p}}) := \frac{\lambda}{2} \|\mathbf{u} - \mathbf{f}\|_2^2 - \langle \alpha\nabla\mathbf{u}, \hat{\mathbf{p}} \rangle_{\mathbf{Y}} + \alpha\delta_{\mathscr{C}_2}(\hat{\mathbf{p}})$, $K(\mathbf{u}) = (\mathbf{t}_1 \circ (I \otimes D)\mathbf{u}, \mathbf{t}_2 \circ (D \otimes I)\mathbf{u})$, and $G(\hat{\mathbf{q}}) := \delta_{\mathscr{C}_1}(\hat{\mathbf{q}})$.
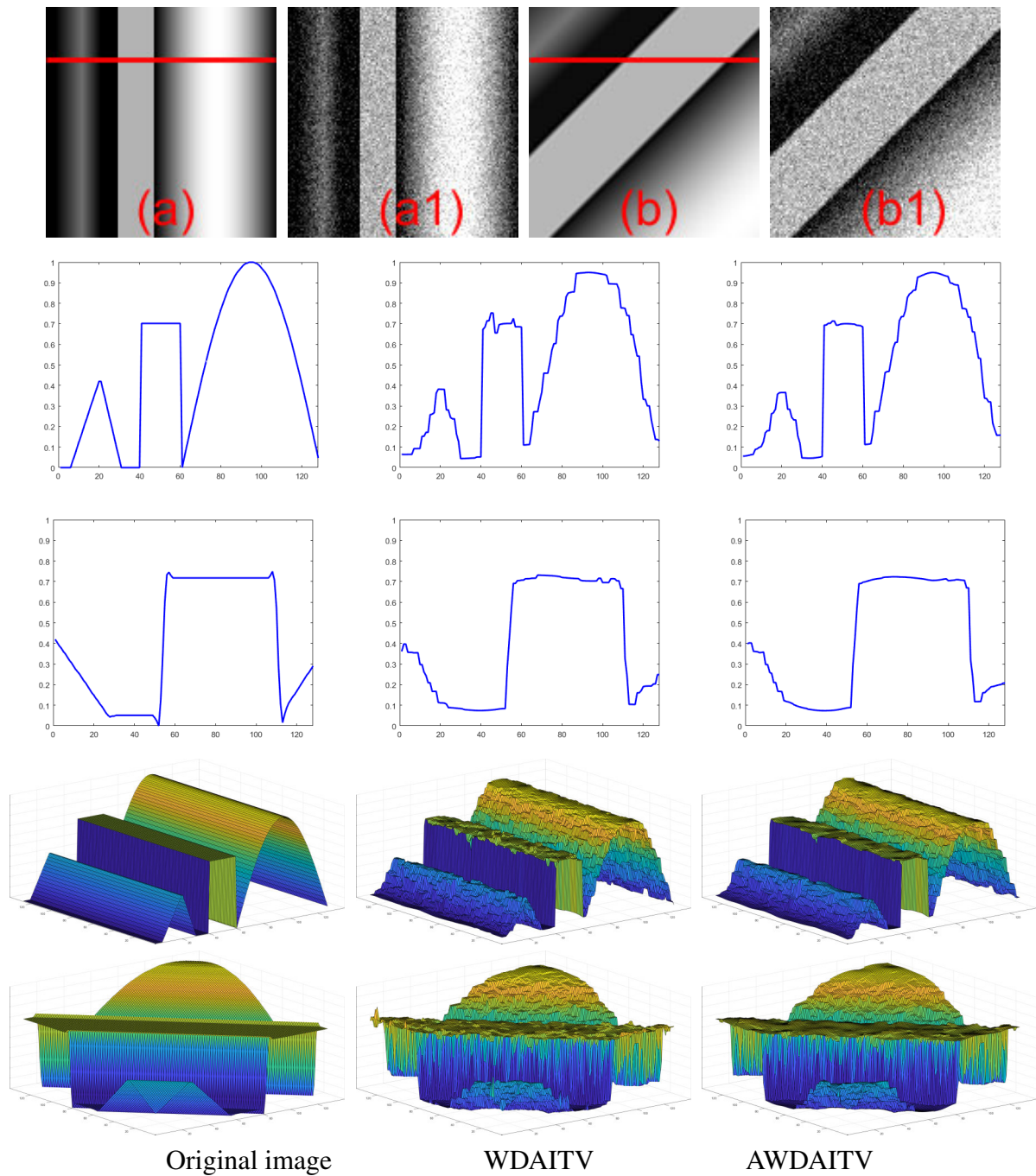
FIGURE 1.   The first row demonstrates synthetic images and their noisy versions with noise variance of $\sigma^2 = 0.01$. The second/third rows plot the horizontal profiles for the ground-truth images and the denoised images by WDAITV and AWDAITV; the corresponding surface plots are provided in the last two rows.

Notice that the function $F(\mathbf{u}, \hat{\mathbf{p}})$ in problem (3.4) is nonconvex due to the coupling between $\mathbf{u}$ and $\hat{\mathbf{p}}$, while $G(\hat{\mathbf{q}})$ is convex. Then problem (3.4) is the nonconvex-concave (but not strongly concave) min-max problem. There are two types of algorithms that solve for the general min-max problem. One is a nested-loop type of algorithms [43, 44] that either employ multiple

gradient ascent steps for updating the dual variables to solve an inner subproblem (exactly or inexactly) or further apply a similar scheme for updating the primal variables. As the name implies, the nested loops are computationally expensive to implement. The second type is single-loop algorithms [34, 45, 46]. For example, gradient descent-ascent (GDA) method [38] performs a gradient descent step on the primal variable and a gradient ascent step on the dual variable simultaneously at each iteration. As an extension of GDA, the primal-dual hybrid gradient (PDHG) proposed in [34] is one of the most popular approaches for solving minimax problem (3.4). To improve the convergence speed of gradient-based methods, some extrapolation techniques [47] are often employed.

Here we focus on an extrapolation technique, referred to as primal dual method (PDM) [35], which consists of two steps: the primal update and the dual update. The primal step minimizes the combination of the Lagrangian function and the proximal terms, i.e.,

$$F(\mathbf{u}, \hat{\mathbf{p}}) + \langle K(\mathbf{u}), \hat{\mathbf{q}} \rangle_{\mathbf{Y}} + \frac{1}{2\tau} \left\| \mathbf{u} - \mathbf{u}^k \right\|^2 + \frac{1}{2\tau} \left\| \hat{\mathbf{p}} - \hat{\mathbf{p}}^k \right\|^2,$$

to update the primal variables $\mathbf{u}$ and $\hat{\mathbf{p}}$, while the dual step involves a dual ascent based on the consensus residual [47]. Specifically, applying the PDM scheme for minimizing (3.4) can be written as

$$\begin{cases} \mathbf{u}^{k+1} = \underset{\mathbf{u}}{\operatorname{argmin}} F\left(\mathbf{u}, \hat{\mathbf{p}}^k\right) + \left\langle K(\mathbf{u}), \hat{\mathbf{q}}^k \right\rangle_{\mathbf{Y}} + \frac{1}{2\tau} \left\| \mathbf{u} - \mathbf{u}^k \right\|^2, & (3.5) \\[2mm] \hat{\mathbf{p}}^{k+1} = \underset{\hat{\mathbf{p}}}{\operatorname{argmin}} F\left(\mathbf{u}^{k+1}, \hat{\mathbf{p}}\right) + \frac{1}{2\tau} \left\| \hat{\mathbf{p}} - \hat{\mathbf{p}}^k \right\|^2, & (3.6) \\[2mm] \bar{\mathbf{u}}^k = 2\mathbf{u}^{k+1} - \mathbf{u}^k, & (3.7) \\[2mm] \hat{\mathbf{q}}^{k+1} = \underset{\hat{\mathbf{q}}}{\operatorname{argmax}} \left\langle K(\bar{\mathbf{u}}^k), \hat{\mathbf{q}} \right\rangle_{\mathbf{Y}} - G(\hat{\mathbf{q}}) - \frac{1}{2\xi} \left\| \hat{\mathbf{q}} - \hat{\mathbf{q}}^k \right\|^2, & (3.8) \end{cases}$$

where $\tau > 0$ and $\xi > 0$ are the stepsize parameters of the primal and dual variables, respectively. In the following, we describe the details for solving the subproblems (3.5)-(3.8).

- As subproblem (3.5) is smooth and convex, the optimality condition yields the closed-form solution as

$$\mathbf{u}^{k+1} = \frac{\mathbf{u}^k + \tau \left(\lambda \mathbf{f} + \operatorname{div}\hat{\mathbf{q}}_{\mathbf{t}}^k - \alpha \operatorname{div}\hat{\mathbf{p}}^k \right)}{1 + \lambda \tau}, \tag{3.9}$$

where $\hat{\mathbf{q}}_{\mathbf{t}} := (\mathbf{t}_1 \circ \mathbf{q}_1, \mathbf{t}_2 \circ \mathbf{q}_2)$.

- As subproblem (3.6) involves an indicator function, we apply the gradient projection method, leading to

$$\hat{\mathbf{p}}^{k+1} = \mathscr{P}_{\mathscr{C}_2}(\hat{\mathbf{p}}^k + \tau \alpha \nabla \mathbf{u}^{k+1}) = \frac{\hat{\mathbf{p}}^k + \tau \alpha \nabla \mathbf{u}^{k+1}}{\max\left(1, \left| \hat{\mathbf{p}}^k + \tau \alpha \nabla \mathbf{u}^{k+1} \right| \right)}, \tag{3.10}$$

in which the division is carried out elementwise.

- As for subproblem (3.8), we can rewrite it as

$$\hat{\mathbf{q}}^{k+1} = \underset{\hat{\mathbf{q}}}{\operatorname{argmin}} G(\hat{\mathbf{q}}) - \left\langle K(\bar{\mathbf{u}}^k), \hat{\mathbf{q}} \right\rangle + \frac{1}{2\xi} \left\| \hat{\mathbf{q}} - \hat{\mathbf{q}}^k \right\|^2,$$
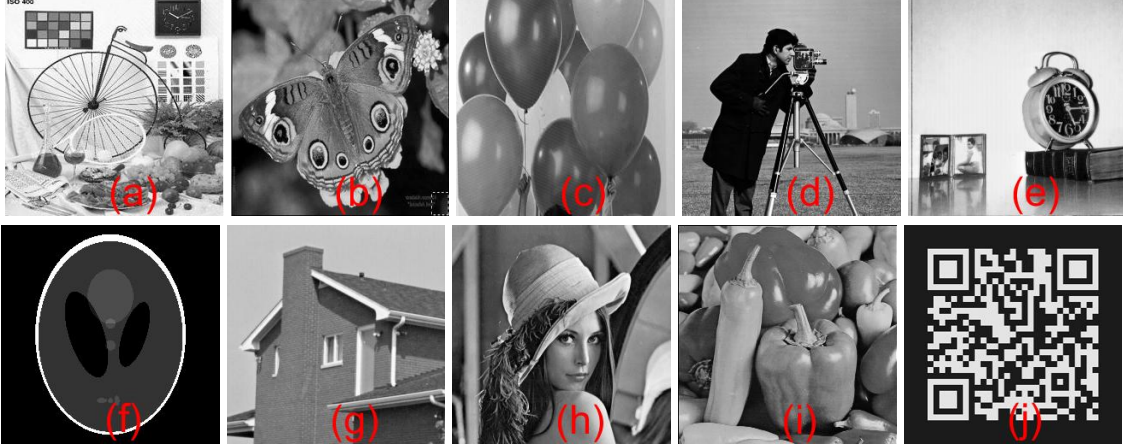
FIGURE 2.   Original images with the size 256×256.

which is convex. We apply the gradient projection method to find the closed-form solution of $\hat{\mathbf{q}}^{k+1} = (\mathbf{q}_1^{k+1}, \mathbf{q}_2^{k+1})$, i.e.,

$$
\begin{cases}
\mathbf{q}_1^{k+1} & = \dfrac{\mathbf{q}_1^k + \xi \mathbf{t}_1 \circ (I \otimes D) \bar{\mathbf{u}}^k}{\max\left(1, \mathbf{q}_1^k + \xi \mathbf{t}_1 \circ (I \otimes D) \bar{\mathbf{u}}^k\right)}, \\
\mathbf{q}_2^{k+1} & = \dfrac{\mathbf{q}_2^k + \xi \mathbf{t}_2 \circ (D \otimes I) \bar{\mathbf{u}}^k}{\max\left(1, \mathbf{q}_2^k + \xi \mathbf{t}_2 \circ (D \otimes I) \bar{\mathbf{u}}^k\right)}.
\end{cases}
\tag{3.11}
$$

Given parameters $\lambda, \alpha, \xi, \tau > 0$, we choose the initial values of $\mathbf{u}^0, \hat{\mathbf{p}}^0, \hat{\mathbf{q}}^0$ to run the iterations of (3.5)-(3.8) until the relative error (RE) reaches to

$$
\text{RE} := \|\mathbf{u}^{k+1} - \mathbf{u}^k\| / \|\mathbf{u}^k\| \le 10^{-5},
\tag{3.12}
$$

or the number of iterations exceeds 500. Due to the nonconvexity of model (3.3), it is challenging to prove the convergence of the algorithm. We empirically validate the convergence of primal/dual variables in Section 4.2. The convergence proof will be left as a future work.

4. EXPERIMENTS

In this section, we test on ten (original) images, shown in Figure 2, to evaluate the effectiveness of the proposed AWDAITV model. For the ease of parameter tuning (which will be elaborated on Section 4.1), we normalize the intensity value of each testing image to $[0, 1]$ before adding the Gaussian noise by using the Matlab function "imnoise." We consider three noise levels of variance σ^2 as 0.01, 0.05, and 0.1. We use the signal to noise ratio (SNR) and the structural similarity index (SSIM) to quantitatively evaluate the denoising performance. All the numerical experiments are performed in Matlab (R2022a) on a windows11 (64bit) desktop computer with an Intel(R) Core(TM) i7-11700 2.50GHz CPU and 16.0GB RAM.

4.1. **Parameter setting.** There are two types of parameters in the proposed algorithm: stepsize parameters τ, ξ and model parameters α, λ. For the stepsize parameters, we observe that a constant value of the product $\tau\xi$ achieves the best denoising results consistently. After fine-tuning, we set $\tau = \xi = \frac{1}{2}$ that works the best for most of the test images, and meanwhile this combination satisfies a convergence condition of the PDM method [35] for convex optimization problems. The value of $\alpha \in [0, 1]$ depends image structures. For an image with simple geometries such as piecewise constant, we set α to be close to 1 in order to enforce the sparseness after

taking the gradient transform. Taking a phantom image of Figure 2(f) and a barcode image of Figure 2(j) for an example, we set α to be 0.5 or 0.7 depending on the noise level. For other images with relatively more complex structures, we set $\alpha = 0.01$. The parameter λ controls how much smoothing is introduced by the regularization in an attempt to filter out the noise without losing too much information when approximating the underlying image. There are a number of automated and semi-automated approaches for finding the optimal regularization parameter λ, such as the L-curve method, the generalized cross validation (GCV), and unbiased predictive risk estimator (UPRE) and Stein's unbiased risk estimator (SURE), all described in Chapter 4 of [6]. However, these methods require the data fitting term and the regularization term to lie in the same functional space, thus not applicable to the proposed AWDAITV model, in which the data fitting term lies in the ℓ^2 space and the regularization term lies in the bounded variation space. In practice, we choose the regularization parameter λ based on trial and error. As for the weight vectors $(\mathbf{t}_1, \mathbf{t}_2)$ used in AWDAITV (1.2), we set them according to (2.2) with $s = 0.5$ and κ as $\{30, 10, 8\}$ respectively for three noise variances from low to high.

4.2. **Validation on convergence.** The convergence of PDM was established in [35] for a convex, closed, and proper objective function, which is unfortunately not applicable to the non-convex AWDAITV model. Instead of a rigorous convergence proof, we analyze the convergence of PDM empirically by examining relative errors (RE) as defined in (3.12) and the objective function values (1.2). We also consider the primal-dual gap (PDG) [48] to evaluate the convergence. Based on the Legendre-Fenchel conjugation, we express the dual formulation of the primal problem (1.2) as

$$\underbrace{\max_{\mathbf{p} \in \delta_{C_2}, \mathbf{q} \in \delta_{C_1}} -\frac{1}{2\lambda} \|\mathrm{div}(T(\hat{\mathbf{q}})) - \alpha \mathrm{div}\hat{\mathbf{p}}\|^2 - \langle (\mathrm{div}(\mathbf{Tq}) - \alpha \mathrm{div}\mathbf{p}), \mathbf{f} \rangle_{\mathbf{X}}}_{\mathscr{D}(\hat{\mathbf{p}}, \hat{\mathbf{q}})}, \qquad (4.1)$$

where δ_{C_1} and δ_{C_2} are defined in (3.1) and (3.2). Assume that $\mathbf{u}^*$ is a solution that minimizes the primal problem $\mathscr{P}(\mathbf{u})$ defined in (2.1) and $(\hat{\mathbf{p}}^*, \hat{\mathbf{q}}^*)$ is a pair of solutions to dual problem (4.1). Then the weak duality always holds [49], i.e.,

$$\mathscr{P}(\mathbf{u}^*) \geq \mathscr{D}(\hat{\mathbf{p}}^*, \hat{\mathbf{q}}^*).$$

If the above inequality becomes equality, namely, the strong duality holds, then it implies that the iterations reach to a saddle point of the problem (3.4). Hence, we can define the primal-dual gap:

$$\mathrm{PDG} = \mathscr{P}\left(\mathbf{u}^k\right) - \mathscr{D}\left(\mathbf{p}^k, \mathbf{q}^k\right).$$

as a measure for the convergence of the iterations (3.5)-(3.8).

We add Gaussian noises to the three testing images of Figure 2 (g), (h), (i) with three noise levels (variance $\sigma^2 = 0.01, 0.05$, and 0.1). Figure 3 presents RE between two consecutive solutions, objective functions, and PDG values with respect to iteration numbers in a logarithmic scale. We observe that all the RE curves decrease till 10^{-4}, which indicates the convergence of

---

Conjugate of a function $h : \mathbf{X} \to \mathbb{R}$ is defined by

$$h^*(\mathbf{y}) = \max_{\mathbf{x} \in \mathbf{X}} \{\langle \mathbf{x}, \mathbf{y} \rangle - h(\mathbf{x})\}.$$
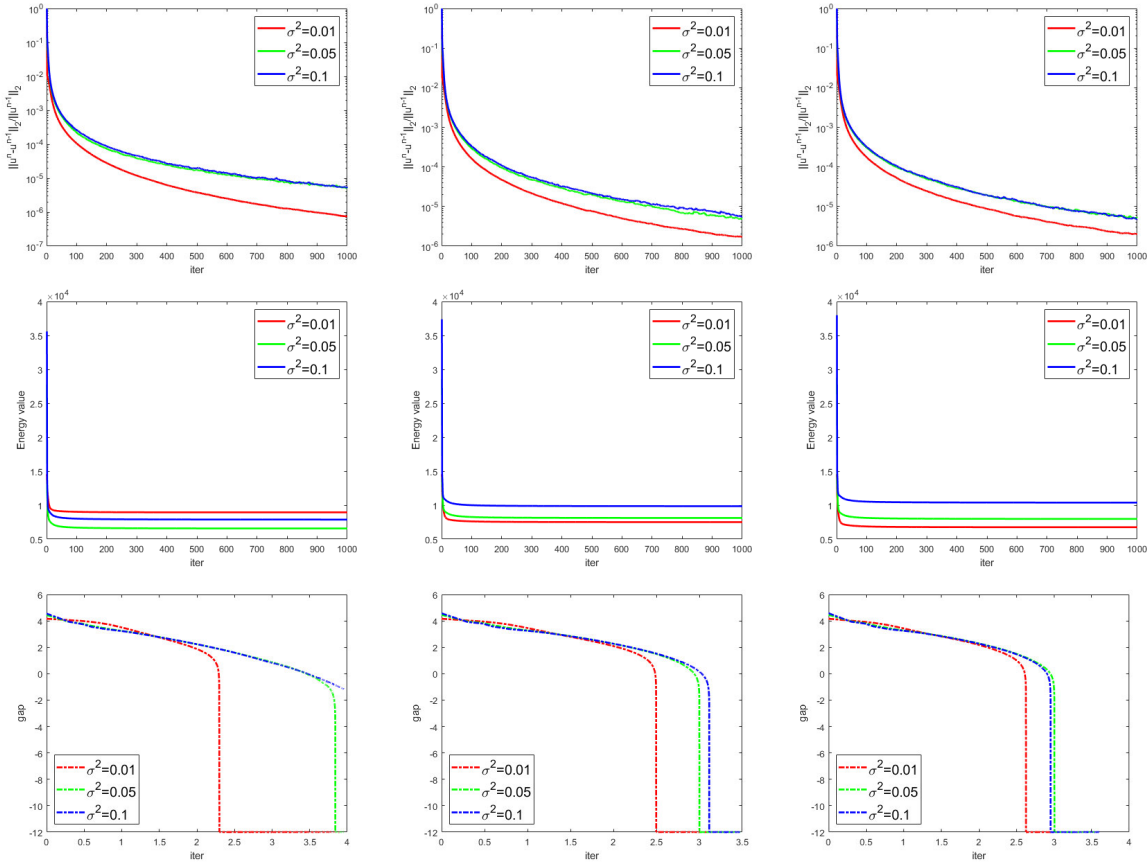
FIGURE 3.   Empirical convergence analysis of the RE curves (top), the objective function values (2.1) (middle), and the PDG plots in a logarithmic axis (bottom) against the iteration number based on three images: Figure 2 (g), (h), (i).

the sequence $\{\mathbf{u}^k\}$. The objective values become flat after 200 iterations, showing that $\{\mathbf{u}^k\}$ may converge to a stationary point of model (1.2). Lastly, PDG values are monotonically decreasing to zero, demonstrating that the algorithm converges to a saddle point. The theoretical proof of PDM for a non-convex problem will be left as future work.

4.3. **Algorithmic comparison.** The proposed AWDAITV model reduces to the weighted difference of anisotropic and isotropic total variation (WDAITV) [27] by setting $\mathbf{t}_1 = \mathbf{t}_2 = (1, 1, \cdots, 1) \in \mathbb{R}^{n^2}$ in (1.2), i.e.,

$$\min_{\mathbf{u}} \frac{\lambda}{2} \|\mathbf{u} - \mathbf{f}\|_2^2 + \|\nabla \mathbf{u}\|_{1,1} - \alpha \|\nabla \mathbf{u}\|_{2,1}, \tag{4.2}$$

which was originally minimized via DCA with a subproblem problem solved by ADMM [27]. We compare the performance of PDM and DCA for solving the same WDAITV model (4.2) using a testing image of Figure 2 (a) under three noise levels. For this example, we set $\alpha = 0.01$ and report the denoising results quantitatively in terms of SNR, SSIM, and computational time in Table 1, which clearly demonstrates the advantages of PDM over DCA. Specifically, PDM always achieves better recovery results in much less time.

TABLE 1. Comparisons between DCA and PDM of denoising Figure 2 (a) under the different noise variance σ^2.

NoiseVariance	Method	λ	SNR	SSIM	TIME
$\sigma^2 = 0.01$	DCA	16	20.9563	0.7700	7.7272
	PDM	16	21.9260	0.8082	0.3078
$\sigma^2 = 0.05$	DCA	7.2	16.5140	0.5795	7.6989
	PDM	7.2	16.9596	0.6145	0.2592
$\sigma^2 = 0.1$	DCA	5.3	14.3930	0.4895	7.8182
	PDM	5.3	14.7296	0.5134	0.2428

4.4. **Comparison with other related models.** Here we compare the proposed AWDAITV with other gradient-based denoising methods, including total variation (TV) [50], high-order total variation (HOTV) [16], total generalized variation (TGV) [15] and the WDAITV model (4.2). For these competing methods, we use the codes provided by the respective authors on their websites. The optimal parameter of λ for each combination of testing images, noise levels, and competing methods is listed in Table 2.

We present the denoising results of ten images as demonstrated in Figure 2 (a)-(j) in terms of SNR and SSIM in Tables 3, 4, and 5 under different noise variances $\sigma^2 = 0.01, 0.05, 0.1$, respectively. For each image, the best SNR/SSIM value is highlighted in bold. In most cases, the AWDAITV achieves the highest SNR and SSIM, especially when the original image is piecewise constant and the noise level is low. The sparsity of the gradient can be severely altered by a high level noise, in which case higher-order TV models are more effective. In addition, we include the average of SNRs and SSIMs obtained from all the test images in the last row of each table, demonstrating that AWDAITV generally yields competitive results.

TABLE 2. The optimal regularization parameter λ in the general model (1.1).

	$\sigma^2 = 0.01$					$\sigma^2 = 0.05$					$\sigma^2 = 0.1$				
	TV	HOTV	TGV	WAITV	AWAITV	TV	HOTV	TGV	WDAITV	AWDAITV	TV	HOTV	TGV	WDAITV	AWDAITV
(a)	21.7	47.0	47.0	27.0	16.0	8.4	16.0	16.3	10.1	7.2	6.0	11.0	11.0	7.2	5.3
(b)	17.2	34.0	34.3	21.0	13.0	7.1	12.0	12.2	8.4	5.9	5.3	8.0	8.2	6.2	4.0
(c)	11.9	20.0	19.7	14.0	11.0	5.3	7.0	6.6	6.2	4.6	4.1	5.0	4.6	4.7	4.0
(d)	15.2	31.0	30.7	18.0	11.0	6.4	11.0	11.0	7.5	4.0	4.7	7.0	7.2	5.5	3.0
(e)	14.9	29.0	29.0	17.8	10.0	6.8	11.0	11.3	8.1	5.0	5.1	8.0	7.9	6.0	3.0
(f)	14.0	36.0	37.0	17.0	4.9	7.0	16.0	16.0	9.0	2.3	6.0	11.0	11.0	6.0	1.7
(g)	12.0	22.0	21.8	13.9	10.0	5.5	8.0	7.8	6.2	4.3	4.3	6.0	5.6	4.8	3.2
(h)	14.5	28.0	27.7	17.6	12.8	6.2	10.0	10.0	7.4	5.3	4.9	7.0	6.8	5.7	4.0
(i)	14.2	28.0	27.9	17.0	12.0	6.3	10.0	10.1	7.4	5.0	4.8	7.0	6.7	5.6	4.1
(j)	15.0	36.0	36.0	16.0	3.6	8.0	19.0	18.0	9.0	1.7	6.0	13.0	13.0	7.0	1.0

We present the denoising results visually in Figure 4 where each testing image is contaminated by the middle level of noise, i.e., $\sigma^2 = 0.05$. For a better visual comparison, we zoom in on a region of interest for each image, as indicated by a red square in Figure 4, and the corresponding results are illustrated in Figure 5.

We can observe that the TV model suffers from staircasing artifacts in the smooth region; see the cheek of Lena (h) in Figure 5. Both HOTV and TGV produce over-smooth outputs. HOTV is based on the fourth-order diffusion that damps oscillations faster than second-order diffusion;

TABLE 3.   SNR (dB) and SSIM for denoising results of $\sigma^2 = 0.01$.

	TV	HOTV	TGV	WDAITV	AWDAITV
(a)	21.7704	21.8540	21.8232	21.7762	**21.9260**
	0.7953	0.7953	0.7944	0.7921	**0.8082**
(b)	18.4554	18.7261	18.7289	18.4715	**18.7737**
	0.7511	0.7583	0.7574	0.7494	**0.7807**
(c)	24.5387	24.3599	24.3284	24.5048	**24.6511**
	0.8378	0.8260	0.8272	0.8380	**0.8572**
(d)	21.3756	21.0825	21.0798	21.3721	**21.5685**
	0.7422	0.7102	0.7130	0.7419	**0.7709**
(e)	25.8937	25.7574	25.7486	26.0205	**26.1341**
	0.8547	0.8216	0.8213	0.8534	**0.8894**
(f)	15.6145	14.3719	14.2776	15.8152	**16.8022**
	0.4489	0.3683	0.3638	0.4445	**0.4806**
(g)	24.7755	24.3922	24.4057	24.7727	**24.8569**
	0.7998	0.7632	0.7647	0.7984	**0.8112**
(h)	20.8174	20.9863	**21.0359**	20.6843	20.7236
	0.7804	0.7771	0.7803	0.7755	**0.7977**
(i)	21.4537	21.4435	21.4409	21.5778	**22.1112**
	0.7995	0.7921	0.7925	0.7998	**0.8335**
(j)	21.9027	20.9235	20.8956	23.1444	**24.1201**
	0.9529	0.8723	0.8717	0.9567	**0.9867**
Average	21.6598	21.3897	21.3765	21.8140	**22.1667**
	0.7763	0.7484	0.7486	0.7750	**0.8016**

as a result, it may blur the edges when keeping the smooth regions. TGV has a coupling effect of the total variation and the high-order total variation, so it can simultaneously preserve edges and piecewise smoothing regions. Comparing WDAITV and AWDAITV, the latter enhances the diffusion strength along with the important features, leading to better denoising results than WDAITV in the areas such as the tripod in the Cameraman (d) and the pointer in Clock (e).

5. CONCLUSIONS

This paper proposed a novel denoising model based on the weighted difference of anisotropic and isotropic total variation, where adaptive weights were incorporated to enhance the robustness for the proposed model. We transformed the nonconvex and nonsmooth model into a saddle point problem and applied the primal dual method to find the model solution. We validated the convergence of the proposed numerical algorithm by examining on the relative errors, objective function values, and primal-dual gaps. Experiments demonstrated the performance of the proposed AWAITV model in comparison to other gradient-based denoising methods. One future work lies in the proof of the PDM convergence for a nonconvex problem.
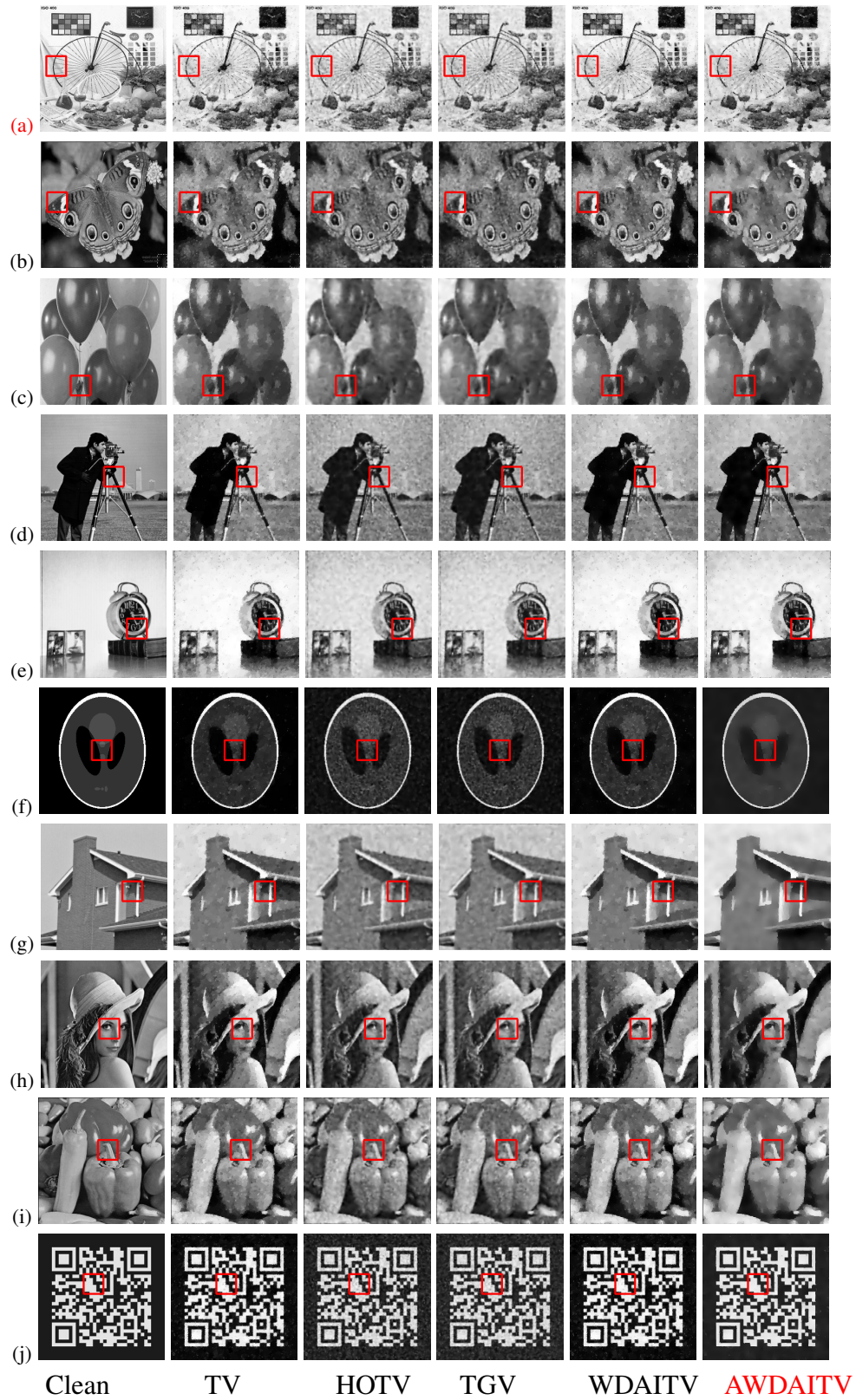
FIGURE 4. The comparison of denoising results for $\sigma^2 = 0.05$ with red patches zoomed in Figure 5.

| Clean | TV | HOTV | TGV | WDAITV | AWDAITV |

FIGURE 5. Zoomed-in patches in Figure 4.

TABLE 4. SNR (dB) and SSIM for denoising results of $\sigma^2 = 0.05$.

	TV	HOTV	TGV	WDAITV	AWDAITV
(a)	16.8869	**17.0521**	17.0285	16.8855	16.9596
	0.6090	**0.6152**	0.6135	0.6061	0.6145
(b)	13.9761	14.1648	**14.1704**	13.9354	14.0102
	0.5555	0.5672	0.5652	0.5525	**0.5851**
(c)	20.4466	**20.6000**	20.5933	20.4161	20.2707
	0.7600	0.7601	0.7663	0.7562	**0.7917**
(d)	16.7447	16.4686	16.4764	**16.7360**	16.5872
	0.5874	0.5386	0.5392	0.5855	**0.6385**
(e)	19.9847	**20.1407**	20.1329	19.9897	19.8366
	0.7464	0.7278	0.7212	0.7402	**0.7924**
(f)	8.9845	8.4277	8.4029	9.0897	**9.4248**
	0.3574	0.2617	0.2618	0.3280	**0.4202**
(g)	**20.4173**	20.2242	20.2267	20.4055	20.3195
	0.7145	0.6742	0.6780	0.7127	**0.7418**
(h)	16.5232	16.7508	**16.7739**	16.4227	16.2548
	0.6397	0.6430	0.6447	0.6311	**0.6688**
(i)	17.0971	17.1189	17.1359	17.1364	**17.4178**
	0.6590	0.6605	0.6599	0.6571	**0.7154**
(j)	14.7053	14.6680	14.6436	15.1568	**16.0131**
	0.8172	0.7081	0.7214	0.8139	**0.9212**
Average	16.5766	16.5616	16.5585	16.6174	**16.7094**
	0.6446	0.6156	0.6171	0.6383	**0.6890**

TABLE 5. SNR (dB) and SSIM for denoising results of $\sigma^2 = 0.1$.

	TV	HOTV	TGV	WAITV	AWAITV
(a)	14.6912	**14.8568**	14.8499	14.6930	14.7296
	0.5075	0.5186	**0.5188**	0.5055	0.5134
(b)	12.2366	**12.4479**	12.4461	12.2053	12.1697
	0.4720	0.4914	0.4887	0.4695	**0.5045**
(c)	18.1395	18.4052	**18.4182**	18.1121	18.1914
	0.7132	0.7137	0.7254	0.7180	**0.7465**
(d)	**14.6799**	14.5404	14.5467	14.6553	14.3534
	0.5279	0.4948	0.4893	0.5225	**0.5808**
(e)	16.9632	**17.1458**	17.1449	16.9605	16.7517
	0.6800	0.6626	0.6664	0.6760	**0.7404**
(f)	6.4059	6.1578	6.1528	**6.4385**	6.3642
	0.2559	0.2254	0.2556	0.3205	**0.3896**
(g)	17.9045	18.0502	**18.0694**	17.9309	17.7738
	0.6587	0.6220	0.6338	0.6569	**0.7011**
(h)	14.5650	14.9298	**14.9477**	14.5025	14.2279
	0.5684	0.5888	0.5942	0.5662	**0.6199**
(i)	15.1912	15.3547	**15.3708**	15.1958	15.2987
	0.5911	0.6022	0.6115	0.5878	**0.6518**
(j)	11.3934	11.5947	11.5955	11.6330	**12.2286**
	0.6967	0.6146	0.6150	0.6817	**0.8285**
Average	14.2170	14.3483	**14.3542**	14.2327	14.2089
	0.5671	0.5534	0.5599	0.5697	**0.6277**

Acknowledgments

References

[1] G. Aubert, P. Kornprobst, Mathematical Problem in Image Processing: Partial Differential Equations and the Calculus of Variations, Springer, 147, 2008.

[2] O. Scherzer, Handbook of Mathematical Methods in Imaging, Springer, 2015.

[3] N. Brás, J. Bioucas-Dias, R. Martins, A. Serra, An alternating direction algorithm for total variation reconstruction of distributed parameters. IEEE Trans. Image Process. 21 (2012), 3004-3016.

[4] K. Papafitsoros K, C. Schonlieb, B. Sengu, Combined first and second order total variation inpainting using split Bregman, Image Processing On Line, 112-136, 2013.

[5] N. Paragios, Y. Chen, O. Faugeras, Handbook of Mathematical Models in Computer Vision, Springer, 2006.

[6] C. Vogel, Computational Methods for Inverse Problems, SIAM, 2002.

[7] Y. Wang, J. Yang, W. Yin, Y. Zhang, A new alternating minimization algorithm for total variation image reconstruction, SIAM J. Imaging Sci. 1 (2008), 248-272.

[8] A. Buades, B. Coll, J. Morel, A non-local algorithm for image denoising, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2 (2005), 60-65.

[9] W. Dong, L. Zhang, G. Shi, X. Li, Nonlocally centralized sparse representation for image restoration, IEEE Trans. Image Process. 22 (2013), 1620-1630.

[10] G. Gilboa, S. Osher, Nonlocal operators with applications to image processing, Multiscale Modeling and Simulation, 7 (2008), 1005-1028.

[11] F. Karam, D. Meskine, K. Sadik, Nonlocal total variation system for the restoration of textured images, Int. J. Comput. Math. 98 (2021), 1749-1768.

[12] G. Peyre, S. Bougleux, L. Cohen, Non-local regularization of inverse problems, Inverse Probl. Imaging 5 (2011), 511-530.

[13] J. Yang, F. Liu, H. Yue, X. Fu, C. Hou, F. Wu, Textured image demoireing via signal decomposition and guided filtering, IEEE Trans. Image Process. 26 (2017), 3528-3541.

[14] H. Yue, X. Sun, S. Member, J. Yang, F. Wu, Image denoising by exploring external and internal correlations, IEEE Trans. Image Process. 24 (2015), 1967-1982.

[15] K. Bredies, K. Kunisch, T. Pock, Total generalized variation, SIAM J. Imaging Sci. 3 (2010), 492-526.

[16] M. Lysaker, A. Lundervold, X. Tai, Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time, IEEE Trans. Image Process. 12 (2013), 1579-1590.

[17] Y. You, M. Kaveh, Fourth-order partial differential equations for noise removal, IEEE Trans. Image Process. 9 (2000), 1723-1730.

[18] W. Bian, X. Chen, A smoothing proximal gradient algorithm for nonsmooth convex regression with cardinality penalty, SIAM J. Numer. Anal. 58 (2020), 858-883.

[19] A. Lanza, S. Morigi, I. Selesnick, F. Sgallari, Sparsity-inducing nonconvex nonseparable regularization for convex image processing, SIAM J. Imaging Sci. 12 (2019), 1099-1134.

[20] Y. Sun, X. Tan, X. Li, L. Lei, G. Kuang, Sparse optimization problem with s-difference regularization, Signal Processing, 168 (2020), 107369.

[21] C. Zhang, X. Chen, A smoothing active set method for linearly constrained non-Lipschitz nonconvex optimization, SIAM J. Optim. 30 (2020), 1-30.

[22] X. Zhang, M. Bai, M. Ng, Nonconvex-TV based image restoration with impulse noise removal, SIAM J. Imaging Sci. 10 (2017), 1627-1667.

[23] R. Chartrand, Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data. IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 262-265, 2019.

[24] W. Jiang, F. Nie, H. Huang, Robust dictionary learning with capped ℓ_1-norm. AAAI Press, 2015.

[25] Z. Xu, H. Zhang, Y. Wang, X. Chang, Y. Liang, $L_{1/2}$ regularization: a thersholding representation theory and a fast solver, IEEE Trans. Neural Netw. Learn. Sys. 23 (2012), 1013-1027.

[26] J. You, Y. Jiao, X. Lu, T. Zeng, A nonconvex model with minimax concave penalty for image restoration, J. Sci. Comput. 78 (2019), 1063-1086.

[27] Y. Lou, T. Zeng, S. Osher, J. Xin, A weighted difference of anisotropic and isotropic total variation model for image processing. SIAM J. Imaging Scie. 8 (2015), 1798-1823.

[28] K. Bui, F. Park, Y. Lou, J. Xin, A weighted difference of anisotropic and isotropic total variation for relaxed Mumford-Shah color and multiphase image segmentation, SIAM J. Imaging Sci. 14 (2021), 1078-1113.

[29] F. Park, Y. Lou, J. Xin. A weighted difference of anisotropic and isotropic total variation for relaxed Mumford-Shah image segmentation, IEEE International Conference on Image Processing, 4314-4318, 2016.

[30] P. Li, W. Chen, H. Ge, M. Ng. $\ell^1 - \alpha\ell^2$ minimization methods for signal and image reconstruction with impulsive noise removal, Inverse Probl. 36 (2020), 055009.

[31] H. Thi, An efficient algorithm for globally minimizing a quadratic function under convex quadratic constraints, Math. Program. 87 (2020), 401-426.

[32] H. Thi, T. Dinh, The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems, Ann. Oper. Res. 133 (2005), 23-46.

[33] H. Thi, T. Dinh, DC programming and DCA: thirty years of developments, Math. Program. 169 (2018), 5-68.

[34] M. Zhu, T. Chan, An efficient primal-dual hybrid gradient algorithm for total variation image restoration, UCLA CAM Report, 34: 08-34, 2008.

[35] A. Chambolle, T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, J. Math. Imaging Vision 40 (2011), 120-145.

[36] G. Steidl, A note on the dual treatment of higher-Order regularization functionals, Computing 76 (2016), 135-148.

[37] Z.F. Pang, Y.M. Zhou, T.T. Wu, D.J. Li, Image denoising via a new anisotropic total-variation-based model, Signal Processing: Image Communication 74 (2019), 140-152.

[38] A. Beck, First-Order Methods in Optimization, SIAM, 2017.

[39] M. Benning, M. Burger, Modern regularization methods for inverse problems, Acta Numer. 27 (2018), 1-111.

[40] B. Chen, Z. Zhang, D. Xia, E. Sidky, X. Pan, Non-convex primal-dual algorithm for image reconstruction in spectral CT, Comput. Medical Imaging 87 (2021), 101821.

[41] C. Clason, S. Mazurenko, T. Valkonen, Acceleration and global convergence of a first-order primal-dual method for nonconvex problems, SIAM J. Optim. 29 (2019), 933-963.

[42] L. Condaty, Discrete total variation: New definition and minimization, SIAM J. Imaging Sci. 10 (2017), 1158-1290.

[43] H. Rafique, M. Liu, Q. Lin, T. Yang, Weakly-convex-concave min-max optimization: provable algorithms and applications in machine learning, Optim. Meth. Softw. 37 (2022), 1087-1121.

[44] K. Thekumparampil, P. Jain, P. Netrapalli, S. Oh, Efficient algorithms for smooth minimax optimization, NeurIPS 32 (2019), 12659-12670.

[45] T. Lin, C. Jin, M. Jordan, On gradient descent ascent for nonconvex-concave minimax problems, Proceedings of the 37th International Conference on Machine Learning, 119 (2020), 6083-6093.

[46] Z. Xu, H. Zhang, Y. Xu, G. Lan, A unified single-loop alternating gradient projection algorithm for nonconvex-concave and convex-nonconcave minimax problems, ArXiv:2006.02032, 2020.

[47] K. Wang, H. He, A double extrapolation primal-dual algorithm for saddle point problems, J. Sci. Computing 85 (2020), 1-30.

[48] N. Komodakis, J. Pesquet. Playing with duality: an overview of recent primal-dual approaches for solving large-scale optimization problems, IEEE Signal Processing Magazine, 32 (2015), 31-54.

[49] Y. Chen, G. Lan, Y. Ouyang, Optimal primal-dual methods for a class of saddle point problems, SIAM J. Optim. 24 (2014), 1779-1814.

[50] L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D, 60 (1992), 259-268.