




A first-order image denoising model for staircase reduction

Wei Zhu¹ 

Received: 16 October 2018 / Accepted: 23 September 2019 /
Published online: 29 November 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

In this paper, we consider a total variation–based image denoising model that is able to alleviate the well-known staircasing phenomenon possessed by the Rudin–Osher–Fatemi model (Rudin et al., *Phys. D* **60**, 259–268, 1992). To minimize this variational model, we employ augmented Lagrangian method (ALM). Convergence analysis is established for the proposed algorithm. Numerical experiments are presented to demonstrate the features of the proposed model and also show the efficiency of the proposed numerical method.

Keywords Image denoising · Augmented Lagrangian method · Variational model

Mathematics Subject Classification (2010) 94A08 · 65K10 · 65M32

1 Introduction

Image denoising aims to remove the noise part of a given image in order to get a clean one inside which meaningful signals should be largely retained. This is a typical inverse problem, and to treat such a problem, an appropriate regularizer should often be employed. In the literature, one of the most famous variational models for this problem was given by Rudin, Osher, and Fatemi (ROF) [30], where the total variation was used as the regularizer. The appealing feature of this regularizer lies in its ability of allowing discontinuous solution of the variational model, and thus object boundary can be captured. Ever since this remarkable work, the total variation–based regularizer has been adopted to dealing with different imaging tasks [8, 14, 15, 17, 19]. Moreover, numerous variational models have been proposed by using different

Communicated by: Raymond H. Chan

✉ Wei Zhu
wzhu7@ua.edu

¹ Department of Mathematics, University of Alabama, Box 870350, Tuscaloosa, AL 35487, USA

forms of regularizers in the field of mathematical imaging [3, 16, 18, 23, 24, 27, 28].

Despite its effectiveness of removing noise while keeping object boundaries, the ROF model bears several undesirable properties. For instance, in [25], Meyer pointed out that the ROF model cannot preserve image contrast and object corners, and also in [7], Bellettini, Caselles, and Novaga studied what kind of shapes can be maintained by the ROF model, which suggests that the ROF model will smear object corners. More importantly, the ROF model also suffers from the staircase effect, that is, the denoised image presents blocky pieces or piecewise constant regions, even for originally smooth parts inside the given image, which surely leads to visually unpleasant denoised results.

To fix the staircasing phenomenon, quite a few of higher order variational models have been developed [9, 18, 23, 31, 35], which is due to the fact that piecewise constant functions could lead to large magnitude of their second-order derivatives. Precisely, in [9], Bredies et al. developed a novel term called total generalized variation as a regularizer for dealing with inverse problems in mathematical imaging, including image denoising. As this new regularizer utilizes higher order derivatives, the proposed model successfully suppress the staircase effect. In [18], Chan et al. proposed adding a nonlinear fourth-order diffusive term in order to reduce the staircase effect. In [23], instead of total variation, second-order derivatives were directly incorporated into the denoising model, and as a result, the blocky phenomenon can also be largely removed. In [31], Euler's elastica was used as the regularizer for the purpose of image denoising, and the staircase issue was also well solved. Moreover, in [35], the authors proposed using the L^1 -norm of mean curvature of image surface as the regularizer for the noise removal problem.

Even though all the above higher order models can successfully ameliorate the staircase effect, one has to deal with more complicated models than the ROF model. It is often challenging to develop efficient numerical methods for these higher order models, which has attracted lots of research attention during the past decade. Many efficient algorithms have been developed to handle these higher order, highly nonlinear, and/or non-differentiable models [4, 13, 20, 31, 34].

In this paper, we intend to develop a first-order image denoising model to reduce the staircase effect while removing noise and keeping object boundaries effectively. More precisely, just like ROF, only the first-order derivative will be employed in the model. As discussed above, the remarkable feature of the ROF model is its capability of preserving jumps, and this is mainly due to its total variation regularizer. Therefore, for the noise removal problem, we still want to keep this nice feature. However, this regularizer cannot avoid the staircase effect, and this issue reminds us that the total variation might not be appropriately applied for each region inside the given noisy image. Note that the staircase often occurs at locations with relatively small gradient of image intensity, since locations with large gradient usually indicate object boundaries. In some sense, the total variation is too weak to suppress the blocky effect in smooth regions. Therefore, to reduce the staircase phenomenon, a regularizer that requires higher regularity than the total variation should be imposed on such regions, i.e., the regions with relatively small gradient. Based on these arguments, we propose employing different types of regularizers for different regions of image

gradient. Specifically, we plan to apply L^p -norm of image gradient with $p > 1$ for regions with small gradient, while retaining the total variation for regions with large gradient. This choice could help alleviate the staircase effect. For instance, just as the celebrated Mumford-Shah model [26], if one chooses L^2 -norm of gradient as the regularizer, a smooth solution could be obtained. In this work, we only apply L^p -norm of image gradient with $p > 1$ for regions with small gradient in order to alleviate the staircase phenomenon.

In fact, similar idea was also proposed in other works [12, 32]. In [12], Chang and Che considered three total variation types of image denoising models, where the L^p -norm of the first-order partial derivatives with $p = 1/2, 1, 2$ were used as regularizers respectively, and then obtained the denoised image by taking a combination of the results from the three models in order to maintain the merits of each model. In [32], Vese studied a class of total variation-based image denoising and deblurring models. Specifically, instead of the total variation $\int |Du|$, she proposed using $\int \phi(|Du|)$ as the regularizer, where ϕ is some specially designed function. In this work, Vese showed the existence and uniqueness of the minimizer of the corresponding models in the space of functions of bounded variation and also numerically tested the proposed models for the case of the potential function $\phi(z) = \sqrt{\epsilon + z^2}$. As a matter of fact, the model that will be discussed in this paper can be regarded as special cases in Vese's work [32]. However, the main theme of our work focuses on the point that our model is able to reduce the staircase effect besides removing noise and keeping edges effectively. Moreover, the potential functions ϕ employed in our model is different from the one that was tested numerically in Vese's work.

To minimize our model, we propose using the augmented Lagrangian method (ALM) [21, 29]. Recently, ALM has been successfully applied for non-differentiable, nonlinear, and higher-order variational models [5, 31, 34, 36–38]. An appealing feature of using ALM lies in the fact that the minimization of our model amounts to the seek of saddle point of some augmented Lagrangian functional, which can be carried out by minimizing several relatively simpler functionals repeatedly and alternately. In fact, these functionals either have closed-form solution or can be solved using fast solvers like FFTs.

As discussed later, the proposed model only involves the first-order derivative, and it can be regarded as a variant of the ROF model. Therefore, one might not expect that the proposed model could achieve as good results as those higher order models [9, 18, 23, 31, 35]. However, with almost the same numerical cost as ROF, the proposed model could alleviate the staircase effect while also removing noise and keeping edges.

The rest of the paper is organized as follows. In the next section, we present our variational model for the noise removal problem, where total variation-based regularizer is piecewisely defined according to image gradient. In Section 3, we address how to minimize the proposed models using ALM, which is followed by the convergence analysis of the numerical algorithm. We then present numerical experiments in Section 4, which validate the features of the proposed models, especially the reduction of staircase effect, and Section 5 is devoted to our conclusions.

2 A total variation–based image denoising model

As discussed in the introduction, by applying the ROF model, the staircasing phenomenon is often unavoidable, which is mainly due to the total variation–based regularizer. Note that this effect often happens in regions with relatively small gradient of image intensity, since locations with large gradient often refer to object boundaries. Therefore, to reduce the staircase effect, some stronger regularizer than the total variation should be imposed. At the same time, we also want to keep total variation for region with large gradient in order to keep object boundaries. Based on this discussion, we propose the minimization of the following functional for the image denoising problem:

$$E(u) = \lambda \int_{\Omega} \phi_{p,a}(|\nabla u|) + \frac{1}{2} \int_{\Omega} (f - u)^2, \quad (1)$$

where $p > 1$ and the potential function $\phi_{p,a}$ is given as follows:

$$\phi_{p,a}(x) = \begin{cases} \frac{1}{pa^{p-1}} |x|^p, & |x| \leq a; \\ |x| - (1 - \frac{1}{p})a, & |x| > a. \end{cases}$$

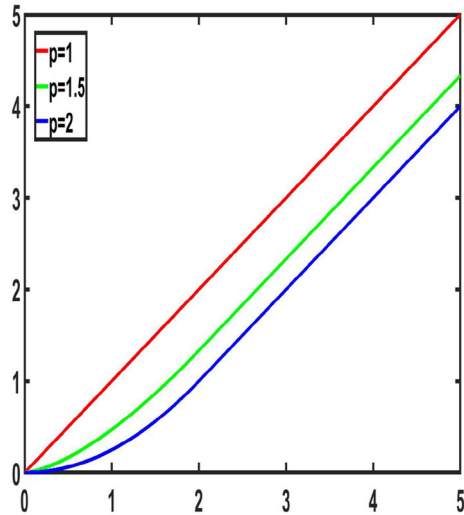
Here, the parameter a defines regions with small or large gradient. With this parameter a , the domain will be decomposed into two parts. On one hand, in the region with $|\nabla u| > a$, the total variation is employed as the regularizer in order to keep edges. In fact, $|\nabla u|$ takes large magnitude along edges. On the other hand, in the region with $|\nabla u| \leq a$, some higher power of $|\nabla u|$ is utilized to impose smooth constraint on u in such a region. For instance, when $p = 2$, in this region, the associated Euler-Lagrange equation reads $-\frac{\lambda}{a} \Delta u + u = f$, and then the function u should be smooth, which explains that this type of regularizer could alleviate the staircase effect. Note that the staircase effect is often observed in regions with relatively small magnitude of gradient.

One can easily check that the function $\phi_{p,a}$ is convex over \mathbb{R} for $p > 1$. We present the graph of this function with $a = 2$ for $p = 1, 3/2, 2$ respectively in Fig. 1 for illustration. The case of $p = 1$ corresponds to the function for the original total variation regularizer. In fact, the case of $p = 2$ was explicitly given in Vese's work [32]. Moreover, for this case, $\phi_{p,a}(x)$ is the Huber function [22], which is also the Moreau envelope of the function $\phi(x) = |x|$.

For an analytical study of the proposed model (Eq. 1), one could follow the discussion in Vese's work [32]. We here only present the key point, and the detailed discussion can be found in [2, 32].

Let $u \in BV(\Omega)$, then its distributional gradient Du defines a Radon measure. By the Radon-Nikodym theorem, one has the decomposition $Du = \nabla u dx + D_s u$, where dx is the Lebesgue measure, $\nabla u = \frac{d(Du)}{dx} \in L^1(\Omega)$, and $D_s u$ and dx are mutually singular. Moreover, as shown by Ambrosio [1], the singular part $D_s u$ can

Fig. 1 The plots of $\phi_{p,a}(x)$ with $a = 2$ for $p = 1, 3/2, 2$ respectively



be further decomposed as a “jump” part and a “Cantor” part. For this, one defines the approximate upper limit $u^+(x)$ and lower limit $u^-(x)$ as follows:

$$u^+(x) = \inf\{t \in [-\infty, \infty] : \lim_{r \rightarrow 0} \frac{dx(\{u > t\} \cap B(x, r))}{r^2} = 0\},$$

$$u^-(x) = \sup\{t \in [-\infty, \infty] : \lim_{r \rightarrow 0} \frac{dx(\{u < t\} \cap B(x, r))}{r^2} = 0\}.$$

Thus, if $u \in L^1(\Omega)$, with respect to the Lebesgue measure, almost all the points in Ω are Lebesgue points, that is, $u(x) = u^+(x) = u^-(x)$, *a.e.x.* Set $S_u = \{x \in \Omega : u^-(x) < u^+(x)\}$, which is the complement of the set of Lebesgue points. S_u is countably rectifiable, and one could define the normal $n_u(x)$ for $H^1 - a.e.x \in \Omega$, where H^1 denotes the Hausdorff measure of dimension one. Then, in [1], Ambrosio showed that

$$Du = \nabla u dx + (u^+ - u^-)n_u H^1|_{S_u} + C_u,$$

where C_u is the Cantor part of $D_s u$, and one could get the following:

$$\int_{\Omega} |Du| = \int_{\Omega} |\nabla u| dx + \int_{S_u} |u^+ - u^-| dH^1 + \int_{\Omega \setminus S_u} |C_u|,$$

As discussed in [32], the proposed model (Eq. 1) can be expressed as follows:

$$RE(u) = \lambda \int_{\Omega} \phi_{p,a}(|\nabla u|) + \lambda \int_{S_u} (u^+ - u^-) dH^1 + \lambda \int_{\Omega - S_u} |C_u| dx + \frac{1}{2} \int_{\Omega} (f - u)^2, \quad (2)$$

It is easy to see that there exist $b > 0, c > 0$ such that $cz - b \leq \phi_{p,a}(z) \leq cz + b$ for any $z > 0$. Then, based on Vese’s work in [32], one could obtain the following existence theorem:

Theorem The functional $\text{RE}(u)$ (Eq. 2) has a unique minimizer in the space $\text{BV}(\Omega)$.

3 Augmented lagrangian method for the proposed model

To minimize the functional (1), one could directly solve its Euler-Lagrange equation, which can be expressed as follows:

$$-\lambda \nabla \cdot \left(\phi'_{p,a}(|\nabla u|) \frac{\nabla u}{|\nabla u|} \right) + u = f. \quad (3)$$

Specifically, one may employ the associated gradient flow to find the solution of this equation.

In this work, we intend to minimize the functional (1) using augmented Lagrangian method (ALM). In fact, during the past years, ALM has been successfully applied for minimizing non-differentiable or/and higher order models [31, 34, 36]. The most appealing feature of using ALM is its simplicity, that is, the original minimization problem can be converted to a constrained optimization one, and the seek of the saddle point of the latter problem amounts to solving several relatively simple functionals repeatedly and alternately. Precisely, for the minimization of functional (1), as in [34], we propose an equivalent constrained optimization problem as follows:

$$\begin{aligned} \min_{q,u} \left\{ \lambda \int_{\Omega} \phi_{p,a}(|q|) + \frac{1}{2} (f - u)^2 \right\}, \\ \text{subject to } q = \nabla u, \end{aligned} \quad (4)$$

and design an augmented Lagrangian functional as follows:

$$\begin{aligned} \mathcal{L}(q, u; \lambda_1) = \lambda \int_{\Omega} \phi_{p,a}(|q|) + \frac{1}{2} (f - u)^2 \\ + \frac{r_1}{2} \int_{\Omega} |q - \nabla u|^2 + \int_{\Omega} \lambda_1 \cdot (q - \nabla u) \end{aligned} \quad (5)$$

where $\lambda_1 \in \mathbb{R}^2$ is a Lagrange multiplier and $r_1 > 0$ is a penalty parameter to be chosen in the numerical implementation.

Based on the theory of optimization, the minimizer of the functional (1) corresponds to the saddle point of the augmented Lagrangian functional (5). To find the saddle point, we employ the standard iterative strategy, that is, we minimize the corresponding subproblem for each of q and u respectively by fixing the other one, and then update the Lagrange multiplier λ_1 ; this process will be repeated until the two variables are convergent, which indicates the saddle point will be approximated. This iterative method for approximating the saddle point of the functional (5) is given in Algorithm 1.

Algorithm 1 Augmented Lagrangian method for the minimization of the proposed model (Eq. 1).

1. Initialization: u^0, q^0 and λ_1^0 . For $k \geq 1$, do the following steps (Step 2 ~ 4):
2. Compute the minimizer u^k, q^k for the associated subproblems with the fixed Lagrangian multiplier λ_1^{k-1} :

$$u^k = \operatorname{argmin}_u \mathcal{L}(u, q^{k-1}; \lambda_1^{k-1}); \tag{6}$$

$$q^k = \operatorname{argmin}_q \mathcal{L}(u^k, q; \lambda_1^{k-1}); \tag{7}$$

3. Update the Lagrangian multiplier λ_1

$$\lambda_1^k = \lambda_1^{k-1} + r_1(q^k - \nabla u^k), \tag{8}$$

4. Measure the relative residual (Eq. 36) and stop the iteration if it is smaller than a given threshold ϵ_r .

Specifically, the two subproblems to be minimized for the variables q and u are given as follows:

$$\varepsilon_1(u) = \frac{1}{2} \int_{\Omega} (f - u)^2 + \frac{r_1}{2} \int_{\Omega} |q - \nabla u|^2 + \int_{\Omega} \lambda_1 \cdot (q - \nabla u); \tag{9}$$

$$\varepsilon_2(q) = \lambda \int_{\Omega} \phi_{p,a}(|q|) + \frac{r_1}{2} \int_{\Omega} |q - \nabla u|^2 + \int_{\Omega} \lambda_1 \cdot (q - \nabla u). \tag{10}$$

As discussed in [31, 34, 36], the minimizer of $\varepsilon_1(u)$ is determined by the corresponding Euler-Lagrange equation as follows:

$$-r_1 \Delta u + u = f - \nabla \cdot (r_1 q + \lambda_1), \tag{11}$$

which is a linear elliptic equation with constant coefficients, and thus can be solved using FFTs.

For the subproblem of q , in this work, we only consider the cases $p = \frac{3}{2}, 2$ and 3 . In fact, the minimizers of $\varepsilon_2(q)$ for these three cases all have closed-form solutions. In what follows, we discuss the minimizer of $\varepsilon_2(q)$ for the case $p = 2$. For $p = 2$, the potential function $\phi_{2,a}$ reads as follow:

$$\phi_{2,a}(x) = \begin{cases} \frac{1}{2a}x^2, & |x| \leq a; \\ |x| - \frac{1}{2}a, & |x| > a. \end{cases}$$

and the minimizer of $\varepsilon_2(q)$ can be described in the following proposition. In fact, this proposition is a well-known fact in optimization [6], where one could derive it using the proximal mapping. For the convenience of reading, we keep a direct proof here.

Proposition If $p = 2$, the minimizer of $\varepsilon_2(q)$ is given as follows:

$$\operatorname{Argmin}_q \varepsilon_2(q) = \begin{cases} \frac{r_1}{r_1 + \lambda/a} q^*, & \text{if } |q^*| \leq a + \frac{\lambda}{r_1}; \\ (1 - \frac{\lambda}{r_1 |q^*|}) q^*, & \text{if } |q^*| > a + \frac{\lambda}{r_1}. \end{cases} \tag{12}$$

where $q^* = \nabla u - \lambda_1/r_1$.

Proof Note that $\varepsilon_2(q)$ can be written as follows:

$$\varepsilon_2(q) = \lambda \int_{\Omega} \phi_{2,a}(|q|) + \frac{r_1}{2} \int_{\Omega} |q - q^*|^2 + \tilde{c},$$

with $q^* = \nabla u - \lambda_1/r_1$ and \tilde{c} is independent of q . Since there is no spatial derivative of q involved, we can get the minimizer of the functional by finding that of the integrand pointwisely. For simplicity, denote the integrand function by $g(q)$, that is, $g(q) = \lambda\phi_{2,a}(|q|) + \frac{r_1}{2}|q - q^*|^2$. As $\phi_{2,a}(x)$ is an increasing function for $x \geq 0$, it is easy to see that the minimizer of $g(q)$ must assume the form sq^* with $s \in [0, 1]$. Therefore, we just need to study the minimizer of a new function $h(s) = g(sq^*)$ with $s \in [0, 1]$. For this, the following two cases need to be considered: (1) $|q^*| \leq a$; (2) $|q^*| > a$.

Case 1. If $|q^*| \leq a$, by the definition of $\phi_{2,a}$, $h(s) = \frac{\lambda}{2a}|q^*|^2s^2 + \frac{r_1}{2}|q^*|^2(s - 1)^2$. One can find the minimum value of $h(s)$ occurs at $s_0 = \frac{r_1}{r_1 + \lambda/a}$, which lies inside $(0, 1)$.

Case 2. If $|q^*| > a$, $h(s)$ can be expressed as follows:

$$h(s) = \begin{cases} \frac{\lambda}{2a}|q^*|^2s^2 + \frac{r_1}{2}|q^*|^2(s - 1)^2, & \text{if } 0 \leq s|q^*| \leq a; \\ \lambda|q^*|s - \frac{\lambda}{2}a + \frac{r_1}{2}|q^*|^2(s - 1)^2, & \text{if } a < s|q^*| \leq |q^*|. \end{cases}$$

It easy to see that $h(s)$ is convex on $[0, 1]$ (note that $\phi_{2,a}(x)$ is convex). Therefore, the minimum value of h occurs at either the two end points or the critical point. As in case 1, if $0 \leq s|q^*| \leq a$, that is, in the interval $[0, a/|q^*|]$, $h(s)$ could take on its minimum value at $s_0 = \frac{r_1}{r_1 + \lambda/a}$. One may check that $s_0 \in [0, a/|q^*|]$ if and only if $|q^*| \leq a + \frac{\lambda}{r_1}$. In another word, if $|q^*| \leq a + \frac{\lambda}{r_1}$, s_0 is a critical point of $h(s)$ in $[0, 1)$, and thus is a minimizer.

Similarly, for $a < s|q^*| \leq |q^*|$ or $s \in (a/|q^*|, 1]$, $h(s)$ assumes its minimum value at $s_1 = 1 - \frac{\lambda}{r_1|q^*|}$ if $s_1 \in (a/|q^*|, 1]$, which is equivalent to be $|q^*| > a + \frac{\lambda}{r_1}$.

Combining both of the above cases, we have if $|q^*| \leq a + \frac{\lambda}{r_1}$, s_0 is a minimizer of $h(s)$ on $[0, 1)$, while if $|q^*| > a + \frac{\lambda}{r_1}$, then s_1 is the minimizer.

Therefore, the minimizer of $\varepsilon_2(q)$ can be expressed as follows:

$$\text{Argmin}_q \varepsilon_2(q) = \begin{cases} \frac{r_1}{r_1 + \lambda/a}q^*, & \text{if } |q^*| \leq a + \frac{\lambda}{r_1}; \\ (1 - \frac{\lambda}{r_1|q^*|})q^*, & \text{if } |q^*| > a + \frac{\lambda}{r_1}. \end{cases} \tag{13}$$

which completes the proof. □

We here only derive the minimizer of $\varepsilon_2(q)$ for the case $p = 2$, in fact, the same procedure can be employed for the cases $p = \frac{3}{2}$ and $p = 3$. Specifically, the minimizer for the case $p = \frac{3}{2}$ reads as follow:

$$\text{Argmin}_q \varepsilon_2(q) = \begin{cases} \left(\frac{2r_1 \sqrt{a|q^*|}}{\sqrt{\lambda^2 + 4r_1^2 a|q^*| + \lambda}} \right)^2 q^*, & \text{if } |q^*| \leq a + \frac{\lambda}{r_1}; \\ (1 - \frac{\lambda}{r_1|q^*|})q^*, & \text{if } |q^*| > a + \frac{\lambda}{r_1}. \end{cases} \tag{14}$$

and the one for $p = 3$

$$\text{Argmin}_q \varepsilon_2(q) = \begin{cases} \frac{2}{\sqrt{1 + 4\lambda|q^*|/(r_1 a^2) + 1}} q^*, & \text{if } |q^*| \leq a + \frac{\lambda}{r_1}; \\ (1 - \frac{\lambda}{r_1|q^*|})q^*, & \text{if } |q^*| > a + \frac{\lambda}{r_1}. \end{cases} \tag{15}$$

Remark In fact, as discussed in [34], the subproblem $\varepsilon_2(q)$ for the ROF model has the form as follows:

$$\varepsilon_2(q) = \lambda \int_{\Omega} |q| + \frac{r_1}{2} \int_{\Omega} |q - \nabla u|^2 + \int_{\Omega} \lambda_1 \cdot (q - \nabla u) \tag{16}$$

and its minimizer reads as follow:

$$\text{Argmin}_q \varepsilon_2(q) = \begin{cases} 0, & \text{if } |q^*| \leq \frac{\lambda}{r_1}; \\ (1 - \frac{\lambda}{r_1|q^*|})q^*, & \text{if } |q^*| > \frac{\lambda}{r_1}. \end{cases} \tag{17}$$

with q^* being defined as in (12). One can see that the ROF model allows 0 for its minimizer when $|q^*| \leq \frac{\lambda}{r_1}$, while for the new model, the minimizer takes on 0 only for the case $q^* = 0$. This shows the major difference between the two models and also demonstrates that the new model is able to reduce the staircasing phenomenon.

In what follows, we provide the convergence result of the proposed ALM discussed in Algorithm 1. By starting with the initial guess of the variables u, q , and λ_1 , one could get a sequence of $\{(u^k, q^k; \lambda_1^k)\}, k = 1, 2, \dots$. In what follows, as in [21, 34], we establish the convergence result for the sequence $\{u^k\}$ for the case $p = 2$. In fact, a similar convergence result can be found in the discrete setting in [10].

Theorem Let $\{(u^k, q^k; \lambda_1^k)\}, k = 1, 2, \dots$ be the sequence generated by the algorithm summarized in Algorithm 1. If $(u^*, q^*; \lambda_1^*)$ is a saddle point of the augmented Lagrangian functional $\mathcal{L}(u, q; \lambda_1)$, then one gets

$$\lim_{k \rightarrow \infty} u^k = u^*. \tag{18}$$

Proof For convenience, in what follows, we denote $\langle f, g \rangle = \int_{\Omega} fg$ and $\|f\|^2 = \int_{\Omega} |f|^2$, where f, g can be vector functions.

Since $(u^*, q^*; \lambda_1^*)$ is a saddle point of the functional $\mathcal{L}(u, q; \lambda_1)$, that is,

$$\mathcal{L}(u^*, q^*; \lambda_1) \leq \mathcal{L}(u^*, q^*; \lambda_1^*) \leq \mathcal{L}(u, q; \lambda_1^*), \text{ for any } u, q, \text{ and } \lambda_1, \tag{19}$$

by using the same argument as in [34], one can get $q^* = \nabla u^*$. As $u^* = \operatorname{argmin}_u \mathcal{L}(u, q^*; \lambda_1^*)$, then for the function $h(t) = \mathcal{L}(u^* + t(u - u^*), q^*; \lambda_1^*)$ for any u and $t \in [0, 1]$, one has $h'(0) \geq 0$, and thus

$$\langle f - u^*, u^* - u \rangle + \langle r_1(q^* - \nabla u^*) + \lambda_1^*, \nabla(u^* - u) \rangle \geq 0, \text{ for any } u. \tag{20}$$

Similarly, as $q^* = \operatorname{argmin}_q \mathcal{L}(u^*, q; \lambda_1^*)$, one gets

$$\lambda \langle \phi_{2,a}(|q|) - \phi_{2,a}(|q^*|), 1 \rangle + \langle r_1(q^* - \nabla u^*) + \lambda_1^*, q - q^* \rangle \geq 0, \text{ for any } q. \tag{21}$$

As $u^k = \operatorname{argmin}_u \mathcal{L}(u, q^{k-1}; \lambda_1^{k-1})$ and $q^k = \operatorname{argmin}_q \mathcal{L}(u^k, q; \lambda_1^{k-1})$, using the same procedure, one obtains the following:

$$\langle f - u^k, u^k - u \rangle + \langle r_1(q^{k-1} - \nabla u^k) + \lambda_1^{k-1}, \nabla(u^k - u) \rangle \geq 0, \text{ for any } u. \tag{22}$$

and

$$\lambda \langle \phi_{2,a}(|q|) - \phi_{2,a}(|q^k|), 1 \rangle + \langle r_1(q^k - \nabla u^k) + \lambda_1^{k-1}, q - q^k \rangle \geq 0, \text{ for any } q. \tag{23}$$

For the convenience of presentation, let us introduce short notations $du^k = u^k - u^*$, $dq^k = q^k - q^*$, and $d\lambda_1^k = \lambda_1^k - \lambda_1^*$. Then, adding the above four inequalities by setting $u = u^k$ in (20), $u = u^*$ in (22), $q = q^k$ in (21), $q = q^*$ in (23) leads to the following:

$$\begin{aligned} & - \langle du^k, du^k \rangle + \langle \nabla(du^k), r_1(q^{k-1} - \nabla u^k) + d\lambda_1^{k-1} \rangle \\ & + \langle dq^k, -r_1(q^k - \nabla u^k) - d\lambda_1^{k-1} \rangle \geq 0. \end{aligned} \tag{24}$$

As $\lambda_1^k = \lambda_1^{k-1} + r_1(q^k - \nabla u^k)$ and $r_1(q^{k-1} - \nabla u^k) = r_1(q^k - \nabla u^k) + r_1(q^{k-1} - q^k)$, this inequality is simplified as follows:

$$- \|du^k\|^2 + \langle \nabla(du^k), d\lambda_1^k + r_1(q^{k-1} - q^k) \rangle + \langle dq^k, -d\lambda_1^k \rangle \geq 0. \tag{25}$$

and then

$$- \|du^k\|^2 + \langle \nabla(u^k - q^k), d\lambda_1^k \rangle + \langle \nabla(du^k), r_1(q^{k-1} - q^k) \rangle \geq 0. \tag{26}$$

As $q^* = \nabla u^*$ and $\nabla(du^k) = dq^k + (\nabla u^k - q^k) = dq^k + \frac{1}{r_1}(\lambda_1^{k-1} - \lambda_1^k)$, one gets

$$\begin{aligned} & - \|du^k\|^2 + \frac{1}{r_1} \langle \lambda_1^{k-1} - \lambda_1^k, d\lambda_1^k \rangle + r_1 \langle dq^k, q^{k-1} - q^k \rangle \\ & + \langle \lambda_1^{k-1} - \lambda_1^k, q^{k-1} - q^k \rangle \geq 0. \end{aligned} \tag{27}$$

As the inequality (23) holds for q^k , we can also write down a similar inequality for q^{k-1} as follows:

$$\lambda \left\langle \phi_{2,a}(|q|) - \phi_{2,a}(|q^{k-1}|), 1 \right\rangle + \left\langle r_1(q^{k-1} - \nabla u^{k-1}) + \lambda_1^{k-2}, q - q^{k-1} \right\rangle \geq 0, \text{ for any } q, \quad (28)$$

Adding the two inequalities by setting $q = q^{k-1}$ in (23) and $q = q^k$ in (28) gives the following:

$$\left\langle \lambda_1^k, q^{k-1} - q^k \right\rangle + \left\langle \lambda_1^{k-1}, q^k - q^{k-1} \right\rangle \geq 0, \quad (29)$$

that is, $\left\langle \lambda_1^{k-1} - \lambda_1^k, q^{k-1} - q^k \right\rangle \leq 0$. With this inequality, the one (27) leads to the following:

$$- \left\| du^k \right\|^2 + \frac{1}{r_1} \left\langle \lambda_1^{k-1} - \lambda_1^k, d\lambda_1^k \right\rangle + r_1 \left\langle dq^k, q^{k-1} - q^k \right\rangle \geq 0 \quad (30)$$

Using the identities

$$\begin{aligned} \left\langle \lambda_1^{k-1} - \lambda_1^k, d\lambda_1^k \right\rangle &= \frac{1}{2} \left(\left\| d\lambda_1^{k-1} \right\|^2 - \left\| d\lambda_1^k \right\|^2 - \left\| \lambda_1^{k-1} - \lambda_1^k \right\|^2 \right) \\ \left\langle dq^k, q^{k-1} - q^k \right\rangle &= \frac{1}{2} \left(\left\| dq^{k-1} \right\|^2 - \left\| dq^k \right\|^2 - \left\| q^{k-1} - q^k \right\|^2 \right) \end{aligned}$$

one gets

$$\begin{aligned} &\left(\frac{1}{2r_1} \left\| d\lambda_1^{k-1} \right\|^2 + \frac{r_1}{2} \left\| dq^{k-1} \right\|^2 \right) - \left(\frac{1}{2r_1} \left\| d\lambda_1^k \right\|^2 + \frac{r_1}{2} \left\| dq^k \right\|^2 \right) \\ &\geq \left\| du^k \right\|^2 + \frac{1}{2r_1} \left\| \lambda_1^{k-1} - \lambda_1^k \right\|^2 + \frac{r_1}{2} \left\| q^{k-1} - q^k \right\|^2, \end{aligned} \quad (31)$$

which shows that the nonnegative sequence $\left\{ \frac{1}{2r_1} \left\| d\lambda_1^k \right\|^2 + \frac{r_1}{2} \left\| dq^k \right\|^2 \right\}_{k=1}^\infty$ is decreasing and convergent, and therefore $\lim_{k \rightarrow \infty} \left\| u^k - u^* \right\| = 0$. \square

Remark In this proof, we only use the convexity of $\phi_{2,a}(x)$, and therefore the conclusion also holds for the cases with $p = \frac{3}{2}, 3$.

4 Numerical implementation

In this section, we present the details of solving the equation (11) and updating the variable q in the numerical implementation, which are similar as those in the works [31, 34, 36].

Let $\Omega = \{(i, j) | 1 \leq i \leq M, 1 \leq j \leq N\}$ be the discretized image domain and each point (i, j) is a grid point. We then define the discrete backward and forward differential operators with periodic boundary condition as follows:

$$\begin{aligned} \partial_1^- u(i, j) &= \begin{cases} u(i, j) - u(i-1, j), & 1 < i \leq M; \\ u(1, j) - u(M, j), & i = 1. \end{cases} \\ \partial_1^+ u(i, j) &= \begin{cases} u(i+1, j) - u(i, j), & 1 \leq i < M - 1; \\ u(1, j) - u(M, j), & i = M. \end{cases} \\ \partial_2^- u(i, j) &= \begin{cases} u(i, j) - u(i, j-1), & 1 < j \leq N; \\ u(i, 1) - u(i, N), & j = 1. \end{cases} \\ \partial_2^+ u(i, j) &= \begin{cases} u(i, j+1) - u(i, j), & 1 \leq j < N; \\ u(i, 1) - u(i, N), & j = N. \end{cases} \end{aligned}$$

and the central difference operators and the gradient operators are defined accordingly:

$$\begin{aligned} \partial_1^c u(i, j) &= (\partial_1^- u(i, j) + \partial_1^+ u(i, j))/2, \\ \partial_2^c u(i, j) &= (\partial_2^- u(i, j) + \partial_2^+ u(i, j))/2, \\ \nabla^\pm u(i, j) &= \langle \partial_1^\pm u(i, j), \partial_2^\pm u(i, j) \rangle. \end{aligned}$$

Note that Eq. (11) is an elliptic equation with constant coefficients, one can use FFTs to solve it. For this, Eq. (11) is discretized as follows:

$$-r_1(\partial_1^+ \partial_1^- u + \partial_2^+ \partial_2^- u) + u = g, \tag{32}$$

where g represents the right-hand side of Eq. (11). By applying Fourier transform to both sides, one gets the following:

$$(1 + 2r_1(2 - \cos z_i^1 - \cos z_j^2))\mathcal{F}u(i, j) = \mathcal{F}g(i, j), \tag{33}$$

where $z_i^1 = 2\pi(i - 1)/M, i = 1, \dots, M$ and $z_j^2 = 2\pi(j - 1)/N, j = 1, \dots, N$. Then once $\mathcal{F}u$ is calculated, u can be obtained using the discrete inverse Fourier transform.

As for the minimizer $q = (q_1, q_2)$ of $\varepsilon_2(q)$, note that its two components are defined on half-grids $(i + \frac{1}{2}, j)$ and $(i, j + \frac{1}{2})$ respectively, in the numerical implementation, we calculate q_1 and q_2 separately. In particular, one has the following:

$$q_{1(i+\frac{1}{2}, j)} = \begin{cases} \frac{r_1}{r_1+\lambda/a} q^*_{1(i+\frac{1}{2}, j)}, & \text{if } |q^*_{i+\frac{1}{2}, j}| \leq a + \frac{\lambda}{r_1}; \\ (1 - \frac{\lambda}{r_1|q^*_{i+\frac{1}{2}, j}|}) q^*_{1(i+\frac{1}{2}, j)}, & \text{if } |q^*_{i+\frac{1}{2}, j}| > a + \frac{\lambda}{r_1}. \end{cases} \tag{34}$$

and

$$q_{2(i, j+\frac{1}{2})} = \begin{cases} \frac{r_1}{r_1+\lambda/a} q^*_{2(i, j+\frac{1}{2})}, & \text{if } |q^*_{i, j+\frac{1}{2}}| \leq a + \frac{\lambda}{r_1}; \\ (1 - \frac{\lambda}{r_1|q^*_{i, j+\frac{1}{2}}|}) q^*_{2(i, j+\frac{1}{2})}, & \text{if } |q^*_{i, j+\frac{1}{2}}| > a + \frac{\lambda}{r_1}. \end{cases} \tag{35}$$

where $q^*_{i+\frac{1}{2},j} = (\partial_1^+ u(i,j), (\partial_2^c u(i,j) + \partial_2^c u(i+1,j))/2) - (\lambda_{11(i+\frac{1}{2},j)}, \lambda_{12(i+\frac{1}{2},j)})/r_1$ and $q^*_{i,j+\frac{1}{2}} = ((\partial_1^c u(i,j) + \partial_1^c u(i,j+1))/2, \partial_2^+ u(i,j)) - (\lambda_{11(i,j+\frac{1}{2})}, \lambda_{12(i,j+\frac{1}{2})})/r_1$.

After the variables u, q are updated, the Lagrange multiplier $\lambda_1 = (\lambda_{11}, \lambda_{12})$ will then be advanced at (i, j) :

$$\begin{aligned} \lambda_{11(i,j)}^k &= \lambda_{11(i,j)}^{k-1} + r_1(q_{1(i,j)}^k - \partial_1^- u_{(i,j)}^k), \\ \lambda_{12(i,j)}^k &= \lambda_{12(i,j)}^{k-1} + r_1(q_{2(i,j)}^k - \partial_2^- u_{(i,j)}^k). \end{aligned}$$

5 Numerical experiments

In this section, we present numerical experiments by applying the proposed model for several images and also compare the results with the ones obtained using the ROF model in order to show how the staircase effect can be alleviated.

In this work, we mainly compare the proposed model with ROF, as the proposed model is merely a variant of ROF with a slight change, and as discussed in Remark 1, the numerical cost for these two models is almost the same. As the proposed model only employs the first-order derivative, one might not expect that it could achieve as good results as those higher order models [9, 18, 23, 31, 35]. In fact, these models have proved to be very successful for image denoising or other imaging problems. To evaluate the performance of the proposed model, we also compare the proposed model with the well-known total generalized variation (TGV) model by Bredies et al. [9], which produces excellent denoising results while also suppressing the staircase effect.

For a fair comparison, we also apply ALM for the minimization of the ROF model. In fact, the algorithm for the ROF model only differs from the one for the proposed model in the update of q in each iteration (see Remark 1). Moreover, we choose the model parameter λ for the two models such that their removed noise parts, $f - u$, have almost the same averaged Frobenius norm, which will be defined later.

As in [31, 36, 38], to monitor the convergence of the iterative process, we check the following relative residual:

$$R^k = \frac{1}{|\Omega|} \|q^k - \nabla u^k\|_{L^1}, \tag{36}$$

where k refers to the iteration number, $|\Omega|$ is the area of domain, and $\|\cdot\|_{L^1}$ denotes the L^1 -norm on Ω . Specifically, for an image $f = [f_{ij}]$, $1 \leq i \leq m, 1 \leq j \leq n$, $\|f\|_{L^1} = \sum_{i=1}^M \sum_{j=1}^N |f_{ij}|$. We also track the relative error of the Lagrange multiplier as following:

$$\frac{\|\lambda_1^k - \lambda_1^{k-1}\|_{L^1}}{\|\lambda_1^{k-1}\|_{L^1}}, \tag{37}$$

the relative error of the solution u^k

$$\frac{\|u^k - u^{k-1}\|_{L^1}}{\|u^{k-1}\|_{L^1}}, \tag{38}$$

and check the energy $E(u^k)$ versus the iteration number k . The value of these relative errors can be used as a stop criterion to terminate the iterative process. For the purpose of presentation, all the above quantities are shown in log scale, as shown in Figs. 7 and 8.

In all of the experiments presented in this paper, the added noise is Gaussian noise with zero mean and standard deviation σ , i.e., the added noise $n \sim N(0, \sigma^2)$. For a fair comparison, we employ the Frobenius norm to measure the removed part. Specifically, for $f = [f_{ij}]$, $1 \leq i \leq m$, $1 \leq j \leq n$, and the averaged Frobenius norm of f reads $\|f\|_F = \sqrt{\sum_{i=1}^M \sum_{j=1}^N |f_{ij}|^2 / MN}$. We then choose the regularization parameters for our model and the ROF model so that their removed noise parts have almost the same Frobenius norm.

We first consider a synthetic image in Fig. 2. The cleaned images show that, just as ROF, the proposed model is able to remove noise effectively, while also keeping the jumps or edges. More importantly, the proposed model alleviates the staircase effect largely. In fact, these features of the proposed model is due to the proposed regularizer. Specifically, along edges with relatively large gradient, the regularizer takes the form of total variation, and hence edge location can be preserved; while in regions with relatively small gradient, we use L^p -norm of gradient as the regularizer with $p > 1$, which imposes higher regularity than total variation on the image. Specifically, when $p = 2$, the associated Euler-Lagrange equation includes the Laplace operator in such regions, which surely helps smooth the image. As discussed previously, the staircase phenomenon is the result of over-sharpening in regions with relatively small magnitude of gradient. Therefore, in these regions, we

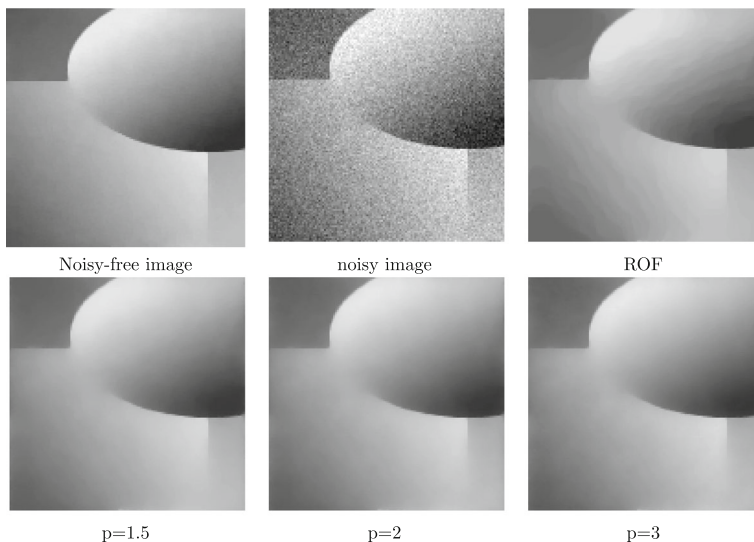


Fig. 2 A noise-free cartoon image, the noisy one, and the denoised images by ROF and by the proposed model with $p = 1.5, 2, 3$, respectively. In this experiment, the standard deviation of added noise is $\sigma = 15$. The additive noise n has the norm $\|n\|_F = 15.03$, and the removed noise part has the norm $\|f - u\|_F = 15.18$ for both our model and ROF

propose adding more higher regularity requirement than the total variation in order to reduce the staircase effect.

In Fig. 3, we apply our model to a real image “Barbara” with $p = 1.5, 2, 3$. The results again demonstrate that the proposed model is capable of removing noise and keeping edge. Moreover, there exists no salient difference among the results for different p . However, the choice of the parameter a is different. Specifically, we used $a = 8$ for $p = 1.5$, $a = 5$ for $p = 2$, and $a = 4$ for $p = 3$. This is because when p is large, we narrow the regions on which we impose the p -Laplace operator in order to avoid the over-smoothness, while when p is small, we expand the region with the new regularizer.

For a better comparison, in Fig. 4, the zoomed-in “forehead” obtained by the two models are presented, which shows that the proposed model ameliorates the staircase effect adequately.

To further test the proposed model, in Fig. 5, we present the zoomed-in “forehead” part obtained by the two models for different level of Gaussian noise, that is, $\sigma = 10, 15$, and 20 , respectively. These results again show that the proposed model is able to alleviate the staircase effect mostly. However, once the noise level (σ) is high, it will affect the output of the proposed model. In these experiments, as we use the same parameter $a = 5$, and therefore if the added noise is high at some region, the total variation-based regularizer will be employed and the smoothing effect will be weakened, and it thus keeps some ridge on the “forehead.” One can surely choose a large regularization parameter λ to remove such ridges; however, it could also over-smooth some meaningful signals. Therefore, in this sense, the proposed model

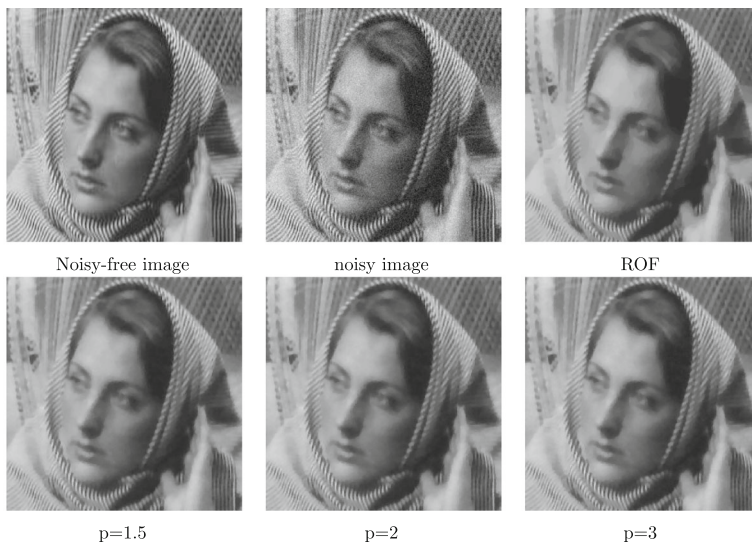


Fig. 3 The noise-free image “Barbara,” the noisy image, and the denoised ones by ROF and by the proposed model with $p = 1.5, 2, 3$, respectively. In this experiment, the standard deviation of added noise is $\sigma = 10$. The additive noise n has the norm $\|n\|_F = 9.98$, and the removed noise part has the norm $\|f - u\|_F = 12.45$ for both our model and ROF

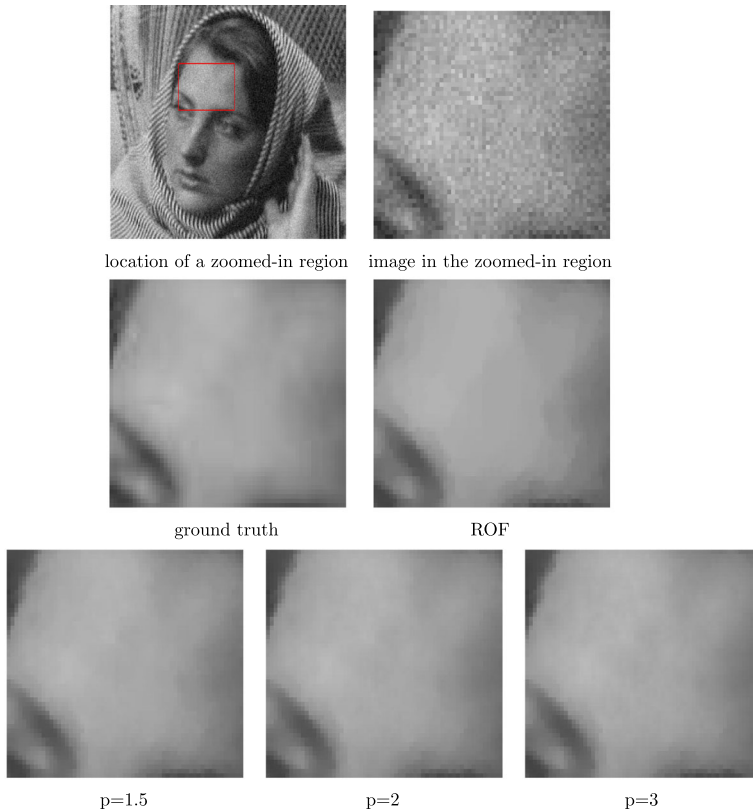


Fig. 4 The comparison of the zoomed-in “forehead” part obtained by the ROF model and the proposed model with $p = 1.5, 2, 3$, respectively. The staircase effect clearly presents in the result by ROF, while it is largely reduced by the new model

is more applicable for images with relatively lower level of noise, for instance, the standard deviation $\sigma = 10$.

In Fig. 6, we compare the performance of the proposed model with $p = 2$ and $\sigma = 10$ by using different parameter a . The results demonstrate that if $a = 2$, there exists some ridges on the “forehead,” while for $a = 8$, the denoised image looks more fuzzy than the choice $a = 5$. In fact, if a is small, the proposed model is close to ROF, since only some region is imposed the new regularizer, and this explains why the denoised image still keeps ridges; and if a is large, the Laplace operator will be applied for a large region, and thus overly smooths signals in that region.

To show the convergence of the iterative process of the proposed ALM, in Fig. 7, we list the plots of relative residual (36), the relative error of Lagrange multiplier (37), the relative error of u^k (38), and the energy $E(u^k)$ versus iteration by applying the proposed model with $p = 2$ for “Barbara” with $\sigma = 10$. Note that, after 1000 iterations, the relative error of u^k is as small as 3.13×10^{-16} , which is close to the precision of floating-point numbers in Matlab. These results illustrate that the

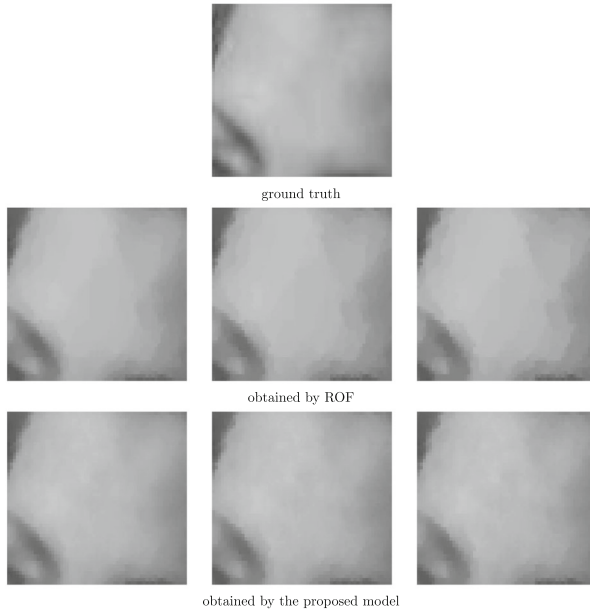


Fig. 5 The comparison of the zoomed-in “forehead” part obtained by the ROF model and the proposed model with $p = 2$ for different Gaussian noise level $\sigma = 10, 15, 20$ (from left to right). The staircase effect clearly presents in the result by ROF, while it is largely reduced by the new model

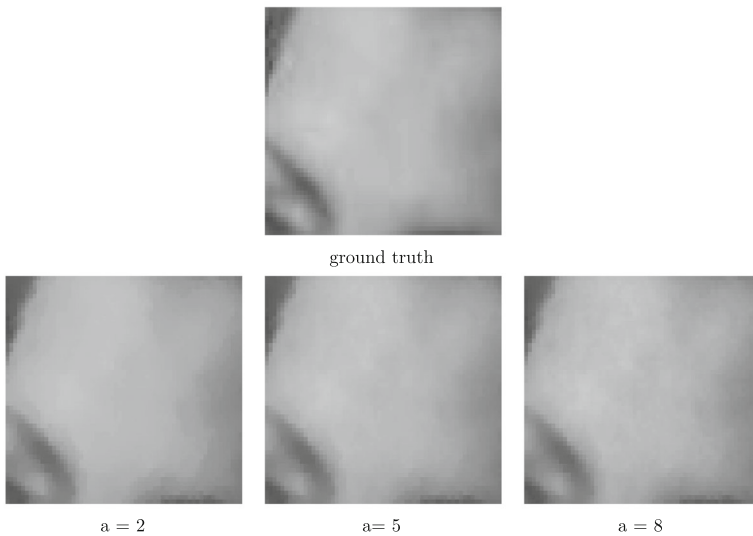


Fig. 6 The comparison of the zoomed-in “forehead” part obtained by the proposed model with $p = 2$ for the image contaminated with Gaussian noise $\sigma = 10$ but with different $a = 2, 5, 8$ (from left to right)

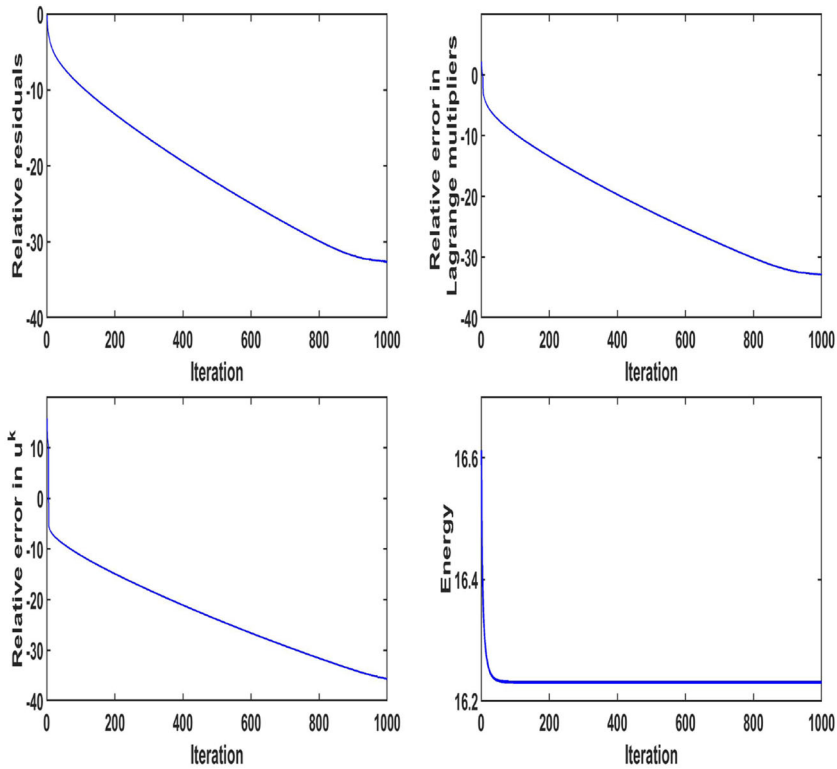


Fig. 7 The plots of relative residual (36), relative error of Lagrange multiplier (37), relative error in u^k (38) and energy $E(u^k)$ for the example “Barbara” contaminated by Gaussian noise with $\sigma = 10$ by using our model with $p = 2$. All the quantities are presented in log scale. Note that the relative error of u^k arrives at 3.13×10^{-16} , which is close to the precision of floating-point numbers in Matlab

iterative process converges to a saddle point of the augmented Lagrangian functional (Eq. 5), and thus the minimizer of the proposed model.

In Fig. 8, we also present the plots of relative error of u^k versus iteration for the ROF model and the proposed model with $p = 2$ when they are applied for “Barbara” with $\sigma = 10$. These plots show that, by using ALM, the proposed model has a strong tendency to converge to the minimizer, while the ROF model could experience slow convergence after some iterations. This is mainly due to the employment of a strongly convex regularizer, that is, the L^2 regularizer, for regions with relative small gradient.

In Figs. 9 and 10, we apply the proposed model for another real image “peppers.” These results again demonstrate that the model is able to remove noise and keep edge efficiently, while also alleviating the staircase effect. Specifically, in the selected zoomed-in region, the proposed model can successfully preserve the shadow on the pepper.

To further compare the quality of the cleaned images by the two models, we list in Table 1 the peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM) [33] for the above experiments. PSNR and SSIM are two commonly

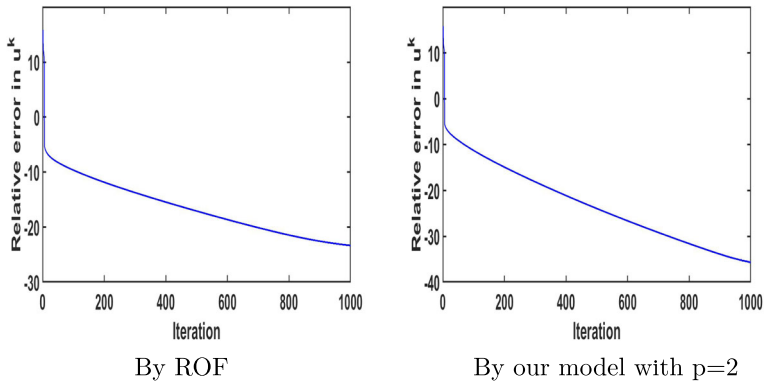


Fig. 8 The comparison of convergence of relative error in u^k (38) for the ROF model and the proposed model with $p = 2$ when applied to “Barbara” with $\sigma = 10$. Note that the vertical quantity is shown in log scale

used objective image metrics to assess the quality of denoised image. In fact, PSNR measures the mean squared difference between a noise-free image and the denoised one. As it is a least square-based measure, it prefers over-smoothed results. On the other hand, SSIM takes into account the following three comparisons: luminance, contrast, and structure [33], and it thus measures the similarity of high-frequency signals like edges between the two images. From Table 1, one can see that ROF

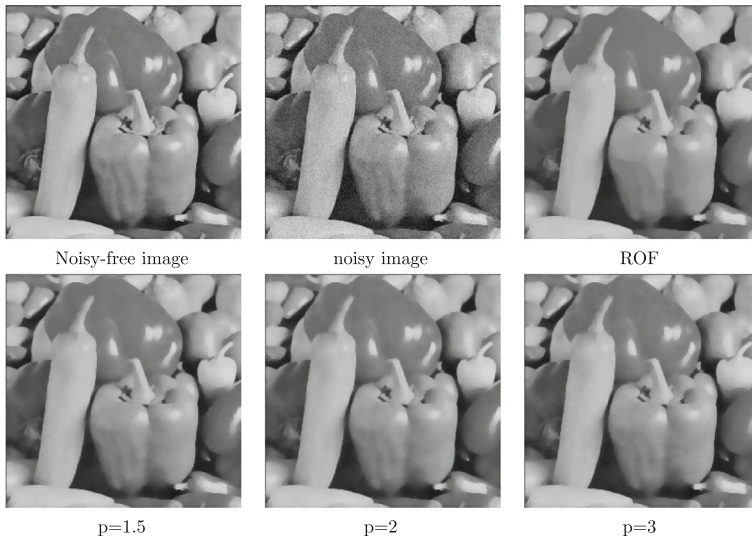


Fig. 9 The noise-free image “peppers,” the noisy image, and the denoised ones by ROF and by the proposed model with $p = 1.5, 2, 3$, respectively. In this experiment, the standard deviation of added noise is $\sigma = 10$. The additive noise n has the norm $\|n\|_F = 9.98$, and the removed noise part has the norm $\|f - u\|_F = 10.6$ for both our model and ROF

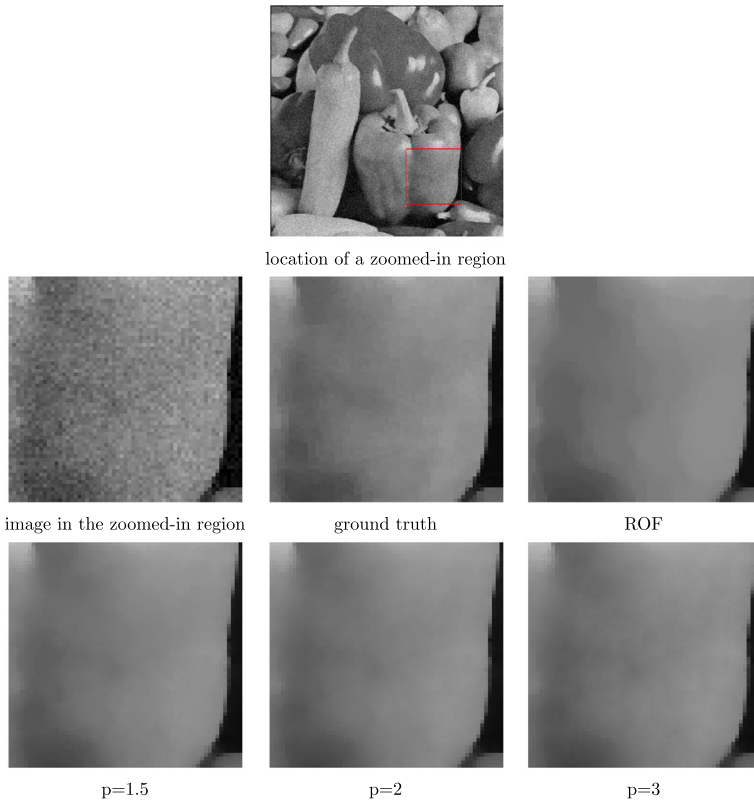


Fig. 10 The comparison of the zoomed-in patch of “peppers” obtained by the ROF model and the proposed model with $p = 1.5, 2, 3$, respectively. The staircase effect clearly presents in the result by ROF, while it is largely reduced by the new model

produces a slightly higher PSNR than the proposed models. However, it is easy to see that the proposed model generates higher SSIM than ROF.

In Figs. 11 and 12, we further compare the results obtained by the proposed method with $p = 2$ and the well-known TGV (total generalized variation) model [9], and the associated values of the PSNR and SSIM in Table 1. To have a fair comparison, for each model and for each experiment, the additive noise and the removed

Table 1 The comparison of PSNR and SSIM for ROF, TGV, and the proposed model

	PSNR			SSIM						
	$p = 3/2$	$p = 2$	$p = 3$	ROF	TGV	$p = 3/2$	$p = 2$	$p = 3$	ROF	TGV
Fig. 2, $\sigma = 15$	34.01	34.01	34.04	33.31	35.18	0.7502	0.7488	0.7440	0.6793	0.7832
Fig. 3, $\sigma = 10$	28.11	28.12	28.10	28.17	28.84	0.7519	0.7513	0.7501	0.7206	0.7710
Fig. 9, $\sigma = 10$	31.84	31.80	31.71	32.02	32.75	0.7562	0.7600	0.7575	0.6553	0.7698

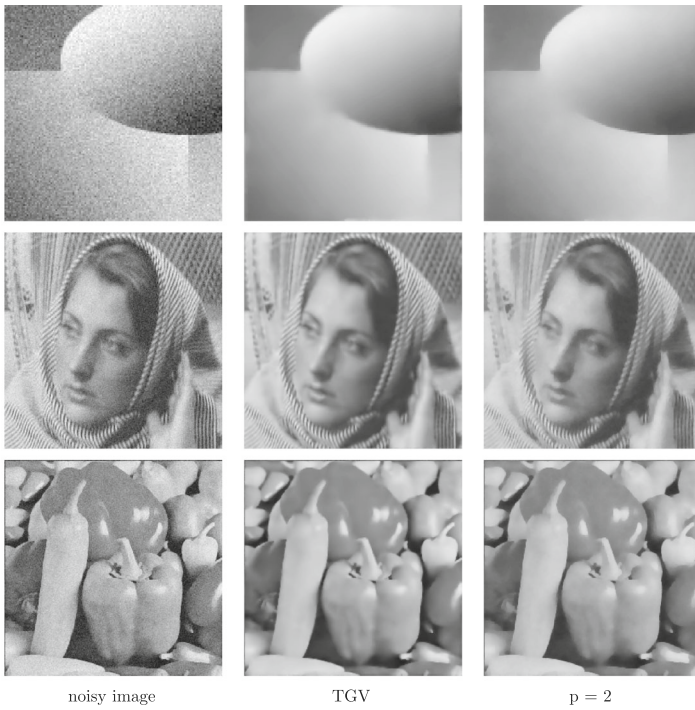


Fig. 11 The comparison of the TGV (total generalized variation) model [9] and the proposed model with $p = 2$. The norms of the additive noise and the removed noise part for both models are as follows: $\|n\|_F = 15.03$, $\|f - u\|_F = 15.18$ for the first example, $\|n\|_F = 9.98$, $\|f - u\|_F = 12.45$ for the second example, and $\|n\|_F = 9.98$, $\|f - u\|_F = 10.65$ for the third example

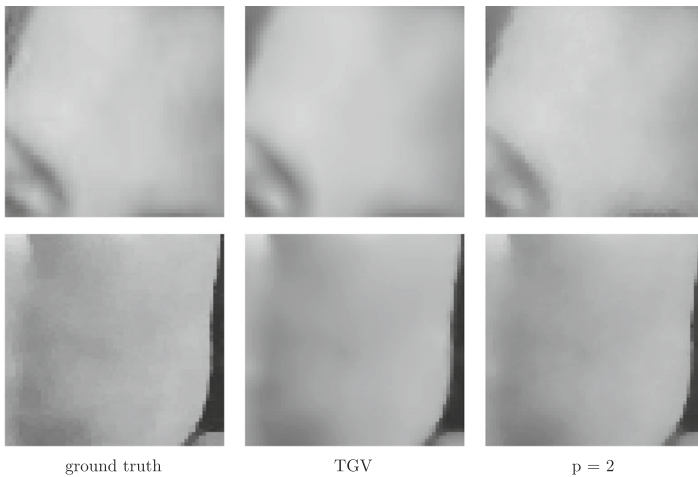


Fig. 12 The comparison of the zoomed-in patches of “Barbara” and “peppers” obtained by the TGV (total generalized variation) model [9] and the proposed model with $p = 2$ for the Gaussian noise level $\sigma = 10$

noise part have almost the same norm. As expected, these figures justify that TGV generates more smooth solution than the proposed model, since it uses higher order derivatives, and as shown in Table 1, TGV also leads to higher values of PSNR and SSIM than the proposed model. On the other hand, with a careful checking of the plots in Figs. 11 and 12, especially for the “Cartoon” image and at the top right part of the “forehead” of “Barbara,” one could see that TGV produces slightly over-smoothed results. Moreover, as the TGV model involves second order derivatives, it often requires more computational effort than the proposed model.

6 Conclusion

In this paper, we consider a variational image denoising model with the aim to ameliorate the staircase effect that is intrinsic to the well-known ROF model. Different from the ROF model, this model applies different type of regularizers for regions with different magnitude of gradient of image intensity, and as a result, edges can be preserved just as the ROF, while smooth regions will not be over-sharpened. To minimize the associated functional, we develop a fast algorithm using the augmented Lagrangian method. The convergence analysis of the proposed algorithm is also provided. Numerical experiments are listed to justify the features of the proposed model and also demonstrate the efficiency of the proposed algorithm.

Acknowledgments The author would like to thank the anonymous referees for their valuable comments and suggestions, which have helped very much to improve the presentation of this paper.

References

1. Ambrosio, L.: A compactness theorem for a new class of functions of bounded variation. *Bollettino della Unione Matematica Italiana* **VII**(4), 857–881 (1989)
2. Aubert, G., Kornprobst, P.: *Mathematical problems in image processing: partial differential equations and the calculus of variations*. Springer Science and Business Media (2006)
3. Aubert, G., Vese, L.: A variational method in image recovery. *SIAM J. Numer. Anal.* **34**(5), 1948–1979 (1997)
4. Brito-Loeza, C., Chen, K.: Multigrid algorithm for high order denoising. *SIAM J. Imaging. Sciences* **3**(3), 363–389 (2010)
5. Bae, E., Tai, X.C., Zhu, W.: Augmented Lagrangian method for an Euler’s elastica based segmentation model that promotes convex contours. *Inverse Probl. Imag.* **11**(1), 1–23 (2017)
6. Beck, A.: *First-Order Methods in Optimization*, vol. 25, SIAM (2017)
7. Bellettini, G., Caselles, V., Novaga, M.: The total variation flow in \mathbb{R}^n . *J. Differ. Equations* **184**, 475–525 (2002)
8. Bertalmio, M., Vese, L., Sapiro, G., Osher, S.: Simultaneous structure and texture image inpainting. *IEEE Trans. on Image Process.* **12**(8), 882–889 (2003)
9. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. *SIAM J. Imaging Sci.* **3**(3), 492–526 (2010)
10. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends[®] in Machine learning* **3**(1), 1–122 (2011)
11. Buttazzo, G.: *Semicontinuity, relaxation and integral representation in the calculus of variations*, Pitman Research Notes in Mathematics 207, Longman Scientific and Technical (1989)

12. Chang, Q.S., Che, Z.Y.: An adaptive algorithm for TV-based model of three norms L_q , ($q = \frac{1}{2}, 1, 2$) in image restoration. *Applied Math and Comput.* **329**, 251–265 (2018)
13. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math Imaging Vis.* **40**, 120–145 (2011)
14. Chan, T., Esedoglu, S.: Aspects of total variation regularized l^1 function approximation. *SIAM J. Appl. Math.* **65**(5), 1817–1837 (2005)
15. Chan, T., Esedoglu, S., Park, F., Yip, M.H.: Recent developments in total variation image restoration. In: Paragios, N., Chen, Y., Faugeras, O. (eds.) *Handbook of Mathematical Models in Computer Vision*. Springer, Berlin (2005)
16. Chambolle, A., Lions, P.L.: Image recovery via total variation minimization and related problems. *Numer. Math.* **76**, 167–188 (1997)
17. Chan, T., Shen, J.: Mathematical models for local nontexture inpaintings. *SIAM J. Appl. Math.* **62**(3), 1019–1043 (2001)
18. Chan, T., Marquina, A., Mulet, P.: High-order total variation-based image restoration. *SIAM J. Sci. Comput.* **22**(2), 503–516 (2000)
19. Chan, T., Wong, C.K.: Total variation blind deconvolution. *IEEE Trans. Image Process.* **7**(3), 370–375 (1998)
20. Goldstein, T., Osher, S.: The split Bregman method for L1-regularized problems. *SIAM J. Imaging Sci.* **2**, 323–343 (2009)
21. Glowinski, R., Le Tallec, P.: *Augmented Lagrangian and Operator-splitting Methods in Nonlinear Mechanics*. SIAM, Philadelphia (1989)
22. Huber, P.J.: Robust estimation of a location parameter. *Ann. Stat.* **53**(1), 73–101 (1964)
23. Lysaker, M., Lundervold, A., Tai, X.C.: Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. *IEEE Trans. Image Process.* **12**, 1579–1590 (2003)
24. Lysaker, M., Osher, S., Tai, X.C.: Noise removal using smoothed normals and surface fitting. *IEEE Trans. Image Process.* **13**(10), 1345–1457 (2004)
25. Meyer, Y.: *Oscillating patterns in image processing and nonlinear evolution equations*, University Lecture Series, Vol 22, Amer. Math. Soc.
26. Mumford, D., Shah, J.: Optimal approximation by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.* **42**, 577–685 (1989)
27. Osher, S., Burger, M., Goldfarb, D., Xu, J.J., Yin, W.T.: An iterative regularization method for total variation-based image restoration. *Multiscale Model. Simul.* **4**, 460–489 (2005)
28. Osher, S., Sole, A., Vese, L.: Image decomposition and restoration using total variation minimization and the h^{-1} norm. *SIAM Multiscale Model. Simul.* **1**, 349–370 (2003)
29. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.* **1**(2), 97–116 (1976)
30. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithm. *Phys. D* **60**, 259–268 (1992)
31. Tai, X.C., Hahn, J., Chung, G.J.: A fast algorithm for Euler’s Elastica model using augmented Lagrangian method. *SIAM J. Imaging Sciences* **4**(1), 313–344 (2011)
32. Vese, L.: A study in the BV space of a denoising-deblurring variational problem. *Appl. Math Optim.* **44**, 131–161 (2001)
33. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)
34. Wu, C., Tai, X.C.: Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, Vectorial TV, and high order models. *SIAM J. Imaging Sciences* **3**(3), 300–339 (2010)
35. Zhu, W., Chan, T.: Image denoising using mean curvature of image surface. *SIAM J. Imaging Sciences* **5**(1), 1–32 (2012)
36. Zhu, W., Tai, X.C., Chan, T.: Augmented Lagrangian method for a mean curvature based image denoising model. *Inverse Probl. Imag.* **7**(4), 1409–1432 (2013)
37. Zhu, W., Tai, X.C., Chan, T.: Image segmentation using Euler’s elastica as the regularization. *J. Sci. Comput.* **57**(2), 414–438 (2013)
38. Zhu, W.: A numerical study of a mean curvature denoising model using a novel augmented Lagrangian method. *Inverse Probl. Imag.* **11**(6), 975–996 (2017)