Commun. Math. Sci.
Vol. 23, No. 2, pp. 329–358

ON VARIATIONAL IMAGE DENOISING: A TOTAL SCALED-GRADIENT VARIATION REGULARIZATION FRAMEWORK AND AN ACCELERATED OPERATOR-SPLITTING SOLVER[*]

CHUNLIN WU[†], HAIBIN SU[‡], AND DAOPING ZHANG[§]

Abstract. In image modeling, piecewise-linear structure is the simplest and important extension of piecewise-constant structure. In this paper, by sparse modeling of piecewise-linear structures, we propose a total scaled-gradient variation (TSGV) regularization framework (with a class of scaling functions) in the context of variational image denoising. A general high-order denoising model using this regularization framework and $\mathscr{L}^p$-based fidelity ($p\in[1,+\infty)$) is studied both theoretically and computationally. We establish two theorems showing its edge and corner contrast preservation under proper assumptions on the scaling function, by introducing some new settings and key inequality estimations. To solve our high-order model, we design an accelerated operator-splitting (AOS) solver by combining inertial extrapolation and some approximation techniques to the operator-splitting (OS) method based on Lie scheme and Marchuk-Yanenko discretization. To the best of our knowledge, this is the first work to combine acceleration techniques with such basic operator-splitting method used widely to solve nonconvex and nonsmooth high-order models. Our AOS solver is easy to implement. It not only inherits the advantages of Lie scheme type of operator-splitting methods, but also becomes more efficient. Numerical experiments demonstrate that our method usually generates denoising results with higher PSNR values by adjusting fewer parameters and using much less running time than the compared high-order methods.

Keywords. Image denoising; piecewise-linear structure; sparse modeling; high-order method; contrast preservation; accelerated operator-splitting.

AMS subject classifications. 68U10; 94A08; 65K99.

1. Introduction

Image denoising is a fundamental task in image processing and computer vision. Let $\Omega\subset\mathbb{R}^2$ be an open rectangular domain and $f:\Omega\to\mathbb{R}$ be a noisy observation of a clean image $\underline{u}:\Omega\to\mathbb{R}$. The aim of image denoising is to recover or estimate $\underline{u}$ from f, and preserve edges and contrasts as much as possible, for better later treatments like segmentation and recognition.

There are several different types of approaches for image denoising, such as linear and nonlinear filtering methods [25,49], linear and nonlinear diffusion methods [40,52], wavelet-based methods [8,20,51], non-local means [7], and variational methods [2,4,11,14,15,42]. Actually there are close connections between these different types of methods as noted and analyzed in [8,46]. In this paper, we focus on variational methods.

In the variational framework, image denoising problem can be formulated as the following form

$$\min_{u}\lambda\mathcal{R}(u)+\mathcal{F}(u;f),$$

where $\mathcal{F}(u;f)$ is a fidelity term to measure the discrepancy between u and f, $\mathcal{R}(u)$ is a regularizer to characterize the prior knowledge on u, and λ is a positive parameter

---

[*] Received: July 26, 2023; Accepted (in revised form): June 27, 2024. Communicated by Lexing Ying.

[†] School of Mathematical Sciences and Key Laboratory of Pure Mathematics and Combinatorics (LPMC), Nankai University, Tianjin 300071, People's Republic of China (wucl@nankai.edu.cn).

[‡] School of Mathematical Sciences, Nankai University, Tianjin 300071, People's Republic of China (1120200027@mail.nankai.edu.cn).

[§] Corresponding author. School of Mathematical Sciences and LPMC, Nankai University, Tianjin 300071, People's Republic of China (daopingzhang@nankai.edu.cn).

to balance these two terms. The fidelity term $\mathcal{F}(u;f)$ is usually derived by maximum likelihood (ML) and thus depends on the statistics of the measurement noise; see, e.g., [14, 43, 44] and references therein. In variational image denoising articles, this noise statistic is usually assumed to be given. The regularizer $\mathcal{R}(u)$ characterizes some prior information and has many choices in the literature. The most fundamental variational model is the total variation (TV) model proposed by Rudin, Osher, and Fatemi [42]:

$$\min_u \lambda \int_\Omega |\nabla u| \mathrm{d}x\mathrm{d}y + \frac{1}{2}\int_\Omega |u-f|^2 \mathrm{d}x\mathrm{d}y, \tag{1.1}$$

where $\nabla u = (u_x, u_y)^\top$ and $|\nabla u| = \sqrt{u_x^2 + u_y^2}$. While efficiently removing Gaussian noise, this TV model (1.1) can preserve image edges very well [47] and recover piecewise-constant images, leading to lots of applications and extensions; see chapters in [2, 15, 39]. However, sometimes TV regularizer has unfavorable properties, like the staircase effect [11, 18], the contrast reduction [16, 33], and the smearing of corners [3].

To overcome the drawbacks of the TV model (1.1), a number of high-order variational models were proposed from the approachs of calculus of variations or geometrical modeling. To the best of our knowledge, [11] by Chambolle and Lions is the first such study and gives an inf-convolution-based regularization framework with the following typical instance

$$\min_{u=u_1+u_2} \left\{ \int_\Omega |\nabla u_1| \mathrm{d}x\mathrm{d}y + \lambda_1 \int_\Omega ||\nabla^2 u_2||_{\mathrm{F}} \mathrm{d}x\mathrm{d}y + \frac{\lambda_2}{2}\int_\Omega |u_1+u_2-f|^2 \mathrm{d}x\mathrm{d}y \right\},$$

where $\nabla^2 u_2$ is the (distributional) Hessian of u_2, $||\cdot||_{\mathrm{F}}$ is the Frobenius norm of a 2×2 matrix, and λ_1, λ_2 are positive parameters. In [4], based on the tensor tool, the authors proposed the total generalized variation (TGV) framework with arbitrary order information, whose second-order instance is as follows [5]

$$\min_{u,\boldsymbol{w}} \left\{ \lambda_1 \int_\Omega |\nabla u - \boldsymbol{w}| \mathrm{d}x\mathrm{d}y + \lambda_2 \int_\Omega \left\| \frac{1}{2}(\nabla \boldsymbol{w} + \nabla \boldsymbol{w}^\top) \right\|_{\mathrm{F}} \mathrm{d}x\mathrm{d}y + \frac{1}{2}\int_\Omega |u-f|^2 \mathrm{d}x\mathrm{d}y \right\}.$$

A fourth-order PDE method was designed in [55], as the Euler-Lagrange equation of the following second-order minimization model

$$\min_u \left\{ \lambda \int_\Omega \phi(|\Delta u|) \mathrm{d}x\mathrm{d}y + \frac{1}{2}\int_\Omega |u-f|^2 \mathrm{d}x\mathrm{d}y \right\}, \tag{1.2}$$

where ϕ is an increasing function and Δ is the Laplacian operator. [29] proposed the following so-called LLT (Lysaker-Lundervold-Tai) denoising model with the Hessian regularizer:

$$\min_u \left\{ \lambda \int_\Omega ||\nabla^2 u||_{\mathrm{F}} \mathrm{d}x\mathrm{d}y + \frac{1}{2}\int_\Omega |u-f|^2 \mathrm{d}x\mathrm{d}y \right\}. \tag{1.3}$$

Both (1.2) and (1.3) use only second-order regularization to avoid the staircase effect. In [13] and [38], their authors proposed to combine the first-order and different second-order regularizers by the means of addition. These high-order models can effectively alleviate the staircase phenomenon of the TV model (1.1). Most of them are convex and can be well analyzed in proper function spaces [4, 11, 26, 38]. They were solved numerically by gradient descent methods [13, 29, 55], or various convex optimization algorithms [5, 12, 23, 38, 53]. Usually, the latter type of algorithms are much more efficient than gradient descent methods based on the Euler-Lagrange equations.

From the perspective of geometrical modeling, there are also several interesting high-order models. By using co-area formula, the TV model (1.1) was extended to the Euler's elastica (EE) denoising/smoothing model [17,48]

$$\min_u \int_\Omega \left(\lambda_1+\lambda_2\left(\nabla\cdot\frac{\nabla u}{|\nabla u|}\right)^2\right)|\nabla u|\mathrm{d}x\mathrm{d}y+\frac{1}{p}\int_\Omega |u-f|^p\mathrm{d}x\mathrm{d}y, \tag{1.4}$$

with the elastica regularization term from [31,32,36,45] and $p\geq 1$, which minimizes both the lengths and curvature integrals of all level curves of the image function $u(x,y)$. By viewing an image as a surface patch $\{(x,y,u(x,y))\in\mathbb{R}^3|(x,y)\in\Omega\}$, Zhu and Chan [57] proposed to use the mean curvature (MC) of the surface as the regularization, leading to the following MC denoising model

$$\min_u \lambda\int_\Omega \left|\nabla\cdot\frac{\nabla u}{\sqrt{1+|\nabla u|^2}}\right|\mathrm{d}x\mathrm{d}y+\frac{1}{2}\int_\Omega |u-f|^2\mathrm{d}x\mathrm{d}y. \tag{1.5}$$

For (1.5), the authors in [57] established two remarkable theorems showing the edge and corner contrast preservation properties. Following this surface geometry perspective [57], [6] and [56] considered the following Gaussian curvature (GC) regularization model

$$\min_u \lambda\int_\Omega \frac{|\det\nabla^2 u|}{(1+|\nabla u|^2)^2}\mathrm{d}x\mathrm{d}y+\frac{1}{2}\int_\Omega |u-f|^2\mathrm{d}x\mathrm{d}y, \tag{1.6}$$

and the Weingarten map (WM) based denoising model

$$\min_u \lambda\int_\Omega \left\|\nabla\left(\frac{\nabla u}{\sqrt{1+|\nabla u|^2}}\right)\right\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y+\frac{1}{2}\int_\Omega |u-f|^2\mathrm{d}x\mathrm{d}y. \tag{1.7}$$

It was shown in [6] and [56] that both the GC and WM models (1.6) (1.7) inherit the same edge and corner contrast preservation properties as the MC model (1.5). However, the proofs in [56] are not complete and seem to us hard to be completed in their proof settings. We mention that [56] also gives a variant of the WM model. These models are nonconvex and nonsmooth, and thus difficult to be solved. Apart from the earlier gradient descent method [45,57], augmented Lagrangian based methods [6,48,54,56,58], relaxed normal two split method [54], and some basic operator-splitting methods using Lie scheme and Marchuk-Yanenko discretization [17,22,28] have been developed. Generally speaking, the latter three types of methods are much more efficient; and for these nonconvex high-order models, basic operator-splitting methods are more stable and with less tuning parameters than the augmented Lagrangian based methods [22,24].

Here, we would like to present an intuitive and uniform understanding about the above variational models from the sparse modeling viewpoint, although they were designed from PDE approaches or geometrical modeling viewpoints. As can be seen, the regularizers in these models are mostly the L_1 norm of some quantities, like the gradient, Laplacian, Hessian, mean curvature, etc. After discretization, we interpret that these models are regularizing the sparsity of those quantities by ℓ_1 norm. According to the theory of compressive sensing [9,19], minimizing these models achieves some sparse recovery effect. From this viewpoint, the TV model (1.1) does tend to generate piecewise-constant images, since it pursuits results with sparse image gradient. Those high-order models, however, produce denoising results with sparse high-order information, like Laplacian, Hessian, curvatures, and others, indicating piecewise-smooth reconstructions with less staircase effects.

Due to the above discussion, we consider in this paper how to regularize piecewise-linear images from the perspective of sparse modeling, as piecewise-linear structure is the simplest and important extension of piecewise-constant structure and this is rarely noticed in existing discussions on the design of variational methods. Our main contributions are as follows

- Benefiting from the sparse modeling viewpoint that provides us more freedom in the design of expression, we propose a total scaled-gradient variation regularization framework (with a class of scaling functions) to construct a general high-order variational denoising model.
- We establish the edge and corner contrast preservation properties of the general model using our regularization framework and $\mathscr{L}^p$-based fidelity ($p\in[1,+\infty)$). A special case of our analysis (with a specific scaling function and squared $\mathscr{L}^2$ fidelity) completes the proofs in a related literature under our settings.
- We provide an accelerated operator-splitting method to solve the proposed model. To the best of our knowledge, this is the first to introduce the inertial extrapolation technique to the operator-splitting method based on Lie scheme and Marchuk-Yanenko discretization in solving restoration models. Numerical experiments demonstrate the advantages of our method in terms of PSNR values, the number of tuning parameters and especially computational speed.

This paper is organized as follows. In Section 2, we propose a total scaled-gradient variation regularization framework to construct a general variational image denoising model, from the sparse modeling viewpoint. In Section 3, we establish its edge and corner contrast preservation properties. In Section 4, an accelerated operator-splitting method with inertial extrapolation technique is devised. The advantages of the proposed method compared to several latest variational approaches are illustrated numerically in Section 5. Section 6 concludes this paper.

2. A total scaled-gradient variation regularization framework and minimization model: from piecewise-constant regularization to piecewise-linear regularization

People often consider clean images in the continuous setting as piecewise-smooth functions [35]. The most simple structure of piecewise-smooth functions is clearly the piecewise-constant structure. The TV model (1.1) characterizes this structure very well and indeed generates piecewise-constant reconstructions, showing good edge preservation property. However, most clean natural images are more complicated than the piecewise-constant structure. *As the simplest extension of the piecewise-constant structure towards the piecewise-smooth structure, the piecewise-linear structure is our central interest to model in this paper*. We start from an intuitive observation by the sparse modeling viewpoint to present a regularization framework for a such structure.

Recall that $f:\Omega\to\mathbb{R}$ is a noisy observation of a clean image $\underline{u}:\Omega\to\mathbb{R}$. We assume the ground truth $\underline{u}(x,y)$ to be with a piecewise-linear structure. It follows that the distributional gradient $\nabla\underline{u}=(\underline{u}_x,\underline{u}_y)^\top$ is piecewise-constant. For any given non-negative and bounded above function $\psi\in C^1([0,\infty))$, we have that $\psi(|\nabla\underline{u}|)\nabla\underline{u}$ is also piecewise-constant. Therefore, by the distributional Jacobian operator ∇ on $\psi(|\nabla\underline{u}|)\nabla\underline{u}$, $\nabla(\psi(|\nabla\underline{u}|)\nabla\underline{u})$ should be sparse in the image domain Ω, and thus the following quantity

$$\int_\Omega \|\nabla(\psi(|\nabla u|)\nabla u)\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y$$

is a natural choice for regularization. Here the ψ function acts as a scaling function and the regularization can be regarded as a total scaled-gradient variation (TSGV). To recover or estimate $\underline{u}$, we hence propose to consider the following general model:

$$\inf_u \left\{ E(u) := \lambda \int_\Omega \|\mathrm{H}_\psi(u)\|_{\mathrm{F}} \mathrm{d}x\mathrm{d}y + \frac{1}{p}\int_\Omega |u-f|^p \mathrm{d}x\mathrm{d}y \right\}, \tag{2.1}$$

with $\mathrm{H}_\psi(u) = \nabla(\psi(|\nabla u|)\nabla u)$ and $p \in [1,+\infty)$, where λ is a positive tuning parameter to balance the fidelity term and the regularization. The fidelity term in (2.1) considers a general data loss for additive measurement noise, such as Gaussian noise ($p=2$) and impulsive noise ($p=1$). The regularization term, in the discrete setting, is essentially the ℓ^1-norm of $\|\mathrm{H}_\psi\|_{\mathrm{F}}$. Note that if f is an affine image, the minimizer of (2.1) is f itself.

In the model (2.1), the Jacobian operator is used. Actually, one can also consider the divergence operator on $\psi(|\nabla u|)\nabla u$, which leads to the following minimization problem:

$$\inf_u \left\{ \lambda \int_\Omega |\nabla\cdot(\psi(|\nabla u|)\nabla u)| \mathrm{d}x\mathrm{d}y + \frac{1}{p}\int_\Omega |u-f|^p \mathrm{d}x\mathrm{d}y \right\}. \tag{2.2}$$

For piecewise-linear images, the regularization term in (2.2) is also a kind of sparsity measure.

Remark 2.1. Here, we start from the sparse modeling viewpoint to present (2.1) and (2.2), which allow to use a general ψ function. Indeed, by choosing some specific ψ functions and p value, (2.1) and (2.2) reduce to four existing high-order denoising models, which were designed by PDE approaches or surface geometry perspective:

- the LLT model (1.3), i.e. (2.1) with $\psi(s)\equiv 1, p=2$;
- the Laplacian regularization model (1.2) with $\phi(s)=s$, i.e. (2.2) with $\psi(s)\equiv 1, p=2$;
- the WM minimization model (1.7), i.e. (2.1) with $\psi(s)=\frac{1}{\sqrt{1+s^2}}, p=2$;
- the MC model (1.5), i.e. (2.2) with $\psi(s)=\frac{1}{\sqrt{1+s^2}}, p=2$.

The above four instance models can be interpreted consistently by the sparse modeling viewpoint. We also note that it is hard for a derivation from the surface geometry viewpoint [56, 57] to give a regularization term with a ψ function rather than $\psi(s)=\frac{1}{\sqrt{1+s^2}}$.

Remark 2.2. We would like to compare (2.1) with (2.2) briefly. We see that the main difference between them is that (2.2) misses some mixed partial derivatives on $\psi(|\nabla u|)\nabla u$ in the regularization term. Of course, in 1-dimensional signal case, (2.1) and (2.2) coincide. In higher-dimensional cases, the Jacobian operator in (2.1) keeps more information than the divergence operator in (2.2). The regularization in (2.1) is a total scaled-gradient variation, while the regularization in (2.2) can be regarded as a certain "partial" scaled-gradient variation. Existing studies on special cases of (2.1) and (2.2) [29, 55–57] show numerically that Jacobian-based models [29, 56] perform better than divergence-based models [55, 57]. Therefore, in the remainder of this paper, we will mainly study the model (2.1).

3. Edge and corner contrast preservation properties of the model (2.1)

The TV model (1.1) can preserve image edges, but suffers from the contrast reduction disadvantage [16, 33]. In [57], the authors proved two remarkable analytical results, which show the perfect edge and corner contrast preservation properties of the MC denoising model ((2.2) with $\psi(s)=\frac{1}{\sqrt{1+s^2}}$ and $p=2$). When extending the MC model to the GC model (1.6) and the WM minimization model ((2.1) with $\psi(s)=\frac{1}{\sqrt{1+s^2}}$ and $p=2$), [6] and [56] also proved two similar theoretical results.

In this section, we will establish similar edge and corner contrast preservation results for model (2.1) with a class of ψ and $p\in[1,+\infty)$. We will give our proof details (with some new and key inequality estimations), which generate existing theoretical results in some sense. We assume ψ to satisfy the following

- Assumption A1: $\psi:[0,+\infty)\to(0,+\infty)$ is a monotonically decreasing C^1 function;
- Assumption A2: $\lim_{s\to+\infty}\psi(s)s=1$ and $(\psi(s)s)'>0$ for $s\in[0,+\infty)$;
- Assumption A3: $\lim_{s\to+\infty}s^2(\psi'(s)s+\psi(s))$ and $\lim_{s\to+\infty}\psi'(s)s^2$ exist.

Here, $\psi'(0)$ is understood as the right derivative of ψ at 0. We mention that a class of functions $\psi(s)=\frac{1}{(b+s^k)^{1/k}}$ with $b>0$ and $k\geq1$ do satisfy the above three assumptions, becasue

1. (1) A1 holds obviously.
2. (2) A2 is true due to $\lim\limits_{s\to+\infty}\frac{s}{(b+s^k)^{1/k}}=1$ and $(\psi(s)s)'=\frac{b}{(b+s^k)^{1+1/k}}>0$ for $s\in[0,+\infty)$.
3. (3) A3 holds since $\lim\limits_{s\to+\infty}s^2(\psi'(s)s+\psi(s))=\lim\limits_{s\to+\infty}\frac{bs^2}{(b+s^k)^{1+1/k}}=b$ if $k=1$ or 0 if $k>1$ and $\lim\limits_{s\to+\infty}\psi'(s)s^2=-\lim\limits_{s\to+\infty}\frac{s^{k+1}}{(b+s^k)^{1+1/k}}=-1$.

Similarly, $\hat{\psi}(s)=\frac{1}{\mathrm{e}^{-s}+s}$ also satisfies the above three assumptions.

Remark 3.1. We can note that when ψ satisfies A1-A3, $\psi(s)s$, $s^2(\psi'(s)s+\psi(s))$ and $\psi'(s)s^2$ are all bounded in $[0,+\infty)$.

We now begin to show the properties of edge preservation and preservation of contrast at edge locations of (2.1). Similar with [6, 56, 57], we consider a specific but important image

$$f_1(x,y)=h\chi_{B(0,R)}(x,y),\quad (x,y)\in\Omega_1=(-2R,2R)\times(-2R,2R), \tag{3.1}$$

where χ is the characteristic function, $B(0,R)$ is an open disk in $\mathbb{R}^2$ centered at the origin with radius R, and $h>0$ is the height. To calculate $\int_{\Omega_1}\|\mathrm{H}_\psi(f_1)\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y$, we also use the natural idea to define $\int_{\Omega_1}\|\mathrm{H}_\psi(f_1)\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y$ as $\lim\limits_{n\to\infty}\int_{\Omega_1}\|\mathrm{H}_\psi(g_n)\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y$, where $\{g_n\}$ is any sequence of smooth functions that approximate f_1 and $\int_{\Omega_1}\|\mathrm{H}_\psi(g_n)\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y$ can be calculated easily (the independence of $\int_{\Omega_1}\|\mathrm{H}_\psi(f_1)\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y$ from the choice of $\{g_n\}$ will be given in the later proof). Since f_1 is radially symmetric, we generate a sequence of smooth radially symmetric functions to approximate it, by rotating a sequence of smooth univariate functions taken from the following set:

$$\begin{aligned}\mathscr{S}=\Big\{\varrho\in C^2([0,\infty))\Big|\,&\varrho'(s)\leq0;\varrho'(s)=\varrho'(2R-s)\text{ if }s\in[0,2R];\varrho'(R)<-\frac{2h}{R};\\&\varrho''(s)\leq0\text{ if }s\in(0,R),\varrho''(s)\geq0\text{ if }s\in(R,+\infty);\\&\text{there exist }R_1,R_2,0<R_1<R<R_2<2R,\\&\text{such that }\varrho(s)=h\text{ if }s\in[0,R_1]\text{ and }\varrho(s)=0\text{ if }s\in[R_2,+\infty)\Big\},\end{aligned}$$

where $\varrho'(0)=0$ is understood as the right derivative. We also note $\varrho(R)=\frac{h}{2}$ by $\varrho'(s)=\varrho'(2R-s)$ if $s\in[0,2R]$, $\varrho(0)=h$ and $\varrho(2R)=0$. This strategy yields the following lemma.

Lemma 3.1. *Let f_1 be an image given in (3.1). Let ψ satisfy Assumption A1-A2. Then we have*

$$\int_{\Omega_1}\|\mathrm{H}_\psi(f_1)\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y=4\pi R.$$

Proof. For any given smooth function $u(x,y)$, we have $\|\mathrm{H}_\psi(u)\|_{\mathrm{F}}$ as

$$\begin{cases} \left\| \psi(|\nabla u|) \begin{pmatrix} u_{xx} & u_{xy} \\ u_{yx} & u_{yy} \end{pmatrix} \right\|_{\mathrm{F}}, \text{ if } \nabla u(x,y)=0; \\ \left\| \dfrac{\psi'(|\nabla u|)}{|\nabla u|} \begin{pmatrix} u_x^2 u_{xx} + u_x u_y u_{yx} & u_x^2 u_{xy} + u_x u_y u_{yy} \\ u_x u_y u_{xx} + u_y^2 u_{yx} & u_x u_y u_{xy} + u_y^2 u_{yy} \end{pmatrix} + \psi(|\nabla u|) \begin{pmatrix} u_{xx} & u_{xy} \\ u_{yx} & u_{yy} \end{pmatrix} \right\|_{\mathrm{F}}, \text{otherwise.} \end{cases} \tag{3.2}$$

Following the strategy mentioned above, we consider choosing a sequence of functions in $\mathscr{S}$ to approach $h\chi_{[0,R]}(s), s\in[0,+\infty)$, and rotate them to get a sequence of smooth radially symmetric functions that approximate the function f_1 over Ω_1. If $\varrho\in\mathscr{S}$, then for a radially symmetric surface represented as $g(x,y)=\varrho(r)$ with $r=\sqrt{x^2+y^2}$, we have

$$g_x=\varrho'\frac{x}{r},\ g_y=\varrho'\frac{y}{r},\ g_{xx}=\varrho''\frac{x^2}{r^2}+\varrho'\frac{y^2}{r^3},\ g_{xy}=g_{yx}=\varrho''\frac{xy}{r^2}-\varrho'\frac{xy}{r^3},\ g_{yy}=\varrho''\frac{y^2}{r^2}+\varrho'\frac{x^2}{r^3}.$$

According to (3.2), we have

$$\begin{aligned} \|\mathrm{H}_\psi(g)\|_{\mathrm{F}} &= \left\| \frac{\psi'(|\varrho'|)}{|\varrho'|} \begin{pmatrix} (\varrho')^2\varrho''\frac{x^2}{r^2} & (\varrho')^2\varrho''\frac{xy}{r^2} \\ (\varrho')^2\varrho''\frac{xy}{r^2} & (\varrho')^2\varrho''\frac{y^2}{r^2} \end{pmatrix} + \psi(|\varrho'|) \begin{pmatrix} \varrho''\frac{x^2}{r^2}+\varrho'\frac{y^2}{r^3} & \varrho''\frac{xy}{r^2}-\varrho'\frac{xy}{r^3} \\ \varrho''\frac{xy}{r^2}-\varrho'\frac{xy}{r^3} & \varrho''\frac{y^2}{r^2}+\varrho'\frac{x^2}{r^3} \end{pmatrix} \right\|_{\mathrm{F}} \\ &= \sqrt{\Big[(\psi(|\varrho'|)\varrho')'\Big]^2 + \left[\frac{\psi(|\varrho'|)\varrho'}{r}\right]^2},\ \text{if } \varrho'(r)\neq 0, \end{aligned}$$

and

$$\begin{aligned} \|\mathrm{H}_\psi(g)\|_{\mathrm{F}} &= \left\| \psi(|\varrho'|) \begin{pmatrix} \varrho''\frac{x^2}{r^2}+\varrho'\frac{y^2}{r^3} & \varrho''\frac{xy}{r^2}-\varrho'\frac{xy}{r^3} \\ \varrho''\frac{xy}{r^2}-\varrho'\frac{xy}{r^3} & \varrho''\frac{y^2}{r^2}+\varrho'\frac{x^2}{r^3} \end{pmatrix} \right\|_{\mathrm{F}} \\ &= \sqrt{\Big[(\psi(|\varrho'|)\varrho')'\Big]^2 + \left[\frac{\psi(|\varrho'|)\varrho'}{r}\right]^2},\ \text{if } \varrho'(r)=0, \end{aligned}$$

where $\varrho'=\varrho'(r)$. For convenience, we denote $\vartheta=\vartheta(r)=\psi(|\varrho'(r)|)\varrho'(r)$ and obtain

$$\int_{\Omega_1}\|\mathrm{H}_\psi(g)\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y = 2\pi\int_0^{2R} r\sqrt{(\vartheta')^2+\left(\frac{\vartheta}{r}\right)^2}\,\mathrm{d}r. \tag{3.3}$$

Since $\vartheta'=\big(\psi'(|\varrho'|)|\varrho'|+\psi(|\varrho'|)\big)\varrho''$, by Assumption A1-A2, we see $\vartheta'\le 0$ and $\vartheta/r\le 0$ for $r\in(0,R]$. One thus gets

$$-\int_0^R r\vartheta'\mathrm{d}r \le \int_0^R r\sqrt{(\vartheta')^2+\left(\frac{\vartheta}{r}\right)^2}\,\mathrm{d}r \le -\int_0^R r\left(\vartheta'+\frac{\vartheta}{r}\right)\mathrm{d}r, \tag{3.4}$$

with the lower bound as

$$-\int_0^R r\vartheta'\mathrm{d}r = -R\psi(|\varrho'(R)|)\varrho'(R) + \int_0^R \psi(|\varrho'|)\varrho'\mathrm{d}r,$$

and the upper bound as

$$-\int_0^R r\left(\vartheta'+\frac{\vartheta}{r}\right)\mathrm{d}r = -\int_0^R r(\psi(|\varrho'|)\varrho')'\mathrm{d}r - \int_0^R \psi(|\varrho'|)\varrho'\mathrm{d}r$$

$$= -R\psi(|\varrho'(R)|)\varrho'(R).$$

For $r \in [R, 2R)$, we have $\vartheta' \geq 0$ and $\vartheta/r \leq 0$, and thus

$$\int_R^{2R} r\vartheta' \mathrm{d}r \leq \int_R^{2R} r\sqrt{(\vartheta')^2 + \left(\frac{\vartheta}{r}\right)^2}\,\mathrm{d}r \leq \int_R^{2R} r\left(\vartheta' - \frac{\vartheta}{r}\right)\mathrm{d}r, \tag{3.5}$$

with the lower bound as

$$\int_R^{2R} r\vartheta' \mathrm{d}r = -R\psi(|\varrho'(R)|)\varrho'(R) - \int_R^{2R} \psi(|\varrho'|)\varrho' \mathrm{d}r,$$

and the upper bound as

$$\begin{aligned}\int_R^{2R} r\left(\vartheta' - \frac{\vartheta}{r}\right)\mathrm{d}r &= \int_R^{2R} r(\psi(|\varrho'|)\varrho')'\mathrm{d}r - \int_R^{2R} \psi(|\varrho'|)\varrho' \mathrm{d}r \\ &= -R\psi(|\varrho'(R)|)\varrho'(R) - 2\int_R^{2R} \psi(|\varrho'|)\varrho' \mathrm{d}r,\end{aligned}$$

by $\varrho'(2R) = 0$. Since $\varrho'(s) = \varrho'(2R - s)$ for $s \in [0, 2R]$, we have $\int_0^R \psi(|\varrho'|)\varrho' \mathrm{d}r = \int_R^{2R} \psi(|\varrho'|)\varrho' \mathrm{d}r$. Therefore, adding (3.4) and (3.5) gives the lower and upper bounds of (3.3) as follows

$$\begin{aligned}-4\pi R\psi(|\varrho'(R)|)\varrho'(R) &\leq \int_{\Omega_1} \|\mathrm{H}_\psi(g)\|_{\mathrm{F}} \mathrm{d}x\mathrm{d}y \\ &\leq -4\pi R\psi(|\varrho'(R)|)\varrho'(R) - 4\pi \int_R^{2R} \psi(|\varrho'|)\varrho' \mathrm{d}r.\end{aligned} \tag{3.6}$$

Let $\{\varrho_n\} \subset \mathscr{S}$ be any sequence of functions such that $g_n(x,y) = \varrho_n(\sqrt{x^2+y^2})$ approaches f_1 pointwisely except $\{(x,y)|x^2+y^2 = R^2\}$ as $n \to +\infty$. Hence, we get $\varrho_n'(R) \to -\infty$ and $\varrho_n'(r) \to 0$ for $r \neq R$. By Assumption A2, we have $\lim_{n\to\infty} \psi(|\varrho_n'(R)|)(-\varrho_n'(R)) = 1$. Moreover, by using the dominated convergence theorem, we get $\lim_{n\to\infty} \int_R^{2R} \psi(|\varrho_n'|)(-\varrho_n')\mathrm{d}r = 0$. We therefore obtain $\int_{\Omega_1} \|\mathrm{H}_\psi(f_1)\|_{\mathrm{F}} \mathrm{d}x\mathrm{d}y = \lim_{n\to\infty} \int_{\Omega_1} \|\mathrm{H}_\psi(g_n)\|_{\mathrm{F}} \mathrm{d}x\mathrm{d}y = 4\pi R$. ◻

Lemma 3.1 indicates that the regularizer in model (2.1) with ψ satisfying A1-A2 is independent of the intensity level h for images like f_1. Therefore the regularization in (2.1) does not penalize contrasts of such images. The next lemma gives a lower bound of the fidelity term of (2.1) with a general $p \in [1, +\infty)$ for f_1.

Lemma 3.2. *Let f_1 be an image given in (3.1) and $p \in [1, +\infty)$. Then we have*

$$\frac{1}{p}\int_{\Omega_1} |g - f_1|^p \mathrm{d}x\mathrm{d}y \geq -\frac{2\pi R h^{p+1}}{2^p p(p+1)\varrho'(R)}, \tag{3.7}$$

where $g(x,y) = \varrho(\sqrt{x^2+y^2})$ for $\varrho \in \mathscr{S}$.

Proof. Since both f_1 and g are radially symmetric, we analyze ϱ to estimate the integral. We see that the tangent line of ϱ at the point $(R, \varrho(R))$ is $T(s) = \varrho(R) +$

$\varrho'(R)(s-R)$. Suppose that $T(s)$ intersects with $z=h$ and $z=0$ at (R_{in},h) and $(R_{\text{out}},0)$, respectively. We thus derive

$$\begin{aligned}\frac{1}{p}\int_{\Omega_1}|g-f_1|^p\mathrm{d}x\mathrm{d}y &= \frac{1}{p}\int_{\Omega_1} r|h\chi_{[0,R]}(r)-\varrho(r)|^p\mathrm{d}r\mathrm{d}\theta\\ &\geq \frac{2\pi}{p}\int_{R_{\text{in}}}^{R_{\text{out}}} r|h\chi_{[0,R]}(r)-\varrho(r)|^p\mathrm{d}r\\ &\geq \frac{2\pi}{p}\int_{R_{\text{in}}}^{R} r(h-T(r))^p\mathrm{d}r+\frac{2\pi}{p}\int_{R}^{R_{\text{out}}} rT(r)^p\mathrm{d}r\\ &= \frac{2\pi}{p}\Big[-\frac{R}{(p+1)\varrho'(R)}(\varrho(R)^{p+1}+(h-\varrho(R))^{p+1})\\ &\qquad +\frac{1}{(p+1)(p+2)(\varrho'(R))^2}(\varrho(R)^{p+2}-(h-\varrho(R))^{p+2})\Big],\end{aligned}$$

where the second inequality is due to the fact that ϱ is concave in $[R_{\text{in}},R]$ and convex in $[R,R_{\text{out}}]$, and the last equality is obtained by applying the integration by parts. Furthermore, by $\varrho(R)=\frac{h}{2}$, we complete the proof. ∎

Combining Lemma 3.1 and 3.2, we establish the following theorem.

THEOREM 3.1. *Consider the model (2.1) with $f=f_1$ given in (3.1). Let ψ satisfy Assumption A1-A3 and $p\in[1,+\infty)$. Then there exists a positive constant*

$$c_1=\frac{h^{p+1}}{2^{p+1}p(p+1)}\frac{1}{\sup\limits_{s\in[2h/R,+\infty)}\{s^2(\psi'(s)s+\psi(s))\}},$$

such that $E(f_1)=\inf_{\varrho\in\mathscr{S}}E(\varrho(r))$ with $r=\sqrt{x^2+y^2}$, when $\lambda<c_1$.

Proof. Take a $\varrho\in\mathscr{S}$ and let $g(x,y)=\varrho(\sqrt{x^2+y^2})$. Combining the inequality (3.6) with (3.7), we obtain the following estimate:

$$E(g)\geq -4\pi\lambda R\psi(|\varrho'(R)|)\varrho'(R)-\frac{2\pi Rh^{p+1}}{2^p p(p+1)\varrho'(R)}.$$

We therefore consider a function $\eta:[2h/R,+\infty)\to\mathbb{R}$ as follows

$$\eta(s)=\lambda 4\pi R\psi(s)s+\frac{2\pi Rh^{p+1}}{2^p p(p+1)}\frac{1}{s},$$

whose derivative is

$$\eta'(s)=\lambda 4\pi R(\psi'(s)s+\psi(s))-\frac{2\pi Rh^{p+1}}{2^p p(p+1)}\frac{1}{s^2}.$$

If $\lambda<c_1:=\frac{h^{p+1}}{2^{p+1}p(p+1)}\inf\limits_{s\in[2h/R,+\infty)}\frac{1}{s^2(\psi'(s)s+\psi(s))}$, then $\eta'(s)<0\ \forall s\in[2h/R,+\infty)$, and thus $\eta(s)$ decreases over $[2h/R,+\infty)$.

Let $s=-\varrho'(R)$. We derive that $\eta(s)$ decreases to $E(f_1)$ when $\varrho'(R)\to-\infty$. Therefore, $E(g)>E(f_1)$ for smooth g generated by any $\varrho\in\mathscr{S}$. Besides, by using the dominated convergence theorem, we know that $E(g)\to E(f_1)$ when g approaches f_1 pointwisely except $\{(x,y)|x^2+y^2=R^2\}$ through $\varrho'(R)\to-\infty$. These mean $E(f_1)=\inf_{\varrho\in\mathscr{S}}E(\varrho(r))$, with $r=\sqrt{x^2+y^2}$, when $\lambda<c_1$. ∎

According to Theorem 3.1, when the regularization parameter λ is less than some constant, solving (2.1) with $f=f_1$ given in (3.1) over the functions determined by $\mathscr{S}$ is f_1 itself. This indicates that the general model (2.1) has edge contrast preservation property in some sense.

We next show that the model (2.1) is able to preserve object corners and contrasts. Again similar with [6, 56, 57], we use the following specific image f_2:

$$f_2(x,y)=h\chi_{(0,R)\times(0,R)}(x,y), \quad (x,y)\in\Omega_2=(-R,R)\times(-R,R). \tag{3.8}$$

We consider two function sets $\mathscr{P}$ and $\mathscr{Q}$ defined below

$$\begin{aligned}\mathscr{P}=\big\{\rho\in C^2(\mathbb{R})\big|\rho(s)=0 \text{ if } s<-1, \rho(s)=1 \text{ if } s>1;\\ 1\le\rho'(0)\le 2;\rho'(s)\ge 0;\rho'(s)=\rho'(-s);\\ \rho \text{ is convex in } (-1,0) \text{ and concave in } (0,1)\big\}\end{aligned}$$

and

$$\mathscr{Q}=\left\{z_{\rho,\mu}\middle|\rho\in\mathscr{P},\mu\in\left(0,\frac{R}{2}\right)\right\},$$

where $z_{\rho,\mu}$ is given by

$$z_{\rho,\mu}(x,y)=\begin{cases} h\rho(\frac{2y}{\mu}), & (x,y)\in[\mu,R)\times(-R,R),\\ h\rho(\frac{2x}{\mu}), & (x,y)\in(-R,\mu)\times[\mu,R),\\ h\rho(2-\frac{2r_\mu}{\mu}), & (x,y)\in(-R,\mu)\times(-R,\mu),\end{cases} \tag{3.9}$$

with $r_\mu=\sqrt{(x-\mu)^2+(y-\mu)^2}$. We also note $\rho(0)=\frac{1}{2}$ by $\rho'(s)=\rho'(-s)$, $\rho(-1)=0$ and $\rho(1)=1$. Let $g\in\mathscr{Q}$. Then there exist ρ and μ such that $g(x,y)=z_{\rho,\mu}(x,y)$. For each sequence of smooth functions $\{g_n\}$ approximating f_2 almost everywhere, we can calculate $\lim_{n\to\infty}\int_{\Omega_2}\|\mathrm{H}_\psi(g_n)\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y$ and define $\int_{\Omega_2}\|\mathrm{H}_\psi(f_2)\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y$ to be $\lim_{n\to\infty}\int_{\Omega_2}\|\mathrm{H}_\psi(g_n)\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y$, as the independence of $\lim_{n\to\infty}\int_{\Omega_2}\|\mathrm{H}_\psi(g_n)\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y$ from $\{g_n\}$ can be shown in the following lemma.

Lemma 3.3. *Let f_2 be an image given in (3.8). Let ψ satisfy Assumption A1-A2. Then we have*

$$\int_{\Omega_2}\|\mathrm{H}_\psi(f_2)\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y=4R.$$

Proof. For $g\in\mathscr{Q}$, we have

$$\|\mathrm{H}_\psi(g)\|_{\mathrm{F}}=\begin{cases}\left|\left(\frac{2h}{\mu}\rho'(\frac{2y}{\mu})\psi\left(\frac{2h}{\mu}\rho'(\frac{2y}{\mu})\right)\right)_y\right|, & (x,y)\in[\mu,R)\times(-R,R);\\ \left|\left(\frac{2h}{\mu}\rho'(\frac{2x}{\mu})\psi\left(\frac{2h}{\mu}\rho'(\frac{2x}{\mu})\right)\right)_x\right|, & (x,y)\in(-R,\mu)\times[\mu,R);\\ \sqrt{\left(\upsilon'\right)^2+\left(\frac{\upsilon}{r_\mu}\right)^2}, & (x,y)\in(-R,\mu)\times(-R,\mu),\end{cases}$$

where $\upsilon=\upsilon(r_\mu)=-\frac{2h}{\mu}\rho'(2-\frac{2r_\mu}{\mu})\psi\left(\frac{2h}{\mu}\rho'(2-\frac{2r_\mu}{\mu})\right)$.

We next estimate three integrals. By Assumption A2 and the property of ρ, we have

$$\int_{[\mu,R)\times(-R,R)}\|\mathrm{H}_\psi(g)\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y=\int_\mu^R\left[\int_{-R}^0\left(\frac{2h}{\mu}\rho'\left(\frac{2y}{\mu}\right)\psi\left(\frac{2h}{\mu}\rho'\left(\frac{2y}{\mu}\right)\right)\right)_y\mathrm{d}y\right.$$

$$
\begin{aligned}
&-\int_0^R\left(\frac{2h}{\mu}\rho'\left(\frac{2y}{\mu}\right)\psi\left(\frac{2h}{\mu}\rho'\left(\frac{2y}{\mu}\right)\right)\right)_y \mathrm{d}y\Big]\mathrm{d}x\\
&=2(R-\mu)\frac{2h}{\mu}\rho'(0)\psi\Big(\frac{2h}{\mu}\rho'(0)\Big),
\end{aligned}
\tag{3.10}
$$

and

$$
\begin{aligned}
\int_{(-R,\mu)\times[\mu,R)}\|\mathrm{H}_\psi(g)\|_\mathrm{F}\mathrm{d}x\mathrm{d}y=&\int_\mu^R\Big[\int_{-R}^0\left(\frac{2h}{\mu}\rho'\left(\frac{2x}{\mu}\right)\psi\left(\frac{2h}{\mu}\rho'\left(\frac{2x}{\mu}\right)\right)\right)_x\mathrm{d}x\\
&-\int_0^\mu\left(\frac{2h}{\mu}\rho'\left(\frac{2x}{\mu}\right)\psi\left(\frac{2h}{\mu}\rho'\left(\frac{2x}{\mu}\right)\right)\right)_x\mathrm{d}x\Big]\mathrm{d}y\\
=&2(R-\mu)\frac{2h}{\mu}\rho'(0)\psi\Big(\frac{2h}{\mu}\rho'(0)\Big).
\end{aligned}
\tag{3.11}
$$

When $(x,y)\in(-R,\mu)\times(-R,\mu)$, one has

$$
\int_{(-R,\mu)\times(-R,\mu)}\|\mathrm{H}_\psi(g)\|_\mathrm{F}\mathrm{d}x\mathrm{d}y=\frac{\pi}{2}\int_0^{\mu+R}r_\mu\sqrt{(\upsilon')^2+\left(\frac{\upsilon}{r_\mu}\right)^2}\mathrm{d}r_\mu.
$$

Again by Assumption A1-A2 and the property of ρ, we have $\upsilon'\le 0,\upsilon/r_\mu\le 0$ for $r_\mu\in(0,\mu)$ and $\upsilon'\ge 0,\upsilon/r_\mu\le 0$ for $r_\mu\in(\mu,\mu+R)$. Similar to the inequalities (3.4) and (3.5), we have

$$
-\int_0^\mu r_\mu\upsilon'\mathrm{d}r_\mu\le\int_0^\mu r_\mu\sqrt{(\upsilon')^2+\left(\frac{\upsilon}{r_\mu}\right)^2}\mathrm{d}r_\mu\le-\int_0^\mu r_\mu\left(\upsilon'+\frac{\upsilon}{r_\mu}\right)\mathrm{d}r_\mu
$$

and

$$
\int_\mu^{\mu+R} r_\mu\upsilon'\mathrm{d}r_\mu\le\int_\mu^{\mu+R} r_\mu\sqrt{(\upsilon')^2+\left(\frac{\upsilon}{r_\mu}\right)^2}\mathrm{d}r_\mu\le\int_\mu^{\mu+R} r_\mu\left(\upsilon'-\frac{\upsilon}{r_\mu}\right)\mathrm{d}r_\mu.
$$

By adding the above two inequalities and calculating four integrals without square roots in the integrands, we get the following lower and upper bound estimates:

$$
\begin{aligned}
&\frac{2h}{\mu}\rho'(0)\psi\Big(\frac{2h}{\mu}\rho'(0)\Big)\pi\mu+\frac{\pi}{2}\left(\int_0^\mu\upsilon\mathrm{d}r_\mu-\int_\mu^{\mu+R}\upsilon\mathrm{d}r_\mu\right)\\
\le&\int_{(-R,\mu)\times(-R,\mu)}\|\mathrm{H}_\psi(g)\|_\mathrm{F}\mathrm{d}x\mathrm{d}y\\
\le&\frac{2h}{\mu}\rho'(0)\psi\Big(\frac{2h}{\mu}\rho'(0)\Big)\pi\mu-\pi\int_\mu^{\mu+R}\upsilon\mathrm{d}r_\mu.
\end{aligned}
\tag{3.12}
$$

Combining (3.10), (3.11) and (3.12) yields

$$
\begin{aligned}
&(4(R-\mu)+\pi\mu)\frac{2h}{\mu}\rho'(0)\psi\Big(\frac{2h}{\mu}\rho'(0)\Big)+\frac{\pi}{2}\left(\int_0^\mu\upsilon\mathrm{d}r_\mu-\int_\mu^{\mu+R}\upsilon\mathrm{d}r_\mu\right)\\
\le&\int_{\Omega_2}\|\mathrm{H}_\psi(g)\|_\mathrm{F}\mathrm{d}x\mathrm{d}y\\
\le&(4(R-\mu)+\pi\mu)\frac{2h}{\mu}\rho'(0)\psi\Big(\frac{2h}{\mu}\rho'(0)\Big)-\pi\int_\mu^{\mu+R}\upsilon\mathrm{d}r_\mu.
\end{aligned}
$$

By noting $\int_{2\mu}^{\mu+R} \upsilon \mathrm{d}r_\mu = 0$, and $\int_0^\mu \upsilon \mathrm{d}r_\mu = \int_\mu^{2\mu} \upsilon \mathrm{d}r_\mu$ (due to $\rho'(s) = \rho'(-s)$), we further estimate

$$\begin{aligned}
&(4(R-\mu)+\pi\mu)\frac{2h}{\mu}\rho'(0)\psi\Big(\frac{2h}{\mu}\rho'(0)\Big)\\
\le& \int_{\Omega_2} \|\mathrm{H}_\psi(g)\|_{\mathrm{F}} \mathrm{d}x\mathrm{d}y\\
\le& (4(R-\mu)+\pi\mu)\frac{2h}{\mu}\rho'(0)\psi\Big(\frac{2h}{\mu}\rho'(0)\Big) - \pi\int_\mu^{2\mu} \upsilon \mathrm{d}r_\mu. \qquad (3.13)
\end{aligned}$$

Let $\{g_n(x,y) = z_{\rho_n,\mu_n}(x,y)\} \subset \mathscr{Q}$ be any sequence of functions that approximate f_2 with $n \to +\infty$ and $\mu_n \to 0^+$. We therefore derive $\int_{\Omega_2} \|\mathrm{H}_\psi(f_2)\|_{\mathrm{F}} \mathrm{d}x\mathrm{d}y = \lim_{n\to+\infty} \int_{\Omega_2} \|\mathrm{H}_\psi(g_n)\|_{\mathrm{F}} \mathrm{d}x\mathrm{d}y = 4R$, by noting $|\int_{\mu_n}^{2\mu_n} \upsilon_n \mathrm{d}r_{\mu_n}| \le \mu_n \sup_{s\in[0,+\infty)}\{\psi(s)s\} \to 0$ due to Assumption A1-A2. ☐

Lemma 3.4. *Let f_2 be an image given in (3.8) and $p \in [1,+\infty)$. Then for $\forall\ g = z_{\rho,\mu} \in \mathscr{Q}$, we have*

$$\frac{1}{p}\int_{\Omega_2} |g-f_2|^p \mathrm{d}x\mathrm{d}y \ge \frac{h^{p+1}}{2^p p(p+1)\frac{2h}{\mu}\rho'(0)}\Big(2(R-\mu)+\frac{\pi\mu}{2}\Big). \qquad (3.14)$$

Proof. Let $g = z_{\rho,\mu} \in \mathscr{Q}$. We establish a lower bound of $\frac{1}{p}\int_{\Omega_2} |g-f_2|^p \mathrm{d}x\mathrm{d}y$ by estimating it over three subregions given in (3.9), respectively.

We first consider $\frac{1}{p}\int_{[\mu,R)\times(-R,R)} |g-f_2|^p \mathrm{d}x\mathrm{d}y$. We note that, on any x-section with $x \in [\mu,R)$, $T(y) = h\rho(0) + \frac{2h}{\mu}\rho'(0)y$ is a tangent line to the curve $z(y) = h\rho(\frac{2y}{\mu})$ at $(0, h\rho(0))$ and intersects with $z = 0$ and $z = h$ at two points $(y_-, 0)$ and (y_+, h), respectively. Thus, we calculate

$$\begin{aligned}
&\frac{1}{p}\int_{[\mu,R)\times(-R,R)} |g-f_2|^p \mathrm{d}x\mathrm{d}y\\
=& \frac{1}{p}\int_\mu^R \int_{-\frac{\mu}{2}}^0 |0-g|^p \mathrm{d}x\mathrm{d}y + \frac{1}{p}\int_\mu^R \int_0^{\frac{\mu}{2}} |h-g|^p \mathrm{d}x\mathrm{d}y\\
\ge& \frac{R-\mu}{p}\int_{y_-}^0 \Big(h\rho(0)+\frac{2h}{\mu}\rho'(0)y\Big)^p \mathrm{d}y + \frac{R-\mu}{p}\int_0^{y_+} \Big(h-h\rho(0)-\frac{2h}{\mu}\rho'(0)y\Big)^p \mathrm{d}y\\
=& \frac{(R-\mu)h^{p+1}}{p(p+1)\frac{2h}{\mu}\rho'(0)}(\rho(0)^{p+1}+(1-\rho(0))^{p+1}). \qquad (3.15)
\end{aligned}$$

Similarly, we have

$$\frac{1}{p}\int_{(-R,\mu)\times[\mu,R)} |g-f_2|^p \mathrm{d}x\mathrm{d}y \ge \frac{(R-\mu)h^{p+1}}{p(p+1)\frac{2h}{\mu}\rho'(0)}(\rho(0)^{p+1}+(1-\rho(0))^{p+1}). \qquad (3.16)$$

To estimate the integral $\frac{1}{p}\int_{(-R,\mu)\times(-R,\mu)} |g-f_2|^p \mathrm{d}x\mathrm{d}y$, we construct a surface patch by rotating the tangent line $T(y) = h\rho(0) + \frac{2h}{\mu}\rho'(0)y$ from $\frac{3\pi}{2}$ to π around the axis $(\mu,\mu,z), z \in (-\infty,+\infty)$. The surface patch is then expressed as $\bar{T}(x,y) = h\rho(0) - \frac{2h}{\mu}\rho'(0)(\sqrt{(x-\mu)^2+(y-\mu)^2} - \mu)$. On any θ-section with $\theta \in [\pi, \frac{3\pi}{2}]$, $\bar{T}(x,y)$ can be written as $T(r_\mu) = h\rho(0) - \frac{2h}{\mu}\rho'(0)(r_\mu - \mu)$, which intersects with $z = h$ and $z = 0$ at

$(R_{\rm in},h)$ and $(R_{\rm out},0)$, respectively. Then we derive

$$\begin{aligned}\frac{1}{p}\int_{(-R,\mu)\times(-R,\mu)}|g-f_2|^p\mathrm{d}x\mathrm{d}y &\geq \frac{1}{p}\int_{(-R,\mu)\times(-R,\mu)}|f_2-\bar{T}|^p\mathrm{d}x\mathrm{d}y\\ &\geq \frac{\pi}{2p}\int_{R_{\rm in}}^{\mu}(h-T(r_\mu))^p r_\mu \mathrm{d}r_\mu+\frac{\pi}{2p}\int_{\mu}^{R_{\rm out}}(T(r_\mu))^p r_\mu\mathrm{d}r_\mu\\ &=\frac{\pi\mu h^{p+1}}{2p(p+1)\frac{2h}{\mu}\rho'(0)}((1-\rho(0))^{p+1}+\rho(0)^{p+1})\\ &\quad+\frac{\pi h^{p+2}}{2p(p+1)(p+2)\left(\frac{2h}{\mu}\rho'(0)\right)^2}(\rho(0)^{p+2}-(1-\rho(0))^{p+2}),\end{aligned} \tag{3.17}$$

where the first inequality is due to the fact that $\rho(s)$ is convex over $s\in(-\infty,0)$ and concave over $s\in(0,+\infty)$, and the second inequality is because of $T(r_\mu)\in[0,\frac{h}{2}]$ for $r_\mu\in[\mu,R_{\rm out}]$.

Combining (3.15), (3.16) and (3.17) and using $\rho(0)=\frac{1}{2}$, we complete the proof. ☐

Theorem 3.2. *Consider the model (2.1) with $f=f_2$ given in (3.8). Let ψ satisfy Assumption A1-A3 and $p\in[1,+\infty)$. Then there exists a positive constant*

$$c_2=\frac{(\pi-2)Rh^{p+1}}{2^p p(p+1)}\frac{1}{4R\sup\limits_{s\in[4h/R,+\infty)}\{s^2(\psi'(s)s+\psi(s))\}+4h(4-\pi)\sup\limits_{s\in[4h/R,+\infty)}\{-s^2\psi'(s)\}},$$

such that $E(f_2)=\inf_{g\in\mathscr{Q}}E(g)$, when $\lambda<c_2$.

Proof. Combining the inequalities (3.13) and (3.14), we obtain the following lower bound estimate

$$E(g)\geq\lambda\psi\Big(\frac{2h}{\mu}\rho'(0)\Big)\frac{2h}{\mu}\rho'(0)(4R+(\pi-4)\mu)+\frac{h^{p+1}}{2^p p(p+1)\frac{2h}{\mu}\rho'(0)}\Big(2R+\Big(\frac{\pi}{2}-2\Big)\mu\Big).$$

We therefore consider a function $\zeta:[4h/R,+\infty)\to\mathbb{R}$ as follows

$$\zeta(s)=\lambda\psi(s)s\left(4R+(\pi-4)\frac{2h\rho'(0)}{s}\right)+\frac{h^{p+1}}{2^p p(p+1)}\left(\frac{2R}{s}+\Big(\frac{\pi}{2}-2\Big)\frac{2h\rho'(0)}{s^2}\right),$$

whose derivative is

$$\zeta'(s)=\lambda(4R(\psi'(s)s+\psi(s))+(\pi-4)2h\rho'(0)\psi'(s))-\frac{h^{p+1}}{2^p p(p+1)}\left(\frac{2R}{s^2}+\frac{2(\pi-4)h\rho'(0)}{s^3}\right).$$

Since $1\leq\rho'(0)\leq 2$ and $s\in[4h/R,+\infty)$, we get $2R+(\pi-4)\frac{2h\rho'(0)}{s}\geq(\pi-2)R$. Let

$$c_2:=\frac{(\pi-2)Rh^{p+1}}{2^p p(p+1)}\frac{1}{4R\sup\limits_{s\in[\frac{4h}{R},+\infty)}\{s^2(\psi'(s)s+\psi(s))\}+4h(4-\pi)\sup\limits_{s\in[\frac{4h}{R},+\infty)}\{-s^2\psi'(s)\}}.$$

Hence, we see that if $\lambda<c_2$, then $\zeta'(s)<0$, $\forall\ s\in[4h/R,+\infty)$ and $\forall\ \rho\in\mathscr{P}$, and thus $\zeta(s)$ decreases over $[4h/R,+\infty)$.

Let $s=\frac{2h}{\mu}\rho'(0)$ with $\rho\in\mathscr{P}$. We derive $E(g)>E(f_2)$ for any smooth $g\in\mathscr{Q}$. Besides, by using the dominated convergence theorem, we know that $E(g)\to E(f_2)$ when $g=z_{\rho,\mu}$ approaches f_2 through $\mu\to 0^+$. These mean $E(f_2)=\inf_{g\in\mathscr{Q}}E(g)$ when $\lambda<c_2$. ☐

Theorem 3.2 indicates that the model (2.1) has corner and contrast preservation ability for f_2 like input images, as the regularization term does not penalize the image height h shown in Lemma 3.3.

Remark 3.2. The above proof procedures are motivated by [57] and also [56], where [57] considered the model (2.2) with $\psi(s)=\frac{1}{\sqrt{1+s^2}}$ and $p=2$, and [56] considered the model (2.1) with $\psi(s)=\frac{1}{\sqrt{1+s^2}}$ and $p=2$. To avoid the incompleteness of (22) and (26) in [56], in this paper, we slightly modify the definitions of $\mathscr{S}$ and $\mathscr{P}$ in [56,57] by introducing natural symmetries $\varrho'(s)=\varrho'(2R-s), s\in[0,2R]$ for $\mathscr{S}$ and $\rho'(s)=\rho'(-s)$ for $\mathscr{P}$, and deriving new inequality estimations for $\int_{\Omega_1}\|\mathrm{H}_\psi(f_1)\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y$ in Lemma 3.1 and $\int_{\Omega_2}\|\mathrm{H}_\psi(f_2)\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y$ in Lemma 3.3. Moreover, our proofs apply to model (2.1) with a general ψ and $p\in[1,+\infty)$.

Remark 3.3. For the model (2.2), by arguments similar with the above, one can show similar theoretical results as Theorems 3.1 and 3.2, and the symmetry assumptions on $\varrho\in\mathscr{S}$ and $\rho\in\mathscr{P}$ can be removed if $1\le p\le 3$.

4. An accelerated operator-splitting method for solving (2.1)

The model (2.1) is clearly nonconvex and nonsmooth, and thus very difficult to be solved efficiently. For related nonconvex high-order models like the EE model (1.4) and GC model (1.6), except gradient descent methods and augmented Lagrangian methods, people recently discovered that basic operator-splitting methods with Lie scheme and Marchuk-Yanenko discretization [17,21,22,28] work surprisingly well with less tuning parameters and stable performances. However, we note that all these studies on such basic operator-splitting methods do not consider acceleration strategies, even when applied to the convex TV model [17,22].

In this section, we devise an efficient operator-splitting method to solve (2.1), by introducing some inertial extrapolation and approximation techniques to the basic operator-splitting method with Lie scheme and Marchuk-Yanenko discretization. Although our theoretical analysis in Section 3 applies to the model (2.1) with any $p\in[1,+\infty)$, here we focus on two special cases, $p=2$ for Gaussian noise removal and $p=1$ for impulsive noise removal, which are the most common in image processing.

4.1. The case of $p=2$.

The model (2.1) with $p=2$ can be reformulated into

$$\inf_u\Big\{E(u)=\lambda\int_\Omega\|\nabla u(\nabla(\psi(|\nabla u|)))^\top+\psi(|\nabla u|)\nabla^2 u\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y+\frac{1}{2}\int_\Omega|u-f|^2\mathrm{d}x\mathrm{d}y\Big\},\tag{4.1}$$

under some smoothness condition on u. We will now discuss how to solve (4.1) efficiently.

We define a function set as

$$\mathscr{W}_1(\Omega)=\mathscr{L}^2(\Omega)\times(\mathscr{L}^2(\Omega))^2\times(\mathscr{L}^1(\Omega))^{2\times 2},$$

and introduce two auxiliary variables

$$\mathbf{p}=\begin{pmatrix}p^1\\ p^2\end{pmatrix}\in(\mathscr{L}^2(\Omega))^2,\quad \mathbf{H}=\begin{pmatrix}H^{11} & H^{12}\\ H^{21} & H^{22}\end{pmatrix}\in(\mathscr{L}^1(\Omega))^{2\times 2}.$$

We thus reformulate (4.1) into the following constrained optimization problem:

$$\begin{cases}\displaystyle\inf_{(u,\mathbf{p},\mathbf{H})\in\mathscr{W}_1(\Omega)}\Big\{\lambda\int_\Omega\big\|\mathbf{p}(\nabla(\psi(|\mathbf{p}|)))^\top+\psi(|\mathbf{p}|)\mathbf{H}\big\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y+\frac{1}{2}\int_\Omega|u-f|^2\mathrm{d}x\mathrm{d}y\Big\},\\ \quad\text{s.t.}\quad \mathbf{p}=\nabla u,\ \mathbf{H}=\nabla^2 u.\end{cases}\tag{4.2}$$

Denote

$$E_1(\mathbf{p},\mathbf{H})=\lambda\int_\Omega \left\|\mathbf{p}(\nabla(\psi(|\mathbf{p}|)))^\top+\psi(|\mathbf{p}|)\mathbf{H}\right\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y, \text{ and } E_2(u)=\frac{1}{2}\int_\Omega |u-f|^2\mathrm{d}x\mathrm{d}y.$$

We also define a set Γ by

$$\Gamma=\left\{(u,\mathbf{p},\mathbf{H})\middle|(u,\mathbf{p},\mathbf{H})\in\mathscr{W}_1(\Omega) \text{ such that } \mathbf{p}=\nabla u,\mathbf{H}=\nabla^2 u\right\}$$

and write its indicator functional as

$$\mathrm{I}_\Gamma(u,\mathbf{p},\mathbf{H})=\begin{cases}0, & \text{if } (u,\mathbf{p},\mathbf{H})\in\Gamma,\\ +\infty, & \text{otherwise.}\end{cases}$$

The problem (4.2) is thus equivalent to

$$\inf_{(u,\mathbf{p},\mathbf{H})\in\mathscr{W}_1(\Omega)}\{E_1(\mathbf{p},\mathbf{H})+E_2(u)+\mathrm{I}_\Gamma(u,\mathbf{p},\mathbf{H})\}. \tag{4.3}$$

To design a stable, easy-to-implement and efficient algorithm for (4.3), we follow the operator-splitting method with Lie scheme and Marchuk-Yanenko discretization [17, 22, 24, 28] and further introduce some acceleration [1, 30, 37, 41] and approximation techniques [10, 54, 56], to construct our accelerated operator-splitting (AOS) solver. Considering the well-known equivalence between the backward Euler scheme and proximal point method, we here derive our solver by splitting the objective functionals, instead of using optimality condition and subgradient flow as in [17, 22, 28]. In particular, we split its objective functional into two parts, i.e., $E_1(\mathbf{p},\mathbf{H})+E_2(u)$ and $\mathrm{I}_\Gamma(u,\mathbf{p},\mathbf{H})$, and optimize them progressively within each iteration. Denote $n=0,1,2,\ldots$ as the iteration number. This strategy leads to the following

- Fractional step 1: from n to $n+\frac{1}{2}$,

$$\begin{aligned}&(u^{n+\frac{1}{2}},\mathbf{p}^{n+\frac{1}{2}},\mathbf{H}^{n+\frac{1}{2}})\\ &=\underset{(u,\mathbf{p},\mathbf{H})\in\mathscr{W}_1(\Omega)}{\arg\min}\left\{\frac{\gamma_1}{2}\int_\Omega(u-u^n)^2\mathrm{d}x\mathrm{d}y+\frac{\gamma_2}{2}\int_\Omega|\mathbf{p}-\mathbf{p}^n|^2\mathrm{d}x\mathrm{d}y\right.\\ &\qquad\left.+\frac{\gamma_3}{2}\int_\Omega\|\mathbf{H}-\mathbf{H}^n\|_{\mathrm{F}}^2\mathrm{d}x\mathrm{d}y+E_1(\mathbf{p},\mathbf{H})+E_2(u)\right\};\end{aligned} \tag{4.4}$$

- Fractional step 2: from $n+\frac{1}{2}$ to $n+1$,

$$\begin{aligned}&(u^{n+1},\mathbf{p}^{n+1},\mathbf{H}^{n+1})\\ &=\underset{(u,\mathbf{p},\mathbf{H})\in\mathscr{W}_1(\Omega)}{\arg\min}\left\{\frac{\gamma_1}{2}\int_\Omega(u-u^{n+\frac{1}{2}})^2\mathrm{d}x\mathrm{d}y+\frac{\gamma_2}{2}\int_\Omega|\mathbf{p}-\mathbf{p}^{n+\frac{1}{2}}|^2\mathrm{d}x\mathrm{d}y\right.\\ &\qquad\left.+\frac{\gamma_3}{2}\int_\Omega\|\mathbf{H}-\mathbf{H}^{n+\frac{1}{2}}\|_{\mathrm{F}}^2\mathrm{d}x\mathrm{d}y+\mathrm{I}_\Gamma(u,\mathbf{p},\mathbf{H})\right\},\end{aligned} \tag{4.5}$$

where $\gamma_1,\gamma_2,\gamma_3$ are positive parameters to control the proximal terms for u, $\mathbf{p}$, $\mathbf{H}$, respectively.

We then introduce the inertial acceleration technique [1, 30, 37, 41] to (4.4) and (4.5) to obtain the following computational scheme:

- Fractional step 1 with extrapolation: from n to $n+\frac{1}{2}$,

$$\begin{aligned}&(u^{n+\frac{1}{2}},\mathbf{p}^{n+\frac{1}{2}},\mathbf{H}^{n+\frac{1}{2}})\\&=\underset{(u,\mathbf{p},\mathbf{H})\in\mathscr{W}_1(\Omega)}{\arg\min}\left\{\frac{\gamma_1}{2}\int_\Omega(u-\bar{u}^n)^2\mathrm{d}x\mathrm{d}y+\frac{\gamma_2}{2}\int_\Omega|\mathbf{p}-\bar{\mathbf{p}}^n|^2\mathrm{d}x\mathrm{d}y\right.\\&\qquad\left.+\frac{\gamma_3}{2}\int_\Omega\|\mathbf{H}-\bar{\mathbf{H}}^n\|_{\mathrm{F}}^2\mathrm{d}x\mathrm{d}y+E_1(\mathbf{p},\mathbf{H})+E_2(u)\right\},\end{aligned}\tag{4.6}$$

where

$$\begin{cases}\bar{u}^n=u^n+t\left(u^n-u^{(n-1)+\frac{1}{2}}\right),\\\bar{\mathbf{p}}^n=\mathbf{p}^n+t\left(\mathbf{p}^n-\mathbf{p}^{(n-1)+\frac{1}{2}}\right),\\\bar{\mathbf{H}}^n=\mathbf{H}^n+t\left(\mathbf{H}^n-\mathbf{H}^{(n-1)+\frac{1}{2}}\right),\end{cases}\tag{4.7}$$

with $\left(\bar{u}^0,\bar{\mathbf{p}}^0,\bar{\mathbf{H}}^0\right)=(u^0,\mathbf{p}^0,\mathbf{H}^0)$ convention;

- Fractional step 2 with extrapolation: from $n+\frac{1}{2}$ to $n+1$,

$$\begin{aligned}&(u^{n+1},\mathbf{p}^{n+1},\mathbf{H}^{n+1})\\&=\underset{(u,\mathbf{p},\mathbf{H})\in\mathscr{W}_1(\Omega)}{\arg\min}\left\{\frac{\gamma_1}{2}\int_\Omega(u-\bar{u}^{n+\frac{1}{2}})^2\mathrm{d}x\mathrm{d}y+\frac{\gamma_2}{2}\int_\Omega|\mathbf{p}-\bar{\mathbf{p}}^{n+\frac{1}{2}}|^2\mathrm{d}x\mathrm{d}y\right.\\&\qquad\left.+\frac{\gamma_3}{2}\int_\Omega\|\mathbf{H}-\bar{\mathbf{H}}^{n+\frac{1}{2}}\|_{\mathrm{F}}^2\mathrm{d}x\mathrm{d}y+\mathrm{I}_\Gamma(u,\mathbf{p},\mathbf{H})\right\},\end{aligned}\tag{4.8}$$

where

$$\begin{cases}\bar{u}^{n+\frac{1}{2}}=u^{n+\frac{1}{2}}+t\left(u^{n+\frac{1}{2}}-u^n\right),\\\bar{\mathbf{p}}^{n+\frac{1}{2}}=\mathbf{p}^{n+\frac{1}{2}}+t\left(\mathbf{p}^{n+\frac{1}{2}}-\mathbf{p}^n\right),\\\bar{\mathbf{H}}^{n+\frac{1}{2}}=\mathbf{H}^{n+\frac{1}{2}}+t\left(\mathbf{H}^{n+\frac{1}{2}}-\mathbf{H}^n\right).\end{cases}\tag{4.9}$$

The t in (4.7) and (4.9) is an extrapolation parameter and belongs to $[0,1)$.

We are now left (4.6) and (4.8) to address. As many experiences show that there is no need to solve all subproblems accurately, we will adopt some proper approximation techniques like reweighting [10, 54, 56], to simplify and further speed up our computation. For the problem (4.6), we see that u is decoupled, but $\mathbf{p},\mathbf{H}$ are still coupled. Thus, we solve (4.6) in Gauss-Seidel flavor, i.e., compute $u^{n+\frac{1}{2}},\mathbf{p}^{n+\frac{1}{2}},\mathbf{H}^{n+\frac{1}{2}}$ progressively. The u-subproblem in (4.6) is trivial and

$$u^{n+\frac{1}{2}}=\frac{f+\gamma_1\bar{u}^n}{1+\gamma_1}.\tag{4.10}$$

The $\mathbf{p}$-subproblem in (4.6) is

$$\mathbf{p}^{n+\frac{1}{2}}=\underset{\mathbf{p}\in(\mathscr{L}^2(\Omega))^2}{\arg\min}\left\{\frac{\gamma_2}{2}\int_\Omega|\mathbf{p}-\bar{\mathbf{p}}^n|^2\mathrm{d}x\mathrm{d}y+\lambda\int_\Omega\left\|\mathbf{p}(\nabla(\psi(|\mathbf{p}|)))^\top+\psi(|\mathbf{p}|)\mathbf{H}^n\right\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y\right\},$$

and approximated as the following

$$\mathbf{p}^{n+\frac{1}{2}}\approx\underset{\mathbf{p}\in(\mathscr{L}^2(\Omega))^2}{\arg\min}\left\{\frac{\gamma_2}{2}\int_\Omega|\mathbf{p}-\bar{\mathbf{p}}^n|^2\mathrm{d}x\mathrm{d}y\right.$$

$$
\begin{aligned}
&\qquad +\lambda\int_{\Omega}\Big[\big\|\mathbf{p}(\nabla(\psi(|\mathbf{p}^n|)))^{\top}\big\|_{\mathrm{F}}+\big\|\psi(|\mathbf{p}^n|)\mathbf{H}^n\big\|_{\mathrm{F}}\Big]\mathrm{d}x\mathrm{d}y\Big\}\\
&=\underset{\mathbf{p}\in(\mathscr{L}^2(\Omega))^2}{\arg\min}\Big\{\frac{\gamma_2}{2}\int_{\Omega}|\mathbf{p}-\bar{\mathbf{p}}^n|^2\mathrm{d}x\mathrm{d}y\\
&\qquad +\lambda\int_{\Omega}\Big[\big|\nabla(\psi(|\mathbf{p}^n|))\big||\mathbf{p}|+\big\|\psi(|\mathbf{p}^n|)\mathbf{H}^n\big\|_{\mathrm{F}}\Big]\mathrm{d}x\mathrm{d}y\Big\},
\end{aligned}
$$

where the handling of the regularization term is inspired by the reweighting and approximation techniques in [10, 54, 56]. Thus we calculate $\mathbf{p}^{n+\frac{1}{2}}$ approximately by the above minimization problem, whose solution has the following closed-form [50, 53]

$$
\mathbf{p}^{n+\frac{1}{2}}=\max\Big\{0,1-\frac{\lambda|\nabla(\psi(|\mathbf{p}^n|))|}{\gamma_2|\bar{\mathbf{p}}^n|}\Big\}\bar{\mathbf{p}}^n. \tag{4.11}
$$

The **H**-subproblem in (4.6), after reweighting approximation, is as follows

$$
\begin{aligned}
\mathbf{H}^{n+\frac{1}{2}}\approx \underset{\mathbf{H}\in(\mathscr{L}^1(\Omega))^{2\times 2}}{\arg\min}&\Big\{\frac{\gamma_3}{2}\int_{\Omega}\|\mathbf{H}-\bar{\mathbf{H}}^n\|_{\mathrm{F}}^2\mathrm{d}x\mathrm{d}y\\
&+\lambda\int_{\Omega}\big\|(\mathbf{p}^{n+\frac{1}{2}})(\nabla(\psi(|\mathbf{p}^{n+\frac{1}{2}}|)))^{\top}+\psi(|\mathbf{p}^{n+\frac{1}{2}}|)\mathbf{H}\big\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y\Big\}.
\end{aligned}
$$

Set $\mathbf{B}^{n+\frac{1}{2}}=(\mathbf{p}^{n+\frac{1}{2}})(\nabla(\psi(|\mathbf{p}^{n+\frac{1}{2}}|)))^{\top}$. Hence we approximate $\mathbf{H}^{n+\frac{1}{2}}$ by the closed-form solution of the above minimization problem, which reads

$$
\mathbf{H}^{n+\frac{1}{2}}=\frac{\mathbf{T}^{n+\frac{1}{2}}-\mathbf{B}^{n+\frac{1}{2}}}{\psi(|\mathbf{p}^{n+\frac{1}{2}}|)}, \tag{4.12}
$$

where $\mathbf{T}^{n+\frac{1}{2}}=\max\Big\{0,1-\frac{(\psi(|\mathbf{p}^{n+\frac{1}{2}}|))^2\lambda}{\gamma_3\|\mathbf{B}^{n+\frac{1}{2}}+\psi(|\mathbf{p}^{n+\frac{1}{2}}|)\bar{\mathbf{H}}^n\|_{\mathrm{F}}}\Big\}(\mathbf{B}^{n+\frac{1}{2}}+\psi(|\mathbf{p}^{n+\frac{1}{2}}|)\bar{\mathbf{H}}^n)$.

As for (4.8), similar to [17], we can convert it into the following equivalent constrained optimization problem

$$
\begin{aligned}
(u^{n+1},\mathbf{p}^{n+1},\mathbf{H}^{n+1})=\underset{(u,\mathbf{p},\mathbf{H})\in\mathscr{W}_1(\Omega)}{\arg\min}&\Big\{\frac{\gamma_1}{2}\int_{\Omega}(u-\bar{u}^{n+\frac{1}{2}})^2\mathrm{d}x\mathrm{d}y+\frac{\gamma_2}{2}\int_{\Omega}|\mathbf{p}-\bar{\mathbf{p}}^{n+\frac{1}{2}}|^2\mathrm{d}x\mathrm{d}y\\
&+\frac{\gamma_3}{2}\int_{\Omega}\|\mathbf{H}-\bar{\mathbf{H}}^{n+\frac{1}{2}}\|_{\mathrm{F}}^2\mathrm{d}x\mathrm{d}y\Big\},\\
&\text{s.t. } \mathbf{p}=\nabla u,\ \mathbf{H}=\nabla^2 u.
\end{aligned}
$$

Substituting the equality constraints into the objective function, we see that $(u^{n+1},\mathbf{p}^{n+1},\mathbf{H}^{n+1})$ can be calculated as follows

$$
\begin{cases}
u^{n+1}=\underset{u\in\mathscr{L}^2(\Omega)}{\arg\min}\Big\{\frac{\gamma_1}{2}\int_{\Omega}(u-\bar{u}^{n+\frac{1}{2}})^2\mathrm{d}x\mathrm{d}y+\frac{\gamma_2}{2}\int_{\Omega}|\nabla u-\bar{\mathbf{p}}^{n+\frac{1}{2}}|^2\mathrm{d}x\mathrm{d}y\\
\qquad\qquad+\frac{\gamma_3}{2}\int_{\Omega}\|\nabla^2 u-\bar{\mathbf{H}}^{n+\frac{1}{2}}\|_{\mathrm{F}}^2\mathrm{d}x\mathrm{d}y\Big\},\\
\mathbf{p}^{n+1}=\nabla u^{n+1},\\
\mathbf{H}^{n+1}=\nabla^2 u^{n+1}.
\end{cases}
$$

Therein, the u-subproblem has the following Euler-Lagrange equation

$$(\gamma_1\mathrm{Id}-\gamma_2\Delta+\gamma_3\Delta^2)u^{n+1}=\gamma_1\bar{u}^{n+\frac{1}{2}}-\gamma_2\,\mathrm{div}\,\bar{\mathbf{p}}^{n+\frac{1}{2}}+\gamma_3\,\mathrm{div}^2\,\bar{\mathbf{H}}^{n+\frac{1}{2}},\tag{4.13}$$

where Id is the identity operator, and div^2 is the adjoint of the Hessian operator. We can use FFT to solve (4.13) as in [50], by simply imposing the periodic boundary condition; otherwise, one can utilize an iterative linear solver for the case of Neumann boundary condition or the approach in [34].

The above derivation leads to a new efficient operator-splitting method, which is summarized in Algorithm 1.

Algorithm 1 Accelerated operator-splitting algorithm to solve the model (2.1) (AOS-TSGV) in the case of $p=2$

```
1. Input: The noisy image $f$, parameters $\lambda,\gamma_1,\gamma_2,\gamma_3,t$, initial value $(u_0,\mathbf{p}_0,\mathbf{H}_0)$, and function $\psi$;
2. For n=0,1,2,...:
    2.1 From $n$ to $n+\frac{1}{2}$: set $(\bar{u}^n,\bar{\mathbf{p}}^n,\bar{\mathbf{H}}^n)$ by (4.7), and compute $(u^{n+\frac{1}{2}},\mathbf{p}^{n+\frac{1}{2}},\mathbf{H}^{n+\frac{1}{2}})$ by (4.10) (4.11) (4.12);
    2.2 From $n+\frac{1}{2}$ to $n+1$: set $(\bar{u}^{n+\frac{1}{2}},\bar{\mathbf{p}}^{n+\frac{1}{2}},\bar{\mathbf{H}}^{n+\frac{1}{2}})$ by (4.9), and compute $(u^{n+1},\mathbf{p}^{n+1},\mathbf{H}^{n+1})=(u^{n+1},\nabla u^{n+1},\nabla^2 u^{n+1})$ by (4.13).
```

4.2. The case of $p=1$. Similarly to Section 4.1, we convert the model (2.1) with $p=1$ to the following

$$\inf_u\Big\{E(u)=\lambda\int_\Omega\|\nabla u(\nabla(\psi(|\nabla u|)))^\top+\psi(|\nabla u|)\nabla^2 u\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y+\int_\Omega|u-f|\mathrm{d}x\mathrm{d}y\Big\}.\tag{4.14}$$

To devise an efficient operator-splitting method as in Section 4.1, we introduce a function space

$$\mathscr{W}_2(\Omega)=\mathscr{L}^1(\Omega)\times(\mathscr{L}^2(\Omega))^2\times(\mathscr{L}^1(\Omega))^{2\times2}\times\mathscr{L}^1(\Omega),$$

and three auxiliary variables

$$\mathbf{p}=\begin{pmatrix}p^1\\p^2\end{pmatrix}\in(\mathscr{L}^2(\Omega))^2,\mathbf{H}=\begin{pmatrix}H^{11}&H^{12}\\H^{21}&H^{22}\end{pmatrix}\in(\mathscr{L}^1(\Omega))^{2\times2},\text{ and }v\in\mathscr{L}^1(\Omega).$$

Then we formulate (4.14) into the following constrained optimization problem

$$\begin{cases}\inf\limits_{(u,\mathbf{p},\mathbf{H},v)\in\mathscr{W}_2(\Omega)}\Big\{\lambda\int_\Omega\big\|\mathbf{p}(\nabla(\psi(|\mathbf{p}|)))^\top+\psi(|\mathbf{p}|)\mathbf{H}\big\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y+\int_\Omega|v|\mathrm{d}x\mathrm{d}y\Big\},\\ \text{s.t.}\qquad \mathbf{p}=\nabla u,\mathbf{H}=\nabla^2u,v=u-f.\end{cases}\tag{4.15}$$

We then still denote

$$E_1(\mathbf{p},\mathbf{H})=\lambda\int_\Omega\big\|\mathbf{p}(\nabla(\psi(|\mathbf{p}|)))^\top+\psi(|\mathbf{p}|)\mathbf{H}\big\|_{\mathrm{F}}\mathrm{d}x\mathrm{d}y,$$

and define

$$E_3(v)=\int_\Omega|v|\mathrm{d}x\mathrm{d}y,$$

and a set Σ as

$$\Sigma=\Big\{(u,\mathbf{p},\mathbf{H},v)\in\mathscr{W}_2(\Omega)\text{ such that }\mathbf{p}=\nabla u,\mathbf{H}=\nabla^2 u,v=u-f\Big\}$$

with its indicator functional

$$\mathrm{I}_\Sigma(u,\mathbf{p},\mathbf{H},v)=\begin{cases}0, & \text{if }(u,\mathbf{p},\mathbf{H},v)\in\Sigma,\\ +\infty, & \text{otherwise.}\end{cases}$$

The problem (4.15) is thus equivalent to

$$\inf_{(u,\mathbf{p},\mathbf{H},v)\in\mathscr{W}_2(\Omega)}\big\{E_1(\mathbf{p},\mathbf{H})+E_3(v)+\mathrm{I}_\Sigma(u,\mathbf{p},\mathbf{H},v)\big\}. \tag{4.16}$$

We split the objective functional of (4.16) into two parts, i.e., $E_1(\mathbf{p},\mathbf{H})+E_3(v)$ and $\mathrm{I}_\Sigma(u,\mathbf{p},\mathbf{H},v)$, and optimize them by the accelerated strategy similar to (4.6) and (4.8) as following

- Fractional step 1 with extrapolation: from n to $n+\frac{1}{2}$,

$$\begin{aligned}&(u^{n+\frac{1}{2}},\mathbf{p}^{n+\frac{1}{2}},\mathbf{H}^{n+\frac{1}{2}},v^{n+\frac{1}{2}})\\ =&\operatorname*{arg\,min}_{(u,\mathbf{p},\mathbf{H},v)\in\mathscr{W}_2(\Omega)}\Big\{\frac{\gamma_1}{2}\int_\Omega(u-\bar{u}^n)^2\mathrm{d}x\mathrm{d}y+\frac{\gamma_2}{2}\int_\Omega|\mathbf{p}-\bar{\mathbf{p}}^n|^2\mathrm{d}x\mathrm{d}y\\ &+\frac{\gamma_3}{2}\int_\Omega\|\mathbf{H}-\bar{\mathbf{H}}^n\|_{\mathrm{F}}^2\mathrm{d}x\mathrm{d}y+\frac{\gamma_4}{2}\int_\Omega(v-\bar{v}^n)^2\mathrm{d}x\mathrm{d}y+E_1(\mathbf{p},\mathbf{H})+E_3(v)\Big\},\end{aligned} \tag{4.17}$$

where

$$\begin{cases}\bar{u}^n=u^n+t\left(u^n-u^{(n-1)+\frac{1}{2}}\right),\\ \bar{\mathbf{p}}^n=\mathbf{p}^n+t\left(\mathbf{p}^n-\mathbf{p}^{(n-1)+\frac{1}{2}}\right),\\ \bar{\mathbf{H}}^n=\mathbf{H}^n+t\left(\mathbf{H}^n-\mathbf{H}^{(n-1)+\frac{1}{2}}\right),\\ \bar{v}^n=v^n+t\left(v^n-v^{(n-1)+\frac{1}{2}}\right),\end{cases} \tag{4.18}$$

with $\left(\bar{u}^0,\bar{\mathbf{p}}^0,\bar{\mathbf{H}}^0,\bar{v}^0\right)=\left(u^0,\mathbf{p}^0,\mathbf{H}^0,v^0\right)$ convention;

- Fractional step 2 with extrapolation: from $n+\frac{1}{2}$ to $n+1$,

$$\begin{aligned}&(u^{n+1},\mathbf{p}^{n+1},\mathbf{H}^{n+1},v^{n+1})\\ =&\operatorname*{arg\,min}_{(u,\mathbf{p},\mathbf{H},v)\in\mathscr{W}_2(\Omega)}\Big\{\frac{\gamma_1}{2}\int_\Omega(u-\bar{u}^{n+\frac{1}{2}})^2\mathrm{d}x\mathrm{d}y+\frac{\gamma_2}{2}\int_\Omega|\mathbf{p}-\bar{\mathbf{p}}^{n+\frac{1}{2}}|^2\mathrm{d}x\mathrm{d}y\\ &+\frac{\gamma_3}{2}\int_\Omega\|\mathbf{H}-\bar{\mathbf{H}}^{n+\frac{1}{2}}\|_{\mathrm{F}}^2\mathrm{d}x\mathrm{d}y+\frac{\gamma_4}{2}\int_\Omega(v-\bar{v}^{n+\frac{1}{2}})^2\mathrm{d}x\mathrm{d}y+\mathrm{I}_\Sigma(u,\mathbf{p},\mathbf{H},v)\Big\},\end{aligned} \tag{4.19}$$

where

$$\begin{cases}\bar{u}^{n+\frac{1}{2}}=u^{n+\frac{1}{2}}+t\left(u^{n+\frac{1}{2}}-u^n\right),\\ \bar{\mathbf{p}}^{n+\frac{1}{2}}=\mathbf{p}^{n+\frac{1}{2}}+t\left(\mathbf{p}^{n+\frac{1}{2}}-\mathbf{p}^n\right),\\ \bar{\mathbf{H}}^{n+\frac{1}{2}}=\mathbf{H}^{n+\frac{1}{2}}+t\left(\mathbf{H}^{n+\frac{1}{2}}-\mathbf{H}^n\right),\\ \bar{v}^{n+\frac{1}{2}}=v^{n+\frac{1}{2}}+t\left(v^{n+\frac{1}{2}}-v^n\right),\end{cases} \tag{4.20}$$

and $\gamma_1,\gamma_2,\gamma_3,\gamma_4$ are positive parameters to control the proximal terms for u, $\mathbf{p}$, $\mathbf{H}$, v, respectively.

(4.17) and (4.19) can be solved similarly as (4.6) and (4.8). In particular, for (4.17),

$$u^{n+\frac{1}{2}}=\bar{u}^n, \tag{4.21}$$

$\mathbf{p}^{n+\frac{1}{2}},\mathbf{H}^{n+\frac{1}{2}}$ can be calculated approximately by (4.11) and (4.12), and $v^{n+\frac{1}{2}}$ is clearly as follows

$$v^{n+\frac{1}{2}}=\max\left\{0,1-\frac{1}{\gamma_4|\bar{v}^n|}\right\}\bar{v}^n. \tag{4.22}$$

In (4.19), $(u^{n+1},\mathbf{p}^{n+1},\mathbf{H}^{n+1},v^{n+1})$ can be calculated by

$$\begin{cases} u^{n+1}=\underset{u\in\mathscr{L}^1(\Omega)}{\arg\min}\Big\{\frac{\gamma_1}{2}\int_\Omega(u-\bar{u}^{n+\frac{1}{2}})^2\mathrm{d}x\mathrm{d}y+\frac{\gamma_2}{2}\int_\Omega|\nabla u-\bar{\mathbf{p}}^{n+\frac{1}{2}}|^2\mathrm{d}x\mathrm{d}y \\ \qquad\qquad +\frac{\gamma_3}{2}\int_\Omega\|\nabla^2u-\bar{\mathbf{H}}^{n+\frac{1}{2}}\|_{\mathrm{F}}^2\mathrm{d}x\mathrm{d}y+\frac{\gamma_4}{2}\int_\Omega(u-f-\bar{v}^{n+\frac{1}{2}})^2\mathrm{d}x\mathrm{d}y\Big\}, \\ \mathbf{p}^{n+1}=\nabla u^{n+1}, \\ \mathbf{H}^{n+1}=\nabla^2u^{n+1}, \\ v^{n+1}=u^{n+1}-f, \end{cases}$$

and the u-subproblem has the following optimality condition

$$((\gamma_1+\gamma_4)\mathrm{Id}-\gamma_2\Delta+\gamma_3\Delta^2)u^{n+1}=\gamma_1\bar{u}^{n+\frac{1}{2}}-\gamma_2\operatorname{div}\bar{\mathbf{p}}^{n+\frac{1}{2}}+\gamma_3\operatorname{div}^2\bar{\mathbf{H}}^{n+\frac{1}{2}}+\gamma_4(f+\bar{v}^{n+\frac{1}{2}}), \tag{4.23}$$

which again can be solved by FFT under the periodic boundary conditions. These discussions give an accelerated operator-splitting method for the model (2.1) with $p=1$, which is summarized in Algorithm 2.

Algorithm 2 Accelerated operator-splitting algorithm to solve the model (2.1) (AOS-TSGV) in the case of $p=1$

```
1. Input: The noisy image $f$, parameters $\lambda,\gamma_1,\gamma_2,\gamma_3,\gamma_4,t$, initial value $(u_0,\mathbf{p}_0,\mathbf{H}_0,v_0)$, and function $\psi$;
2. For n=0,1,2,...:
    2.1 From $n$ to $n+\frac{1}{2}$: set $(\bar{u}^n,\bar{\mathbf{p}}^n,\bar{\mathbf{H}}^n,\bar{v}^n)$ by (4.18), and compute $(u^{n+\frac{1}{2}},\mathbf{p}^{n+\frac{1}{2}},\mathbf{H}^{n+\frac{1}{2}},v^{n+\frac{1}{2}})$ by (4.21) (4.11) (4.12) (4.22);
    2.2 From $n+\frac{1}{2}$ to $n+1$: set $(\bar{u}^{n+\frac{1}{2}},\bar{\mathbf{p}}^{n+\frac{1}{2}},\bar{\mathbf{H}}^{n+\frac{1}{2}},\bar{v}^{n+\frac{1}{2}})$ by (4.20), and compute $(u^{n+1},\mathbf{p}^{n+1},\mathbf{H}^{n+1},v^{n+1})=(u^{n+1},\nabla u^{n+1},\nabla^2u^{n+1},u^{n+1}-f)$ by (4.23).
```

Remark 4.1. Algorithms 1 and 2 (AOS-TSGV) are made up of some simple calculations like soft-shrinkage and FFT. Thus they are easy to be implemented. It looks that AOS-TSGV has multiple proximal parameters (γ_i, $i\in\{1,2,3,4\}$). Actually, they can be fixed in our experiments, like those time step size related parameters in basic operator-splitting method based on Lie scheme and Marchuk-Yanenko discretization [17, 21, 22, 28]. Our AOS-TSGV not only inherits the advantages of the basic operator-splitting method on less tuning parameters and stable performances, but also can speed it up quite a lot. See Section 5. A drawback of our AOS method is that the theoretical convergence analysis is currently unclear, like the basic operator-splitting method based on Lie scheme and Marchuk-Yanenko discretization without acceleration [17, 21, 22, 28]. However, such operator-splitting methods can be regarded as some numerical implementations of subgradient descent flow, and at least the sequence of objective values can be expected to be convergent by using small time step sizes (or equivalently, large proximal parameters). In Section 5, we will validate the numerical convergence of Algorithms 1 and 2.

4.3. Spatial discretization. Let the image domain Ω be discretized as $M\times N$ grids with spatial step size $\delta>0$. For simplicity, we denote by $u_{i,j}$ the approximate value of u at $(i\delta,j\delta),1\le i\le M,1\le j\le N$. Therefore a digital image ($u(x,y)$ after discretization) can be regarded as an element of $V=\mathbb{R}^{M\times N}$. For a given digital image $u_{i,j},1\le i\le M,1\le j\le N$, we first define, under the periodic boundary condition, the forward(+) and backward(-) finite differences by

$$(\mathring{\partial}_x^+ u)_{i,j}=\begin{cases}(u_{i,j+1}-u_{i,j})/\delta, & 1\le j<N,\\ (u_{i,1}-u_{i,N})/\delta, & j=N;\end{cases}\qquad (\mathring{\partial}_x^- u)_{i,j}=\begin{cases}(u_{i,j}-u_{i,j-1})/\delta, & 1<j\le N,\\ (u_{i,1}-u_{i,N})/\delta, & j=1;\end{cases}$$

$$(\mathring{\partial}_y^+ u)_{i,j}=\begin{cases}(u_{i+1,j}-u_{i,j})/\delta, & 1\le i<M,\\ (u_{1,j}-u_{M,j})/\delta, & i=M;\end{cases}\qquad (\mathring{\partial}_y^- u)_{i,j}=\begin{cases}(u_{i,j}-u_{i-1,j})/\delta, & 1<i\le M,\\ (u_{1,j}-u_{M,j})/\delta, & i=1.\end{cases}$$

With the above notation, the discrete gradient ∇u and Hessian $\nabla^2 u$ of a digital image $u\in V$ are calculated as, for $1\le i\le M,1\le j\le N$,

$$(\nabla u)_{i,j}=\Big((\mathring{\partial}_x^+ u)_{i,j},(\mathring{\partial}_y^+ u)_{i,j}\Big)^\top,\ \text{and}\ (\nabla^2 u)_{i,j}=\begin{pmatrix}(\mathring{\partial}_x^-(\mathring{\partial}_x^+ u))_{i,j} & (\mathring{\partial}_x^+(\mathring{\partial}_y^+ u))_{i,j}\\ (\mathring{\partial}_y^+(\mathring{\partial}_x^+ u))_{i,j} & (\mathring{\partial}_y^-(\mathring{\partial}_y^+ u))_{i,j}\end{pmatrix}.$$

For $\mathbf{p}=(p^1,p^2)^\top,p^1,p^2\in V$, its discrete divergence is defined by

$$(\operatorname{div}\mathbf{p})_{i,j}=(\mathring{\partial}_x^- p^1)_{i,j}+(\mathring{\partial}_y^- p^2)_{i,j},1\le i\le M,1\le j\le N,$$

with div as the adjoint of $-\nabla$ under the standard inner products in V and $V\times V$. For $\mathbf{H}=\begin{pmatrix}H^{11} & H^{12}\\ H^{21} & H^{22}\end{pmatrix}$ with $H^{kl}\in V,k,l\in\{1,2\}$, we calculate, for $1\le i\le M,1\le j\le N$, that

$$(\operatorname{div}^2\mathbf{H})_{i,j}=(\mathring{\partial}_x^+(\mathring{\partial}_x^- H^{11}))_{i,j}+(\mathring{\partial}_y^-(\mathring{\partial}_x^- H^{12}))_{i,j}+(\mathring{\partial}_x^-(\mathring{\partial}_y^- H^{21}))_{i,j}+(\mathring{\partial}_y^+(\mathring{\partial}_y^- H^{22}))_{i,j},$$

with div^2 as the adjoint of ∇^2 under the standard inner products in V and $V\times V\times V\times V$. See, e.g., [53, 58] for more details.

5. Numerical experiments

In this section, we provide some numerical experiments to demonstrate the effectiveness and efficiency of our method. We will compare our Algorithm 1 (AOS-TSGV in the case of $p=2$) to several typical or latest existing variational methods. We will also report the experimental results of our Algorithm 2 (AOS-TSGV in the case of $p=1$). All the experiments were performed on a laptop with Windows 7, 32-bit, 4 GB RAM, Intel Core i3-2348M CPU (2.30GHz) and Matlab R2012b.

5.1. Experimental settings. In our experiments, we test two synthetic images ("Disk" and "Square") and six natural images ("Peppers", "Lena", "Cameraman", "Barbara", "Boat", and "Goldhill"). All the test images are with 256×256 resolution. We use the Matlab built-in function *imnoise* to add the Gaussian noise with variance 0.005 and 10% impulsive noise to synthetic images, and the Gaussian noise with variance 0.01 and 20% impulsive noise to natural images, respectively. The clean and noisy images are all displayed in Figure 5.1.

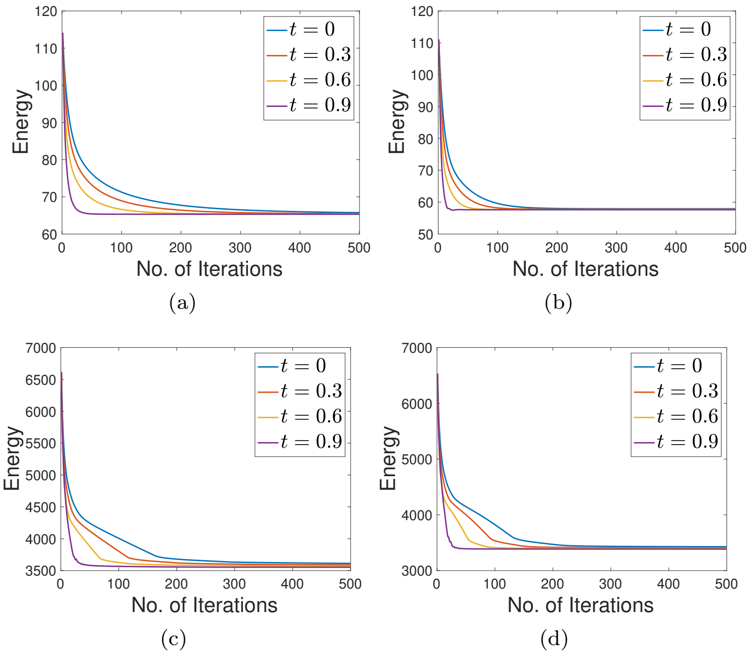
FIG. 5.1. *The test images used in this paper. The first row: the clean images. The second row: the observations with Gaussian noise. The third row: the observations with impulsive noise.*

We use the initialization method as in [28]. In particular, we compute $u_0=(\mathrm{Id}-\Delta)^{-1}f$ and set the algorithm initial value for Algorithm 1 as $(u^0,\mathbf{p}^0,\mathbf{H}^0)=(u_0,\nabla u_0,\nabla^2 u_0)$, and for Algorithm 2 as $(u^0,\mathbf{p}^0,\mathbf{H}^0,v^0)=(u_0,\nabla u_0,\nabla^2 u_0,u_0-f)$. The stopping criterion $\frac{|E(u^{n+1})-E(u^n)|}{|E(u^{n+1})|}<\mathrm{Tol}=10^{-4}$ based on energy values is used for both Algorithm 1 and Algorithm 2, unless otherwise stated. The spatial step size δ is just set as 1. We set $\gamma_1=\gamma_2=\gamma_3=5$ for Algorithm 1 (AOS-TSGV for Gaussian noise removal), and $\gamma_1=\gamma_2=\gamma_3=\gamma_4=50$ for Algorithm 2 (AOS-TSGV for impulsive noise removal). Since there are a class of scaling functions ψ satisfying Assumptions A1-A3, i.e. $\psi(s)=\frac{1}{(b+s^k)^{1/k}}$ with $b>0$ and $k\geq 1$, we choose four of them $\psi_1(s)=\frac{1}{1+s}$, $\psi_2(s)=\frac{1}{0.1+s}$, $\psi_3(s)=\frac{1}{\sqrt{1+s^2}}$, $\psi_4(s)=\frac{1}{\sqrt{0.1+s^2}}$ to test our methods.

For all the compared methods, we use the peak signal-to-noise ratio (PSNR, in dB), running time (Time, in seconds) and/or the number of iterations (#It) to evaluate their performances. To better display residual images, we use $f-u+0.5$ for the Gaussian noise removal problems, and $u-\underline{u}+0.5$ for the impulsive noise removal problems, by considering the statistical characteristics of different kinds of noises. Note that the intensity values of f and algorithm output u are all rescaled into $[0,1]$, if needed.

5.2. Numerical energy convergence and acceleration effect by the extrapolation parameter t. In this subsection, we test the numerical energy convergence of Algorithm 1 (AOS-TSGV for Gaussian noise removal) and Algorithm 2 (AOS-TSGV for impulsive noise removal) and acceleration effect by the extrapolation parameter t. The parameter t is set among $\{0,0.3,0.6,0.9\}$, where $t=0$ means the original basic operator-splitting method without any acceleration. We use "Disk" and "Square" as test images.

To see the numerical energy convergence, we force Algorithms 1 and 2 to run 500 iterations with different t. Therein, the scaling function ψ is set as ψ_1 for "Disk" and Gaussian noise removal, ψ_2 for "Square" and Gaussian noise removal, ψ_3 for "Disk" and impulsive noise removal, and ψ_4 for "Square" and impulsive noise removal. Figure 5.2 shows the curves of the objective energy versus the number of iterations. We see that for each test, the sequences of energy values by Algorithms 1 or 2 with different $t\in\{0,0.3,0.6,0.9\}$ converge to almost the same value. The convergent speed becomes faster as t increases.



(a) (b) (c) (d)

FIG. 5.2. *Objective energy values versus number of iterations. (a) Algorithm 1 with ψ_1 on "Disk"; (b) Algorithm 1 with ψ_2 on "Square"; (c) Algorithm 2 with ψ_3 on "Disk"; (d) Algorithm 2 with ψ_4 on "Square".*

To further investigate the acceleration effect by the extrapolation parameter t in terms of both effectiveness and efficiency, we use the default stopping criterion mentioned before and do more tests. In particular, we apply Algorithms 1 and 2 with $\psi_1,\psi_2,\psi_3,\psi_4$ to the denoising problems on "Disk" and "Square". The PSNR value and number of iterations by Algorithms 1 and 2 are listed in Tables 5.1 and 5.2, respectively. We can see that, the number of iterations consistently decreases when t increases, and Algorithms 1 and 2 with $t=0.9$ give the best PSNR values among those with $t\in\{0,0.3,0.6,0.9\}$. This illustrates that our AOS-TSGV in the cases of $p=2$ and $p=1$ with $t=0.9$ achieves good performances with high efficiency. Therefore, in the rest of our experiments, we set $t=0.9$ for Algorithms 1 and 2.

5.3. Tests on numerical edge and corner contrast preservation properties by AOS-TSGV. Theorems 3.1 and 3.2 show that the model (2.1) with suitable ψ and $p\in[1,\infty)$ can preserve edge and corner contrasts for "Disk" and "Square" type of images. In this subsection, we investigate whether Algorithms 1 and 2 possess the ability to preserve edge and corner contrasts numerically. To mimic Theorems 3.1 and 3.2, we use clean "Disk" and "Square" as the algorithm inputs. Both Algorithms 1 and 2 with $\psi_1,\psi_2,\psi_3,\psi_4$ are tested. The algorithm outputs and residuals are displayed in Figure 5.3. We can see that the restored images are visually identical to the clean inputs, which validate that our Algorithms 1 and 2 can preserve edge and corner contrast numerically.

5.4. Test on the influence of the tuning parameter λ in AOS-TSGV. Since the tuning parameter λ is a trade-off between the fidelity term and the regularization term, it affects the denoising results. To test the sensitivity of our AOS-TSGV with

	Disk								Square							
	ψ_1		ψ_2		ψ_3		ψ_4		ψ_1		ψ_2		ψ_3		ψ_4	
t	psnr	#It	psnr	#It	psnr	#It	psnr	#It	psnr	#It	psnr	#It	psnr	#It	psnr	#It
0	29.89	316	30.18	318	29.73	303	30.04	335	30.10	206	30.24	173	30.03	209	30.25	184
0.3	29.97	245	30.23	240	29.82	238	30.11	257	30.16	153	30.28	127	30.10	156	30.31	136
0.6	30.05	164	30.29	155	29.91	162	30.18	171	30.21	97	30.33	79	30.15	100	30.35	85
0.9	30.15	53	30.36	47	30.03	54	30.28	54	30.34	27	30.45	24	30.29	28	30.51	23

TABLE 5.1. *The PSNR values (dB) and number of iterations of Algorithm 1 with different ψ and extrapolation parameter $t \in \{0, 0.3, 0.6, 0.9\}$ for Gaussian noise removal of "Disk" and "Square".*

	Disk								Square							
	ψ_1		ψ_2		ψ_3		ψ_4		ψ_1		ψ_2		ψ_3		ψ_4	
t	psnr	#It	psnr	#It	psnr	#It	psnr	#It	psnr	#It	psnr	#It	psnr	#It	psnr	#It
0	31.88	327	32.25	200	31.11	290	32.24	252	35.12	281	35.40	207	34.46	302	35.95	231
0.3	31.98	235	32.67	162	31.18	208	32.53	197	35.51	211	35.65	157	34.68	221	36.18	168
0.6	32.18	146	33.81	160	31.36	130	33.22	152	35.91	131	35.88	98	35.08	139	36.78	110
0.9	32.94	57	34.06	64	32.13	57	33.98	76	37.33	49	36.97	52	36.50	57	38.24	46

TABLE 5.2. *The PSNR values (dB) and number of iterations of Algorithm 2 with different ψ and extrapolation parameter $t \in \{0, 0.3, 0.6, 0.9\}$ for impulsive noise removal of "Disk" and "Square".*

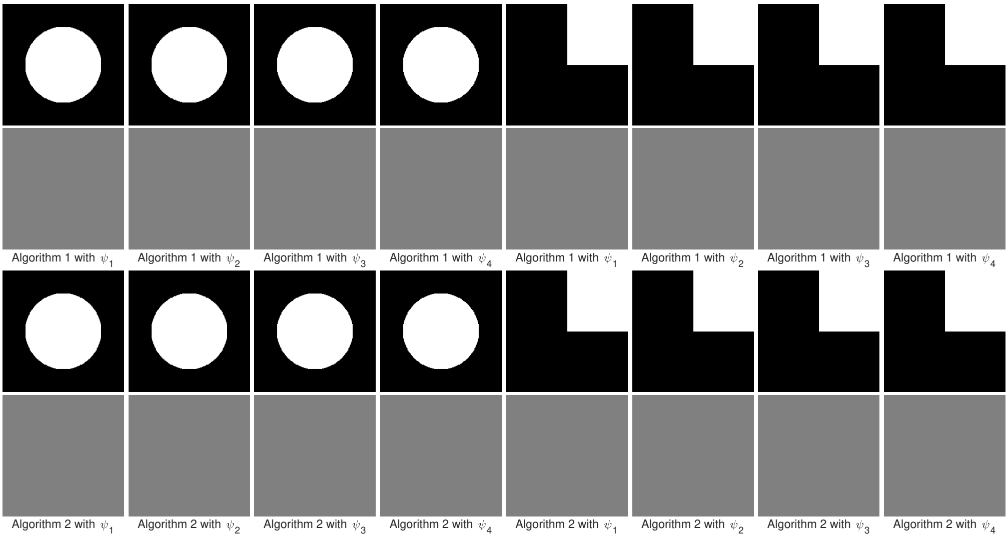


FIG. 5.3. *Tests on edge and corner contrast preservation of AOS-TSGV (Algorithms 1 and 2) with $\psi_1, \psi_2, \psi_3, \psi_4$ scaling functions, by using clean "Disk" and "Square" as inputs. The first and third rows give the algorithm outputs. The second and fourth rows give the residual images, shown in the way described in Section 5.1.*

respect to the tuning parameter λ, we run AOS-TSGV on all six natural images and record PSNR values of the denoising results by using different λ values. Figure 5.4 displays PSNR vs λ curves of AOS-TSGV with $p = 2$ and ψ_1, ψ_2 functions (the first two rows), as well as PSNR vs λ curves of AOS-TSGV with $p = 1$ and ψ_3, ψ_4 functions (the last two rows). From Figure 5.4, we have two observations in each example: first, the proposed algorithm can produce good PSNR when λ belongs to a relatively large range; second, the variation of PNSR with respect to λ is almost smooth.

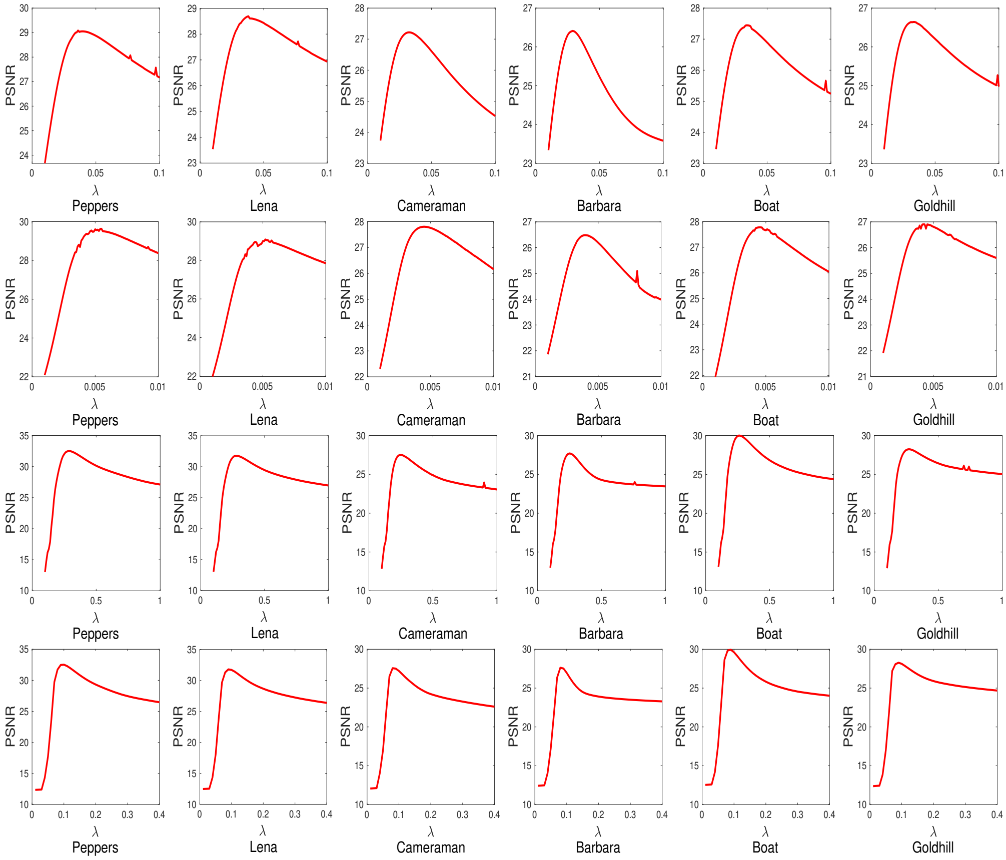


FIG. 5.4. *Test on the influence of the tuning parameter* λ *in AOS-TSGV: PSNR vs* λ *by* ψ_1, ψ_2 *for Gaussian noise removal* ($p=2$) *shown in the first two rows, and PSNR vs* λ *by* ψ_3, ψ_4 *for impulsive noise removal* ($p=1$) *shown in the last two rows.*

5.5. Comparisons of AOS-TSGV in the case of $p=2$ (Algorithm 1) to several typical or latest variational denoising methods. In this subsection, we compare Algorithm 1 with ψ_2 scaling function to several typical or latest variational denoising methods, including

- ALM-TV, i.e., an augmented Lagrangian based method [53] for the TV model [42];
- ALM-LLT, i.e., an augmented Lagrangian based method [53] for the LLT model [29];
- PD-TGV, i.e., a primal-dual method [27] for the TGV model [4];
- ALM-MC, i.e., an augmented Lagrangian based method [58] for the MC model [57];
- OS-EE, i.e., an operator-splitting method for the EE model [17];
- OS-GC, i.e., an operator-splitting method for the GC model [28];
- SA-TV-TV^2, i.e., an alternating direction method of multipliers for the reformulated spatially adapted first and second order variational model [56].

Note that we choose SA-TV-TV^2 instead of the WM model from [56] for comparison, because SA-TV-TV^2 has been demonstrated with better performances than the WM model in [56]. We use their default stopping criterions, except PD-TGV. For PD-TGV, since [27] did not point out the stopping criterion, we terminate this algorithm when it satisfies $\frac{|u^{n+1}-u^n|}{|u^n|} < 10^{-5}$ or reaches 1000 iterations. Most algorithm parameters of these methods are also set as default values in their codes or papers. For each of these methods, we tune the model parameter(s) and important algorithm parameter(s) manually for each test image. The numbers of tuning parameters of all compared methods are listed in Table 5.3. Benefiting from the basic operator-splitting method with Lie scheme and Marchuk-Yanenko discretization, our AOS-TSGV has only one tuning parameter, which is the least among these methods.

	ALM-TV	ALM-LLT	PD-TGV	ALM-MC	OS-EE	OS-GC	SA-TV-TV^2	AOS-TSGV
Model para.(s)	1	1	2	1	2	2	1	1
Algorithm para.(s)	1	1	0	4	0	0	2	0

Table 5.3. *The number of tuning parameters (tuned manually for each test image) of different methods for Gaussian noise removal problems.*

We first compare the computational effectiveness between these variational methods. In Figure 5.5, we show the denoising results and residuals with respect to the noisy observations for "Lena" and "Cameraman". We can see that for these two test images, all compared methods can effectively remove the Gaussian noise. If zooming in on the details, one can see a little staircase effect in the ALM-TV results, and the edges retained in the residual images by ALM-MC, SA-TV-TV^2 and AOS-TSGV are a little weaker than others. For quantitative comparisons, in Table 5.4, we list the PSNR values of denoising results by compared methods for all test images. We can see, overall speaking, that Algorithm 1 and SA-TV-TV^2 perform better than ALM-TV, ALM-LLT, PD-TGV, ALM-MC, OS-EE and OS-GC; and furthermore, the denoising results by our Algorithm 1 have slightly higher PSNR values than SA-TV-TV^2. Here, the model parameter λ in our method was set 5.5×10^{-3} for "Peppers", 5.2×10^{-3} for "Lena", 4.5×10^{-3} for "Cameraman", 3.9×10^{-3} for "Barbara", 4.5×10^{-3} for "Boat", and 4.2×10^{-3} for "Goldhill", respectively.

	ALM-TV		ALM-LLT		PD-TGV		ALM-MC		OS-EE		OS-GC		SA-TV-TV^2		AOS-TSGV	
	psnr	time	psnr	time	psnr	time	psnr	time	psnr	time	psnr	time	psnr	time	psnr	time
Peppers	28.66	0.3	28.94	3.8	29.21	4.7	29.27	35.3	29.05	107.9	28.90	446.7	$\underline{29.60}$	18.6	**29.63**	0.7
Lena	28.46	0.3	28.53	4.9	28.74	4.8	28.56	32.1	28.78	94.9	28.69	430.0	$\underline{29.05}$	18.8	**29.08**	0.8
Cameraman	27.33	0.3	27.06	4.6	27.37	12.4	27.66	42.2	27.60	58.6	27.15	227.9	**27.94**	18.7	$\underline{27.80}$	1.7
Barbara	25.94	0.3	26.30	2.2	26.30	4.3	25.76	32.4	26.12	101.4	$\underline{26.47}$	444.9	26.43	18.6	**26.48**	1.8
Boat	27.20	0.3	27.29	2.5	27.32	3.8	27.36	35.0	27.38	82.5	27.29	495.5	$\underline{27.73}$	18.6	**27.78**	0.8
Goldhill	26.74	0.3	26.58	4.5	26.82	10.1	26.70	40.3	26.84	72.4	$\underline{26.86}$	216.5	26.76	18.8	**26.90**	0.9

Table 5.4. *The PSNR values (dB) of denoising results and running time (in seconds) of compared methods for Gaussian noise removal of all natural test images. For each test, the highest PSNR value is shown in bold and the second highest is underlined.*

We next compare the computational efficiency between these methods. Table 5.4 records the running time of compared methods for all natural test images. As the only first-order method, ALM-TV is faster than all the others. However, among all the tested high-order methods, our Algorithm 1 (AOS-TSGV in the case of $p=2$) is the fastest one. It is even faster than the two convex high-order methods, i.e., ALM-LLT and PD-TGV. Compared to the nonconvex methods, i.e., ALM-MC, OS-EE, OS-GC and SA-TV-TV^2,

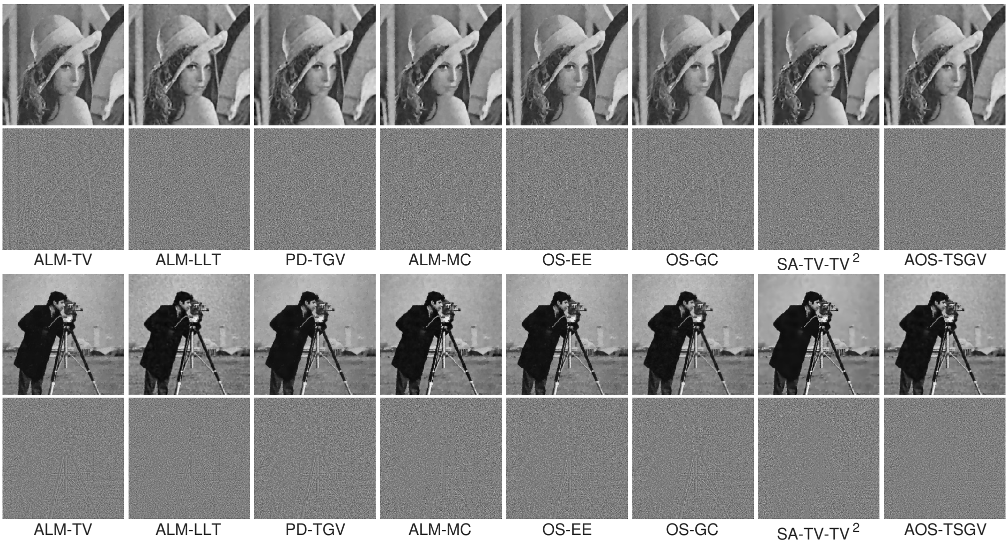


FIG. 5.5. *Gaussian noise removal results of compared methods for "Lena" and "Cameraman". The first and third rows: the recovered images. The second and fourth rows: the residual images shown in the way described in Section 5.1.*

this advantage of computational efficiency of our Algorithm 1 is more obvious (at least ten times faster than them).

5.6. Numerical tests of AOS-TSGV in the case of $p=1$ (Algorithm 2). In this subsection, to save space, we briefly report the numerical tests about AOS-TSGV in the case of $p=1$ (Algorithm 2) with ψ_4 scaling function. In Figure 5.6, we show the denoising results and residuals for all natural test images. This demonstrates that our AOS-TSGV in the case of $p=1$ (Algorithm 2) can effectively remove impulsive noise. Only little edge information is left in the residual images $u-\underline{u}+0.5$. The quantitative evaluations in terms of PSNR values and running time are recorded in Table 5.5.

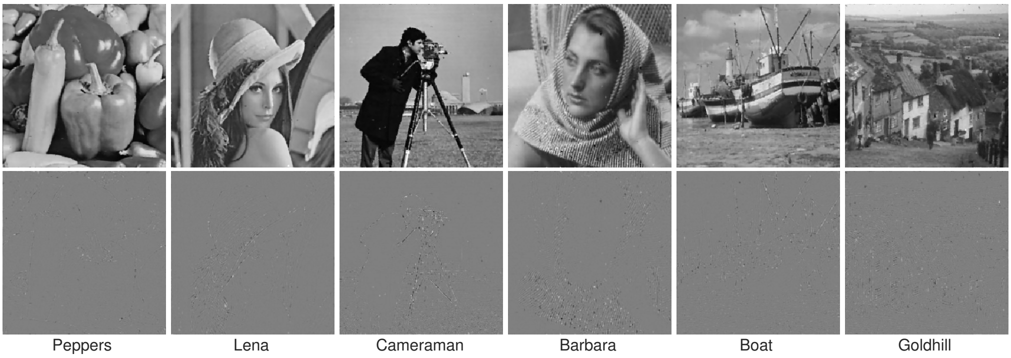


FIG. 5.6. *Impulsive noise removal results by Algorithm 2 (AOS-TSGV in the case of $p=1$). The first row: the recovered images. The second row: the residual images shown in the way described in Section 5.1.*

Peppers		Lena		Cameraman		Barbara		Boat		Goldhill	
psnr	time	psnr	time	psnr	time	psnr	time	psnr	time	psnr	time
32.55	2.4	31.81	2.4	27.59	3.6	27.67	3.4	29.98	2.6	28.27	2.5

Table 5.5. *The PSNR values (dB) of denoising results and running time (in seconds) of Algorithm 2 (AOS-TSGV in the case of $p=1$) for impulsive noise removal of all natural test images.*

6. Conclusion

In this paper, motivated by sparse modeling of piecewise-linear image structures, we proposed a total scaled-gradient variation regularization framework for image denoising problem. This regularization framework was combined with $\mathscr{L}^p$-based fidelity term ($p\in[1,+\infty)$) to construct a general high-order variational model. Under proper assumptions on the scaling function ψ, we established two theorems showing the edge and corner contrast preservation properties of the model. To solve the proposed nonconvex and nonsmooth model, we devised an accelerated operator-splitting method by introducing the inertial extrapolation and some approximation techniques. As far as we know, such inertial acceleration had not been combined with the basic operator-splitting method based on Lie scheme and Marchuk-Yanenko discretization in solving image denoising models. Comparison experiments showed that our method with less tuning parameters achieved best PSNR values in most tests, and it was almost ten times faster than the methods giving the second best PSNR values. A future work is to theoretically study the convergence of the proposed algorithms or design other efficient and convergent algorithms for our TSGV-based models.

Acknowledgments. We would like to thank Prof. Liangjian Deng, Prof. Yuping Duan, Dr. Yiming Gao, Prof. Hao Liu, and Prof. Wei Zhu for providing us related Matlab codes or making the codes available online. The work of the first author was supported by the National Natural Science Foundation of China (Nos. 12271273 and 11871035) and a Key Program (No. 21JCZDJC00220) of Natural Science Foundation of Tianjin, China. The work of the third author was supported by National Natural Science Foundation of China (No. 12201320) and the Fundamental Research Funds for the Central Universities, Nankai University (Nos. 63221039 and 63231144).

REFERENCES

[1] F. Alvarez and H. Attouch, *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping*, Set-Valued Analysis, 9(1):3–11, 2001. 4.1, 4.1

[2] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, New York: Springer, 147, 2006. 1, 1

[3] G. Bellettini, V. Caselles, and M. Novaga, *The total variation flow in $\mathbb{R}^n$*, J. Differ. Equ., 184(2):475–525, 2002. 1

[4] K. Bredies, K. Kunisch, and T. Pock, *Total generalized variation*, SIAM J. Imag. Sci., 3(3):492–526, 2010. 1, 1, 1, 5.5

[5] K. Bredies and T. Valkonen, *Inverse problems with second-order total generalized variation constraints*, Proceedings of SampTA 2011 9th International Conference on Sampling Theory and Applications, Singapore, 2011. 1, 1

[6] C. Brito-Loeza, K. Chen, and V. Uc-Cetina, *Image denoising using the Gaussian curvature of the image surface*, Numer. Meth. Partial Differ. Equ., 32(3):1066–1089, 2016. 1, 1, 3, 3, 3

[7] A. Buades, B. Coll, and J.-M. Morel, *A review of image denoising algorithms, with a new one*, Multiscale Model. Simul., 4(2):490–530, 2005. 1

[8] J.-F. Cai, B. Dong, S. Osher, and Z. Shen, *Image restoration: total variation, wavelet frames, and beyond*, J. Amer. Math. Soc., 25(4):1033–1089, 2012. 1

[9] E.J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction*

from highly incomplete frequency information, IEEE Trans. Inf. Theory, 52(2):489–509, 2006. 1

[10] E.J. Candès, M.B. Wakin, and S.P. Boyd, *Enhancing sparsity by reweighted ℓ_1 minimization*, J. Fourier Anal. Appl., 14:877–905, 2008. 4.1, 4.1, 4.1

[11] A. Chambolle and P.-L. Lions, *Image recovery via total variation minimization and related problems*, Numer. Math., 76(2):167–188, 1997. 1, 1, 1

[12] A. Chambolle and T. Pock, *An introduction to continuous optimization for imaging*, Acta Numer., 25:161–319, 2016. 1

[13] T. Chan, A. Marquina, and P. Mulet, *High-order total variation-based image restoration*, SIAM J. Sci. Comput., 22(2):503–516, 2000. 1

[14] T.F. Chan and J. Shen, *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*, SIAM, 2005. 1

[15] K. Chen, C.-B. Schönlieb, X.-C. Tai, and Y. Laurent, *Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision*, Springer, 2021. 1, 1

[16] C.-A. Deledalle, N. Papadakis, and J. Salmon, *On debiasing restoration algorithms: Applications to total-variation and nonlocal-means*, in J.F. Aujol, M. Nikolova, and N. Papadakis (eds.), Scale Space and Variational Methods in Computer Vision, Springer, 129–141, 2015. 1, 3

[17] L.-J. Deng, R. Glowinski, and X.-C. Tai, *A new operator splitting method for the Euler elastica model for image smoothing*, SIAM J. Imag. Sci., 12(2):1190–1230, 2019. 1, 1, 4, 4.1, 4.1, 4.1, 5.5

[18] D.C. Dobson and F. Santosa, *Recovery of blocky images from noisy and blurred data*, SIAM J. Appl. Math., 56(4):1181–1198, 1996. 1

[19] D.L. Donoho, *Compressed sensing*, IEEE Trans. Inf. Theory, 52(4):1289–1306, 2006. 1

[20] D.L. Donoho and I.M. Johnstone, *Adapting to unknown smoothness via wavelet shrinkage*, J. Amer. Stat. Assoc., 90(432):1200–1224, 1995. 1

[21] Y. Duan, Q. Zhong, X.-C. Tai, and R. Glowinski, *A fast operator-splitting method for Beltrami color image denoising*, J. Sci. Comput., 92(3):89, 2022. 4, 4.1

[22] R. Glowinski, S. Luo, and X.-C. Tai, *Fast operator-splitting algorithms for variational imaging models: Some recent developments*, in R. Kimmel and X.-C. Tai (eds.), Handbook of Numerical Analysis, Elsevier, 20:191–232, 2019. 1, 4, 4.1, 4.1

[23] R. Glowinski, S.J. Osher, and W. Yin, *Splitting Methods in Communication, Imaging, Science, and Engineering*, Springer, 2016. 1

[24] R. Glowinski, T.-W. Pan, and X.-C. Tai, *Some facts about operator-splitting and alternating direction methods*, in R. Glowinski, S. Osher, and W. Yin (eds.), Splitting Methods in Communication, Imaging, Science, and Engineering, Springer International Publishing, 19–94, 2016. 1, 4.1

[25] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, Pearson Education India, 2009. 1

[26] W. Hinterberger and O. Scherzer, *Variational methods on the space of functions of bounded Hessian for convexification and denoising*, Computing, 76:109–133, 2006. 1

[27] F. Knoll, K. Bredies, T. Pock, and R. Stollberger, *Second order total generalized variation (TGV) for MRI*, Magn. Reson. Med., 65(2):480–491, 2011. 5.5

[28] H. Liu, X.-C. Tai, and R. Glowinski, *An operator-splitting method for the Gaussian curvature regularization model with applications to surface smoothing and imaging*, SIAM J. Sci. Comput., 44(2):A935–A963, 2022. 1, 4, 4.1, 4.1, 5.1, 5.5

[29] M. Lysaker, A. Lundervold, and X.-C. 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(12):1579–1590, 2003. 1, 1, 2.2, 5.5

[30] S. Ma, J. Yang, and C. Chen, *A general inertial proximal point algorithm for mixed variational inequality problem*, SIAM J. Optim., 25(4):2120–2142, 2015. 4.1, 4.1

[31] S. Masnou, *Disocclusion: a variational approach using level lines*, IEEE Trans. Image Process., 11(2):68–76, 2002. 1

[32] S. Masnou and J.M. Morel, *Level lines based disocclusion*, Proceedings 1998 International Conference on Image Processing, IEEE, 259–263, 1998. 1

[33] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*, Amer. Math. Soc., 22, 2001. 1, 3

[34] L. Moisan, *Periodic plus smooth image decomposition*, J. Math. Imag. Vis., 39:161–179, 2011. 4.1

[35] D.B. Mumford and J. Shah, *Optimal approximations by piecewise smooth functions and associated variational problems*, Commun. Pure Appl. Math., 42(5):577–685, 1989. 2

[36] M. Nitzberg, D. Mumford, and T. Shiota, *Filtering, Segmentation and Depth*, Springer, 662, 1993. 1

[37] P. Ochs, Y. Chen, T. Brox, and T. Pock, *iPiano: Inertial proximal algorithm for nonconvex*

optimizationm for mixed variational inequality problem, SIAM J. Imag. Sci., 7(2):1388–1419, 2014. 4.1, 4.1

[38] K. Papafitsoros and C.-B. Schönlieb, *A combined first and second order variational approach for image reconstruction*, J. Math. Imag. Vis., 48(2):308–338, 2014. 1

[39] N. Paragios, Y. Chen, and O.D. Faugeras, *Handbook of Mathematical Models in Computer Vision*, Springer Science & Business Media, 2006. 1

[40] P. Penrona and J. Malik, *Scale space and edge detection using anisotropic diffusion*, IEEE Trans. Pattern Anal. Mach. Intell., 12(4):629–639, 1990. 1

[41] B.T. Polyak, *Some methods of speeding up the convergence of iteration methods*, USSR Comput. Math. Math. Phys., 4(5):1–17, 1964. 4.1, 4.1

[42] L.I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60(1-4):259–268, 1992. 1, 5.5

[43] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational Methods in Imaging*, Springer, 2009. 1

[44] F. Sciacchitano, Y. Dong, and T. Zeng, *Variational approach for restoring blurred images with cauchy noise*, SIAM J. Imag. Sci., 8(3):1894–1922, 2015. 1

[45] J. Shen, S.H Kang, and T.F. Chan, *Euler's elastica and curvature-based inpainting*, SIAM J. Appl. Math., 63(2):564–592, 2003. 1, 1

[46] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk, *On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides*, SIAM J. Numer. Anal., 42(2):686–713, 2004. 1

[47] D. Strong and T. Chan, *Edge-preserving and scale-dependent properties of total variation regularization*, Inverse Probl., 19(6):S165–S187, 2003. 1

[48] X.-C. Tai, J. Hahn, and G.J. Chung, *A fast algorithm for Euler's elastica model using augmented Lagrangian method*, SIAM J. Imag. Sci., 4(1):313–344, 2011. 1, 1

[49] C. Tomasi and R. Manduchi, *Bilateral filtering for gray and color images*, Sixth International Conference on Computer Vision, IEEE, 839–846, 1998. 1

[50] Y. Wang, J. Yang, W. Yin, and Y. Zhang, *A new alternating minimization algorithm for total variation image reconstruction*, SIAM J. Imag. Sci., 1(3):248–272, 2008. 4.1, 4.1

[51] J.B. Weaver, Y. Xu, D.M. Healy Jr., and L.D. Cromwell, *Filtering noise from images with wavelet transforms*, Magn. Reson. Med., 21(2):288–295, 1991. 1

[52] J. Weickert, *Anisotropic Diffusion in Image Processing*, Teubner Stuttgart, 1998. 1

[53] C. Wu and X.-C. Tai, *Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models*, SIAM J. Imag. Sci., 3(3):300–339, 2010. 1, 4.1, 4.3, 5.5

[54] M. Yashtini and S.H. Kang, *A fast relaxed normal two split method and an effective weighted TV approach for Euler's elastica image inpainting*, SIAM J. Imag. Sci., 9(4):1552–1581, 2016. 1, 4.1, 4.1, 4.1

[55] Y.-L. You and M. Kaveh, *Fourth-order partial differential equations for noise removal*, IEEE Trans. Image Process., 9(10):1723–1730, 2000. 1, 1, 2.2

[56] Q. Zhong, R.W. Liu, and Y. Duan, *Spatially adapted first and second order regularization for image reconstruction: from an image surface perspective*, J. Sci. Comput., 92(2):33, 2022. 1, 1, 2.1, 2.2, 3, 3, 3, 3.2, 4.1, 4.1, 4.1, 5.5

[57] W. Zhu and T. Chan, *Image denoising using mean curvature of image surface*, SIAM J. Imag. Sci., 5(1):1–32, 2012. 1, 1, 1, 2.1, 2.2, 3, 3, 3, 3.2, 5.5

[58] W. Zhu, X.-C. Tai, and T. Chan, *Augmented Lagrangian method for a mean curvature based image denoising model*, Inverse Probl. Imag., 7(4):1409–1432, 2013. 1, 4.3, 5.5