

A Projected Gradient Method for a High-order Model in Image Restoration Problems

Zhi-Feng Pang¹, Baoli Shi¹ and Lihong Huang^{2,3}

¹ College of Mathematics and Information Science, Henan University, Kaifeng, 475004, China

² College of Mathematics and Econometrics, Hunan University, Changsha, 410082, China

³ Hunan Women's University, Changsha, 410004, China

Received: Sep 3, 2011; Revised Oct. 4, 2011; Accepted May. 11, 2012

Published online: 1 Sep. 2012

Abstract: Based on the augmented Lagrangian strategy, we propose a projected gradient method for solving the high-order model in image restoration problems. Based on the Bermúdez and Moreno (BM) algorithm, the convergence of the proposed method is proved. We also give the relationship that the semi-implicit gradient descent method can be deduced from the projected gradient method. Some numerical experiments are arranged to demonstrate the efficiency of the proposed method for restoring the gray-level and color images.

Keywords: Augmented Lagrangian strategy, high-order model, Kronecker product, projected gradient method, semi-implicit gradient descent method.

1. Introduction

During the past two decades, the restoration of digital images based on variational models and optimization techniques has been extensively studied in many areas of image processing and computer vision such as image denoising, image deblurring, image zooming and image inpainting, etc. [1, 9]. Among these variational models, a frequently used model proposed by Rudin, Osher, and Fatemi (ROF) [19] considers to solve the following problem

$$\min_u \frac{\alpha}{2} \int_{\Omega} (f - u)^2 dx + |u|_{BV(\Omega)}, \quad (1)$$

where $\Omega \subset \mathbb{R}^2$ with Lipschitz boundary, $BV(\Omega)$ is the space of functionals with bounded variation and α is the regularization parameter. Since the ROF model has the ability to preserve image edges, this model and its variants have been widely used in the image restoration problems [1, 9, 18].

However, in the course of restoring the deteriorated image, the ROF model is well known to make the solution be piecewise constant (called the staircasing effect). To overcome this effect, some high-order PDEs [5, 10, 17, 21, 25] have been introduced. Its reason is that the high-

order PDEs can damp the oscillations much faster and require much stronger smoothness. One of high-order PDEs has been proposed by Lysaker, Lundervold, and Tai [17] as follows:

$$\min_u \frac{\lambda}{2} \int_{\Omega} (f - u)^2 dx + |u|_{BV^2(\Omega)}, \quad (2)$$

where λ is the regularization parameter and $BV^2(\Omega)$ called the second-order bounded variation space is defined by

$$|u|_{BV^2(\Omega)} = \sup \left\{ \int_{\Omega} u \operatorname{div}^2 \xi dx \mid \xi \in C_c^2(\Omega), \|\xi\|_{L^\infty(\Omega)} \leq 1 \right\} < \infty.$$

In fact, when $u \in W^{2,1}(\Omega)$, we can get $|u|_{BV^2(\Omega)} = \int_{\Omega} |\nabla^2 u| dx$, where $\nabla^2 u$ denotes the Hessian of u . Obviously, we can extend the model (2) to a general high-order model as

$$\min_u \frac{\gamma}{2} \int_{\Omega \setminus D} (f - Ku)^2 dx + |u|_{BV^2(\Omega)}, \quad (3)$$

where γ is the regularization parameter and $D \subseteq \Omega$. It is obvious that the model (3) is the deblurring problem when K is the convolution operator and $D = \emptyset$ [24]. When K

* Corresponding author: zhifengpang@163.com.

is the identity operator and $D \neq \emptyset$, this model turns to the image zooming or inpainting problem [12].

From the numerical point of view, the high-order model similar to the ROF model is not straightforward to be minimized, which is due to the fact that $|u|_{BV^2(\Omega)}$ is not differentiable. Thus a regularization approach is devised by employing $\int_{\Omega} |\nabla^2 u|_{\varepsilon} dx = \sqrt{u_{xx}^2 + u_{xy}^2 + u_{yx}^2 + u_{yy}^2} + \varepsilon dx$ for a small positive ε to replace $|u|_{BV^2(\Omega)}$. Following this regularization, an artificial time marching algorithm [17, 26] or a fixed point iteration [22] is used to compute the related solution. But both of these methods suffer from some difficulties such as choosing the suitable parameter ε . On the other hand, it is obvious that the solution obtained by these two methods is not the true solution of the high-order model (3). Another approaches, which can avoid dealing with the nondifferentiable term $|u|_{BV^2(\Omega)}$, follow from the motivation of Chambolle's work [6]. These methods use the Legendre-Fenchel transformation to transform the original problem into the related dual problem so that we can easily obtain the solution of the original problem [8, 21, 20].

The augmented Lagrangian strategy, which can efficiently combine the advantages of the Lagrangian method and the penalty method, has been successfully applied to the image processing problem [14–16, 23, 24]. Recently, different from the methods by solving some subproblems to find the saddle point of the Lagrangian functions [16, 23, 24], Ito and Kunisch in their papers [14, 15] designed a smoothing function to approximate the original nonsmooth function in Hilbert space and then obtained the related solution based on an active set method. Motivated by the work of Ito and Kunisch [14, 15], in this paper, we propose a projected gradient method as a basic method to solve the high-order model (2). This projected gradient method requires a relatively small memory footprint and is easier to solve the high-order model (2) than the active set method used in [14, 15]. Thanks to the Bermúdez and Moreno(BM) algorithm [3], the convergence of the proposed algorithms is proved. We also give the relationship that the typical semi-implicit gradient descent algorithm [8] can be naturally deduce from this projected gradient method. Furthermore, different from the augmented Lagrangian method used to solve the image deblurring problem [24] and the artificial time marching method to solve image zooming [14] or the image inpainting problem [12], we also extend to this projected gradient method to efficiently solve these aforementioned image restoration problems.

The plan of the paper is as follows. We first recall some results about the augmented Lagrangian strategy in Section 2. Then we propose a projected gradient method to solve the high-order model based on the augmented Lagrangian strategy and give the convergence result of this method in Section 3. Furthermore, we also refer to the relationship between the projected gradient method and the typical semi-implicit gradient descent method in this section. We present experimental results in support of our pro-

posed algorithms in Section 4, followed by some conclusions in Section 5.

2. Augmented Lagrangian method

Augmented Lagrangian method has many advantages over the Lagrangian method and the penalty method, which was recently applied to solve nonsmooth, convex optimization problems in image processing [14–16, 23, 24]. However, different from its applications in [16, 23, 24], Ito and Kunisch [14, 15] proposed a Lagrangian smoothing regularization strategy in Hilbert space, which gives us equivalent but regularized strategy. Now, we recall this strategy and give some related results.

Set X, Y be Hilbert spaces. Assume that $g : X \rightarrow R$ is a continuously differentiable, convex function and $\varphi : Y \rightarrow R$ is a proper, lower semicontinuous, convex function. Furthermore, we also assume that φ is nondifferentiable in origin and $X = X^*$. Let us consider the following optimization problem

$$\min_{u \in X} g(u) + \varphi(\Lambda u), \quad (4)$$

where $\Lambda : X \rightarrow Y$ is a bounded linear operator. Then, based on the results [11], the following statement holds.

Lemma 1. *The necessary and sufficient condition for $u^* \in X$ to be the minimizer of (4) is given by*

$$\begin{cases} \xi^* \in \partial\varphi(\Lambda u^*), \\ g'(u^*) + \Lambda^* \xi^* = 0, \end{cases} \quad (5)$$

where $\partial\varphi(x)$ denotes the subdifferential at the point x [11] which is defined by

$$\partial\varphi(x) = \{x^* \in X : \varphi(y) - \varphi(x) \geq \langle x^*, y - x \rangle \forall y \in X\}.$$

It is obvious that u^* satisfying (5) implies that u^* is the solution of (4). However, due to the nondifferentiable of φ , it is difficult to find some efficiently numerical methods to obtain it. To overcome this drawback, a regularization method [14, 15] was proposed by designing a smooth approximation based on the augmented Lagrangian strategy. That is to say, we can equivalently convert the problem (4) into the following constrained problem

$$\begin{cases} \min g(u) + \varphi(\Lambda u - v), \\ \text{subject to } v = 0 \text{ in } X. \end{cases} \quad (6)$$

The equality constraint $v = 0$ in (6) can be treated by the augmented Lagrangian method. By employing the augmented Lagrangian strategy [13], we can transform the constrained optimization problem (6) into the following unconstrained problem

$$\min g(u) + \varphi(\Lambda u - v) + (\xi, v) + \frac{c}{2} |v|_X^2, \quad (7)$$

where $\xi \in X$ is a multiplier and c is a positive penalty parameter. Let

$$\varphi_c(\Lambda u, \xi) = \inf_{v \in X} \left\{ \varphi(\Lambda u - v) + (\xi, v) + \frac{c}{2} |v|_X^2 \right\}, \quad (8)$$

then $\varphi_c(\Lambda u, \xi)$ is a C^1 approximation of φ [14, 15].

Definition 1. Let $G : X \rightarrow R \cup \{\infty\}$ be a proper, convex function, then the conjugate function of G is defined by

$$G^*(x^*) = \sup_{x \in X} \left\{ \langle x, x^* \rangle - G(x) \right\} \text{ for } x^* \in X.$$

Lemma 2. The infimum problem (8) can be rewritten as

$$\varphi_c(\Lambda u, \xi) = \sup_{y^* \in X} \left\{ \langle \Lambda u, y^* \rangle - \varphi^*(y^*) - \frac{1}{2c} |y^* - \xi|^2 \right\} \quad (9)$$

Furthermore, the supremum of (9) is attained at a unique point

$$\xi_c(u, \xi) = \varphi'(\Lambda u, \xi) = H_c(\Lambda u + c^{-1}\xi).$$

where $H_c = c(I - J_{1/c})$.

If set $\delta_{C^*}(y^*)$ denote the characteristic function of a closed convex set $C^* \subset X$, i.e., $\delta_{C^*}(y^*) = 0$ if $y^* \in C^*$, else $\delta_{C^*}(y^*) = \infty$. Then, for the supremum problem (9), we have the following assertions [14, 15].

Lemma 3. Assume that $\varphi^*(y^*) = \delta_{C^*}(y^*)$ in the supremum problem (9), we have the following assertions:

(1) The supremum of (9) is attained at a unique point

$$\xi_c(u, \xi) = \mathcal{P}_{C^*}(\xi + c\Lambda u),$$

where $\mathcal{P}_{C^*}(u)$ denotes the projection of $u \in X$ onto C^* .

(2) If $\xi \in \partial\varphi(\Lambda u)$ for $\xi, u \in X$, then $\xi = \mathcal{P}_{C^*}(\xi + c\Lambda u)$ for all $c > 0$.

3. Projected gradient method

To overcome the nonsmoothness of the high-order model (2), some numerical methods have been proposed in [8, 17, 21, 26, 22]. In this section, based on the augmented Lagrangian strategy, we first propose a projected gradient method to solve the high-order model (2) and give the convergence analysis by using the results of the Bermúdez and Moreno(BM) algorithm [3]. Then we arrange some implementation details for the proposed method and also mention that the typical semi-implicit gradient descent algorithm [8] can be deduced from the projected gradient method.

3.1. Projected gradient method for the high-order model

Set $g(u) = \frac{\lambda}{2} \|f - u\|_{L^2(\Omega)}^2$, $\varphi(\Lambda u) = \int_{\Omega} |\Lambda u| dx$ and $\Lambda = \nabla^2$. If assume that u is in the Hilbert space, then the high-order model (2) can be rewritten as a special form of the minimization problem (4). Following from Lemma 2.4, we then have the following result.

Theorem 1. The solution u^* of the high-order model (2) satisfies

$$\begin{cases} u^* = f - \frac{1}{\lambda} \text{div}^2 \xi^*, \\ \xi^* = \mathcal{P}_{C^*}(\xi^* + c\nabla^2 u^*), \end{cases} \quad (10a)$$

where C^* is given by

$$C^* = \{y \in X : |y| \leq 1 \text{ a.e. in } \Omega\}.$$

It is not difficult to find that Theorem 3.1 implies the optimality condition of the high-order model (2). Especially, Eq. (10a) is based on the fact that the conjugate function of $\varphi(\Lambda u)$ is a characteristic function on the closed and convex set C^* . If we define the following Lagrangian functional

$$L_c(u, \xi) = \frac{\lambda}{2} \int_{\Omega} (f - u)^2 dx + \varphi_c(\nabla^2 u, \xi), \quad (11)$$

where the definition of $\varphi_c(\nabla^2 u, \xi)$ is similar to (8), we can deduce that (u^*, ξ^*) is the saddle point of the Lagrangian functional (11). To get this optimal point (u^*, ξ^*) , we propose the following iterative method.

• **Algorithm 3.1.** Projected Gradient Method.

(I) Set $\xi^0 = 0$ and $c > 0$;

(II) Compute (u^n, ξ^{n+1}) by

$$\begin{cases} u^n = f - \frac{1}{\lambda} \text{div}^2 \xi^n, \\ \xi^{n+1} = \mathcal{P}_{C^*}(\xi^n + c\nabla^2 u^n); \end{cases}$$

(III) If the stop criterion is not satisfied, set $n := n + 1$ and go to step (II).

In order to get the convergence result of Algorithm 3.1, we need to recall some results of the Bermúdez and Moreno(BM) algorithm [3]. Assume that V and E are the Hilbert spaces, $B : E \rightarrow V$ is a bounded linear operator and B^* is the adjoint operator of B . With choosing an arbitrarily original value v^0 , the BM algorithm then employs the iterative strategy

$$\begin{cases} v^n = A^{-1}(h - B y^n), \\ y^{n+1} = H_{\eta}(B^* v^n + \eta y^n) \end{cases} \quad (12)$$

to solve the minimization problem

$$\min_{v \in V} \frac{1}{2} (Av, v) - (h, v) + \phi(v), \quad (13)$$

where $\phi = \psi \circ B^*$ and $\psi : E \rightarrow R$. Bermúdez and Moreno(BM) [2, 3] gave the following convergence result of the algorithm (12).

Theorem 2. If the following assumptions hold:

(1) A is a linear symmetric coercive operator: $(Av, v)_V \geq \alpha \|v\|_V^2$ and is continuous on the finite dimensional subspaces of V ;

(2) There exists v_0 in $\text{dom}(\phi)$ satisfying $\frac{(Av, v - v_0) + \phi(v)}{\|v\|} \rightarrow +\infty$ if $\|v\| \rightarrow +\infty$.

(3) The parameter η satisfies

$$0 \leq \eta \leq \frac{2\alpha}{\|B^*\|^2}.$$

where $\|B^*\|$ denotes the L^2 norm of the operator B^* .

Then the sequence $\{v^n\}$ generated by (12) satisfies $\lim_{n \rightarrow \infty} v^n = v^*$, here v^* is the solution of (13). Furthermore, $y^n \rightharpoonup y$ in E with $y \in \partial\psi(B^*v)$.

If set $V = L^2(\Omega)$, $E = (L^2(\Omega))^4$, $B = \nabla^2$, $A = \lambda I$ and $h = \lambda f$, then all of the assumptions in Theorem 3.2 are satisfied for the high-order model (2). This implies that Algorithm 3.1 is a special case of the BM algorithm. So we get the following result based on Theorem 3.1 and 3.2.

Corollary 1. If the parameter c satisfies

$$0 \leq c \leq \frac{2\lambda}{\|\nabla^2\|^2},$$

where $\|\nabla^2\|$ denotes the L^2 norm of the Hessian operator ∇^2 . Then the sequence $\{u^n\}$ generated by Algorithm 3.1 converges to the unique solution u^* of the high-order model (2).

3.2. Projected gradient method and semi-implicit gradient descent method

We start by giving a discrete algorithm for solving the high-order model (2). For convenience, we first consider the gray-level image. Let an $N \times N$ image be denoted by $u \in R^{N^2}$ with the column lexicographical ordering. Denote the first-order forward difference matrix as D_0 , then backward difference matrix can be denoted as $-D_0^T$. Assume that u satisfies the zero Neumann boundary condition, then the discretization of (2) can be written as

$$\min_u \frac{\lambda}{2} \|f - u\|_{L^2(\Omega)}^2 + |Mu|_{L^1(\Omega)}, \tag{14}$$

where

$$M = \begin{pmatrix} (I \otimes D_0^T)(I \otimes D_0) \\ (I \otimes D_0^T)(D_0 \otimes I) \\ (D_0 \otimes I)(I \otimes D_0^T) \\ (D_0^T \otimes I)(D_0 \otimes I) \end{pmatrix} = \begin{pmatrix} I \otimes (D_0^T D_0) \\ D_0^T \otimes D_0 \\ D_0 \otimes D_0^T \\ (D_0^T D_0) \otimes I_0 \end{pmatrix} \tag{15}$$

and $|Mu|_1$ is given by

$$|Mu|_1 = \sum_{i=1}^{N^2} \sqrt{[M_1 u]_i^2 + [M_2 u]_i^2 + [M_3 u]_i^2 + [M_4 u]_i^2}.$$

where $M_1 = I \otimes (D_0^T D_0)$, $M_2 = D_0^T \otimes D_0$, $M_3 = D_0 \otimes D_0^T$ and $M_4 = (D_0^T D_0) \otimes I$ and I denotes the N -order identity matrix. Here \otimes denotes the Kronecker product. Based on the fact that $(\text{div}^2 u, w)_X = (u, \nabla^2 w)_Y$, it is easy to denote the discrete form of div^2 . Furthermore, the

projection of x on the closed convex set $\mathcal{C} := \{y : |y| \leq 1\}$ can be written as

$$\mathcal{P}_{\mathcal{C}}(x) = \frac{x}{\max\{1, |x|\}}.$$

By the above facts, Theorem 3.1 can be reformulated as the discrete form below.

Theorem 3. The solution u^* of (14) satisfies the following equation

$$\begin{cases} u^* = f - \frac{1}{\lambda}(M^T \xi^*), & (16a) \\ \xi^* = \frac{\xi^* + \tau(Mu^*)}{\max(1, |\xi^* + \tau(Mu^*)|)}. & (16b) \end{cases}$$

for each $\tau > 0$.

Then the discretization of Algorithm 3.1 can be rewritten as the following strategy.

• **Algorithm 3.2.** Projected Gradient Method(PGM).

- (I) Set $\xi^0 = 0$ and $\tau > 0$;
- (II) Compute (u^n, ξ^{n+1}) by

$$\begin{cases} u^n = f - \frac{1}{\lambda}(M^T \xi^n), \\ \xi^{n+1} = \frac{\xi^n + \tau(Mu^n)}{\max(1, |\xi^n + \tau(Mu^n)|)}; \end{cases}$$
- (III) If the stop criterion is not satisfied, set $n := n + 1$ and go to step (II).

In fact, Algorithm 3.2 is convergent when the parameter τ satisfies some conditions which are based on the spectral radius of M . For the Kronecker product, we recall some of its basic properties if the related matrices are assumed as the N -order real-value square matrices.

Lemma 4. For matrices A, B, C and D , the following assertions hold [27]:

- (1) $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$.
- (2) $(A \otimes B)^T = A^T \otimes B^T$, here T denotes the conjugate transpose of the matrix.
- (3) $A \otimes B$ is unitary if A and B are unitary.
- (4) Let σ_i and $\mu_i, i = 1, \dots, N$, are the eigenvalues of A and B respectively. Then the eigenvalues of $A \otimes B$ are $\sigma_i \mu_i, i = 1, \dots, N$.

Lemma 5. For the matrix A , there exists N -order unitary matrices U and V such that

$$A = U \text{diag}(s_i) V,$$

where $\text{diag}(s_i)$ is a N -order diagonal matrix with (i, i) entries the singular values of A .

Theorem 4. The spectral radius of M satisfies $\rho(M) \leq 8$.

Proof. Set $S = D_0^T D_0$. By Lemma 3.1, we can deduce that S has the singular value decomposition $S = U \text{diag}(\sigma_i^2) V$, here σ_i is the singular value of D_0 , U and V are N -order unitary matrices.

We notice that D_0 is a sparse matrix with 1 and -1 as the only two nonzero elements per row, it is not difficult to deduce that $\|D_0\|_\infty = 2$. On the other hand, since the largest eigenvalue of $M^T M$ is the ℓ^2 norm of M , we consider the following matrix:

$$\begin{aligned} M^T M &= (I \otimes S)^T (I \otimes S) + (S \otimes I)^T (S \otimes I) + (D_0^T \otimes D_0)^T (D_0^T \otimes D_0) + (D_0 \otimes D_0^T)^T (D_0 \otimes D_0^T) \\ &= I \otimes (S^T S) + (S^T S) \otimes I + (D_0 \otimes D_0^T) (D_0^T \otimes D_0) + (D_0^T \otimes D_0) (D_0 \otimes D_0^T) \\ &= I \otimes (S^T S) + (S^T S) \otimes I + S^T \otimes S + S \otimes S^T \\ &= (V^T V) \otimes (V^T \text{diag}(\sigma_i^4) V) + (V^T \text{diag}(\sigma_i^4) V) \otimes (V^T V) + (V^T \text{diag}(\sigma_i^2) U^T) \otimes (U \text{diag}(\sigma_i^2) V) \\ &\quad + (U \text{diag}(\sigma_i^2) V) \otimes (V^T \text{diag}(\sigma_i^2) U^T) \\ &= (V^T V) \otimes (V^T \text{diag}(\sigma_i^4) V) + (V^T \text{diag}(\sigma_i^4) V) \otimes (V^T V) + (V^T \text{diag}(\sigma_i^2) U^T) \otimes (U \text{diag}(\sigma_i^2) V) \\ &\quad + (U \text{diag}(\sigma_i^2) V) \otimes (V^T \text{diag}(\sigma_i^2) U^T) \\ &= (V \otimes V)^T [I \otimes \text{diag}(\sigma_i^4) + \text{diag}(\sigma_i^4) \otimes I] (V \otimes V) + (V \otimes U)^T [\text{diag}(\sigma_i^2) \otimes \text{diag}(\sigma_i^2)] (V \otimes U) \\ &\quad + (V \otimes U) [\text{diag}(\sigma_i^2) \otimes \text{diag}(\sigma_i^2)] (V \otimes U)^T. \end{aligned}$$

It is easy to find that the matrices included in $[\cdot]$ of the last equation of $M^T M$ are diagonal. Then, with the help of Lemma 3.1, we can deduce that the ℓ^2 norm $\sigma_{max}(M^T M)$ of M satisfies

$$\sigma_{max}(M^T M) = 2\sigma_{max} \text{diag}(\sigma_i^4) + 2\sigma_{max}(\text{diag}(\sigma_i^2) \otimes \text{diag}(\sigma_i^2)) = 4\|D_0\|_2^4 \leq 4\|D_0\|_\infty^4 = 64.$$

Therefore, we have $\rho(M) \leq 8$.

Theorem 5. Set $\tau = \frac{c}{\lambda}$. Assume that $0 \leq \tau \leq \frac{1}{32}$, then the sequence $\{u^n\}$ generated by Algorithm 2 converges to the solution u^* of (13).

Proof. With the help of $\lambda_{max}(M^T M) < 64$ and Corollary 3.1, it is easy to deduce the assertion.

Now we consider the relationship between Algorithm 3.2 and semi-implicit gradient descent method [8]. Actually, Eq. (16b) can be equivalently viewed as

$$\xi_i^*(Mu^*)_i = |(Mu^*)_i| \text{ and } |\xi_i^*| \leq 1, \text{ for all } i \tag{17}$$

and

$$\xi_i^* = \frac{(Mu^*)_i}{|(Mu^*)_i|} \text{ if } |(Mu^*)_i| \neq 0. \tag{18}$$

By the fact that $\xi_i^*(Mu^*)_i \leq |\xi_i^*| |(Mu^*)_i|$ and $|\xi_i^*| \leq 1$, (17) and (18) can be obviously summarized to

$$\xi_i^* |(Mu^*)_i| = (Mu^*)_i.$$

Hence, (16) can be written as

$$\begin{cases} u_i^* = f_i - \frac{1}{\lambda} (M^T \xi^*)_i, \\ \xi_i^* |(Mu^*)_i| = (Mu^*)_i \end{cases} \tag{19}$$

for $i = 1, 2, \dots, N^2$. We then can get

$$\xi_i^* \left| \left(\lambda M f - (M M^T \xi^*) \right)_i \right| = \left(\lambda M f - (M M^T \xi^*) \right)_i,$$

which can be solved by semi-implicit gradient descent (or fixed point) method [8] as follows

$$\xi_i^{n+1} = \xi_i^n + \beta \left(\left(\lambda M f - (M M^T \xi^n) \right)_i - \left| \left(\lambda M f - (M M^T \xi^n) \right)_i \right| \xi_i^{n+1} \right).$$

So we have the following algorithm to solve the problem (2).

Algorithm 3.3. Semi-implicit Gradient Descent Method (SGDM).

(I) Set $\xi^0 = 0$ and $\gamma > 0$;

(II) Compute (u^n, ξ^{n+1}) by

$$\begin{cases} u^n = f - \frac{1}{\lambda} (M^T \xi^n), \\ \xi_i^{n+1} = \frac{\xi_i^n + \beta (\lambda M f - (M M^T \xi^n))_i}{1 + \beta |(\lambda M f - (M M^T \xi^n))_i|}; \end{cases}$$

(III) If the stop criterion is not satisfied, set $n := n + 1$ and go to step (II).

Theorem 6. Assume that $0 \leq \gamma \leq \frac{1}{64}$. Then the sequence $\{u^n\}$ generated by Algorithm 3.3 converges to the solution u^* of (14).

Remark. Following from the work in [8] and using the fact that $\rho(M) \leq 8$, it is easy to deduce that the above assertion holds. Actually, Algorithm 3.2 can be viewed as a modified form of Algorithm 3.3 [13, 2]. But here we show that Algorithm 3.3 can be naturally deduced from Algorithm 3.2.

Remark. For the color image restoration problems, we can denote that $u \in R^{3N^2}$. Then the Hessian matrix can be written as $\mathbf{M} = (M_r; M_g; M_b)$, here $M_j = M$ given in (15) for $j = r, g, b$. In this case, the minimization problem (14) can be transformed into the following form

$$\min_{\mathbf{u}} \frac{\lambda}{2} \|\mathbf{f} - \mathbf{u}\|_{\ell^2(\Omega)}^2 + |\mathbf{M}\mathbf{u}|_{\ell^1(\Omega)}.$$

Then we can get the similar results as for the gray-level image.

4. Numerical Results

In this section, we present some numerical experiments based on platform by using Windows 7 and Matlab 2009(a) on a desktop with Intel Core i5 processor at 2.4 GHz and 4 GB memory. For Algorithm 3.2, we choose the time step $\tau = 1/32$. In order to illustrate the effectiveness of the proposed algorithms, we show the Signal to Noise Ratio (SNR) of the restored image and the L^2 -norm of the difference between the restored image and the original image. For a given true image u_0 and its noisy observation f , the noise is denoted as $f - u_0$, then SNR and L^2 -norm are defined by

$$SNR = 10 \log \left(\frac{\|f - \bar{f}\|_{\ell^2}^2}{\|\eta - \bar{\eta}\|_{\ell^2}^2} \right) \text{ and } L^2\text{-norm} = \|u - u_0\|_{\ell^2}.$$

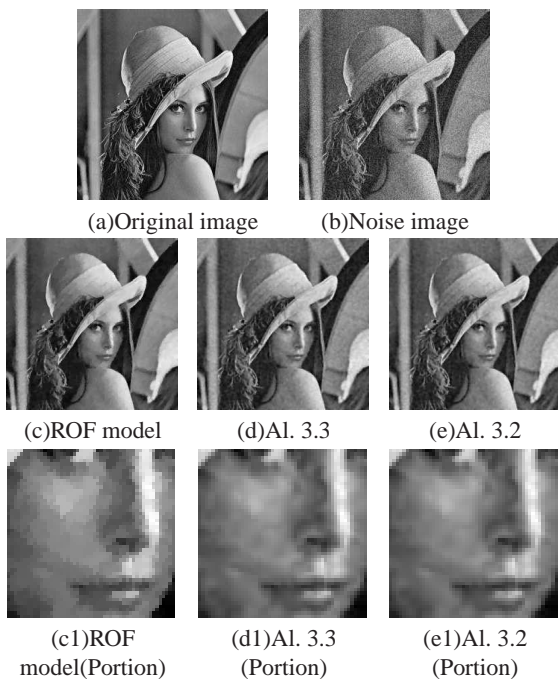


Figure 1 (b) $SNR = 10.0363$; (c) $SNR = 15.7436$; (d) $SNR = 16.0530$; (e) $SNR = 16.1383$.

4.1. Image denoising

In this subsection, we present some numerical examples to compare Algorithm(AI.) 3.2 with Algorithm 3.3.

Example 1. We first consider the Lena image shown in Figure 1 which is contaminated by the random noise with the standard deviation variance $\sigma = 60$. Before processing, the noise image has $SNR = 10.0363$. The algorithms will be terminated if the condition $\|u^{k+1} - u^k\|/\|u^k\| \leq 1.0 \times 10^{-3}$ is met. We compare Algorithm 3.2 with Algorithm 3.3 for recovering the noisy image and set $\lambda = 35$ for these two algorithms. In order to show the ability for overcoming the staircasing phenomenon of the the high-order model, we also employ the semi-implicit gradient descent method to solve the ROF model (1) where we set $\alpha = 12.5$. From Figure 1, it is obvious that the ROF model and the high-order model can recover the noisy image efficiently. As we can see SNR from the notes in Figure 1, Algorithm 3.2 has a better restoration than Algorithm 3.3 and the ROF model. Simultaneously, from the convergence curves of L^2 -norm and SNR , we can deduce that Algorithm 3.2 has faster convergence than Algorithm 3.3. Furthermore, we also notice that the ROF model almost has the same curve in the first steps as Algorithm 3.2 and it has even a higher L^2 -norm than Algorithm 3.3 after some more iterations. Now, we also extend the algorithms to recover the degraded color pepper image. The noise is added the Gaussian noise with the standard variance $\sigma = 30$. The stop conditions is also chose

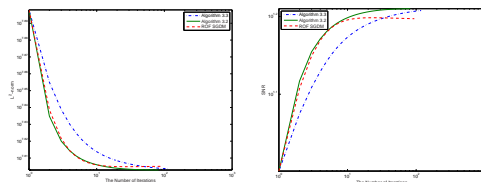


Figure 2 The L^2 -norm curve and SNR curve for Figure 1.

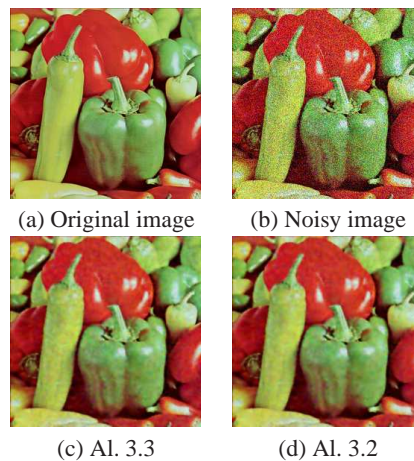


Figure 3 (a) $SNR = 18.4383$; (c) $SNR = 42.1765$; (d) $SNR = 42.6673$.

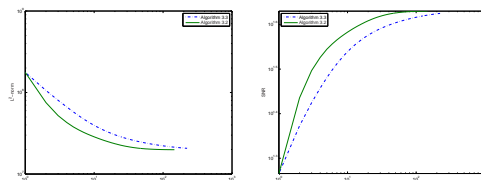


Figure 4 The L^2 -norm curve and SNR curve for Figure 2.

$\|u^{k+1} - u^k\|/\|u^k\| \leq 1.0 \times 10^{-3}$. As shown in the grey-scale case, Algorithm 3.2 and 3.3 can efficiently recover the noisy image and we also get the same conclusions for the convergence curves of two algorithms. The relate images are shown in Figure 3 and Figure 4.

4.2. Other applications

In this subsection, we extend Algorithm 3.2 to solve the general high-order model (3) including image deblurring, image zooming and image inpainting. In order to use this algorithm, we have to transform the high-order model (3) into the formula of the model (2) by using some numerical techniques.



Figure 5 Left: deteriorated image; Right: restored image.

4.2.1. Image deblurring

In this subsection, we consider to solve the deblurring problem (3) by using the following strategy

$$\begin{cases} v^n = u^n + \mu K^*(f - Ku^n), \\ u^{n+1} := \arg \min_u \frac{\gamma}{2} \int_{\Omega} (u - v^n)^2 dx + \mu |u|_{BV^2(\Omega)}, \end{cases} \quad (20)$$

where K^* is the adjoint of the operator K . Based on (20), thus we can employ Algorithm 3.2 to solve the deblurring problem (3).

Remark. For the above strategy, Aujol [2] proposed to solve the image deblurring problem based on the ROF model. He also gave the result that the sequence $\{u^n\}$ converges to the solution of the responding problem when $\mu < \frac{1}{\|K^*K\|}$. Since the high-order model is similar to the ROF model, so we can employ this strategy to the image deblurring image problem. Furthermore, we also get the results that the sequence $\{u^n\}$ converges to the solution of (3) when $\mu < \frac{1}{\|K^*K\|}$.

Example 2. In this example, we use the gray-level and color Lena images as the test images. The related deteriorated images shown in Figure 5 are blurred with a Gaussian kernel of $hsize = 3$ and added the Gaussian white noise with the standard deviation $\sigma = 0.02$ for gray image and $\sigma = 0.05$ for the color image. We choose parameters $\gamma = 0.001$, $\mu = 0.8$ for the gray image and $\gamma = 0.008$, $\mu = 0.02$ for the color image. From the restored images in Figure 5, we can find that the high-order model (3) can efficiently suppresses the blur and noise.

4.2.2. Image zooming

Image zooming is the problem of increasing the resolution of a given image to higher resolution. In this subsection,

we consider to use the model (3) to solve the image zooming problem. Given the low resolution image u_0 of the size $M \times N$, we want to get a zoomed image of size $M \cdot k \times N \cdot k$ for a positive integer k . Usually, we can use bilinear interpolation to extend u_0 to all the $M \cdot k \times N \cdot k$ mesh grid points, but this method can not get the suitable zooming image. So we have to employ other methods to improve the bilinear interpolation image f . Here we can use f as the initial inputting image and then use the higher-order model to update it. That is to say, we can consider the following formulation

$$\min_u \frac{\bar{\gamma}}{2} \int_{\Omega \setminus D} (u - f)^2 dx + |u|_{BV^2(\Omega)}. \quad (21)$$

Here $\bar{\gamma}$ satisfies

$$\bar{\gamma} = \begin{cases} \gamma & \text{if } u \in \Omega \setminus D, \\ 0 & \text{if } u \in D, \end{cases}$$

where D is the zooming domain, which corresponds to the set of the interpolation pixel points. If we introduce an auxiliary variable z , based on the penalty method [4], then the solution of the problem (21) can be approximated by solving the following strategy

$$\begin{cases} u^n := \frac{z^n + \bar{\gamma}\mu f}{1 + \bar{\gamma}\mu}, \\ z^{n+1} := \min_z \frac{1}{2\mu} \int_{\Omega} (z - u^n)^2 dx + |u|_{BV^2(\Omega)}. \end{cases}$$

Example 3. In this example, we use a gray-level synth image and the color strawberries as the test images for the image zooming problem. For the gray-level image, Figure 6 shows the zoomed images obtained by the bilinear interpolation, the ROF model and the high-order model. We use $\gamma = 150$ and $\mu = 0.02$ for the ROF model and $\gamma = 150$ and $\mu = 0.05$ for the high-order model. They have the same size with 100×100 . Among these restored images, we can see the quality of the high-order model has the highest SNR and the most suitable edges. Furthermore, in order to show the efficiency of our proposed method for the color image zooming. The original strawberry has the size of 256×256 . We also use the downsample method to get the deteriorated image shown in Figure 7 (b) and (d) with size of 128×128 and 64×64 . We set the parameter $\lambda = 50$, $\mu = 0.03$ for zooming 2 times and $\lambda = 50$, $\mu = 0.1$ for zooming 4 times. From the zoomed images by a factor of 2 and 4 shown in Figure 7 (c) and (d), we also find that the high-order model can efficiently zooms image.

4.2.3. Image inpainting

For the image inpainting problem, we recover degraded or missing parts denoted D in the image domain Ω by using the proposed method. Not similar to the image zooming problem, here the inpainting domain D is determined by

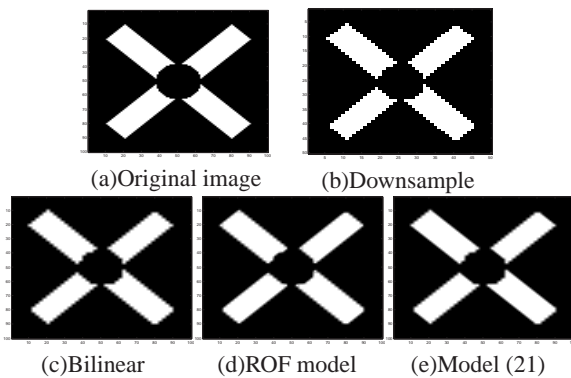


Figure 6 (a) 100×100 ; (b)Downsample image: 50×50 ; (c) $SNR = 12.0064$; (d) $SNR = 12.5176$ (e) $SNR = 12.5294$.

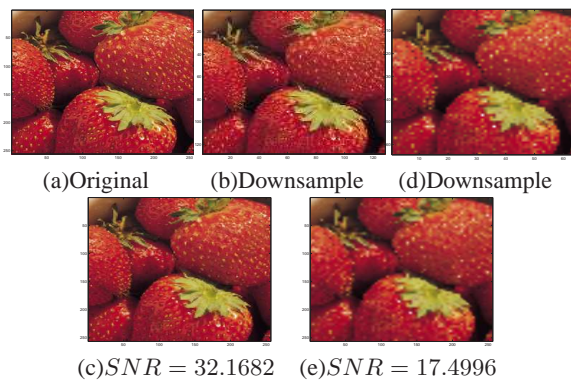


Figure 7 (a) 256×256 ; (b)Downsample image: 128×128 ; (c) $s56 \times 256$; (d)Downsample image: 64×64 ; (e) 256×256 ; the images (c) and (e) restored by the high-order model (21).

the mask image. Based on the model (3), the image inpainting model can be written as

$$\min_u \frac{\bar{\gamma}}{2} \int_{\Omega \setminus D} (u - f)^2 dx + |u|_{BV^2(\Omega)}, \quad (23)$$

where D is the inpainting domain and $\bar{\gamma}$ satisfies

$$\bar{\gamma} = \begin{cases} \gamma & \text{if } u \in \Omega \setminus D, \\ 0 & \text{if } u \in D. \end{cases}$$

By using the penalty method [4] again, the solution of the problem (23) can be approximated by solving the following problem

$$\begin{cases} u^n := \frac{z^n + \bar{\gamma}\mu f}{1 + \bar{\gamma}\mu} \\ z^{n+1} := \min_z \frac{1}{2\mu} \int_{\Omega \setminus D} (z - u^n)^2 dx + |z|_{BV^2(\Omega)}. \end{cases}$$

Example 4. In this example, we also consider the gray-level image and the color image shown in Figure 8 and 9. We

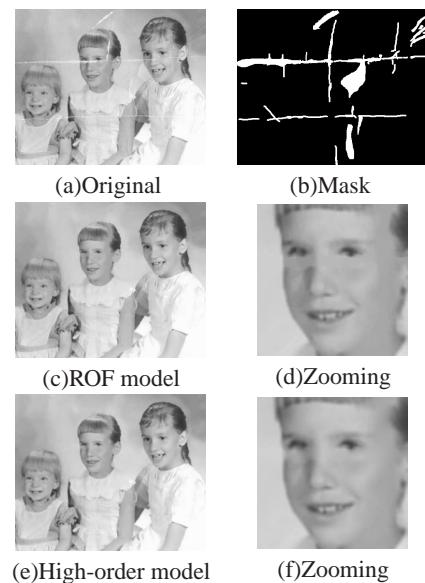


Figure 8 (d) The zooming image corresponds to the ROF model; (f)The zooming image corresponds to the high-order model;



Figure 9 The related images in example 4.4.

use $\gamma = 150$, $\mu = 0.005$ for the gray tested image and the color tested image. The algorithm will be stopped until the iteration attains 2000 times. It is obvious to see from Figure 8 and 9 that the high-order model can efficiently restore the deteriorated image. Furthermore, for comparison, we also show the restored gray image by the ROF model. It is not difficult to find that the restored image by the high-order model looks more natural than that restored by the ROF model. In fact, this is because high-order linear or nonlinear diffusion damps oscillations much faster than the second order diffusion.

5. Conclusion

Based on the augmented Lagrangian strategy and by introducing a smooth approximation function, we proposed a projected gradient algorithm to solve the high-order model in image restoration problem. The convergence of the proposed algorithm is proved based on the BM algorithm.

Furthermore, we give the fact that the semi-implicit gradient descent algorithm in [8] can be deduced from our proposed algorithm naturally. The numerical experiments indicate that our proposed algorithm is efficient not only for the image denoising problem but also for other image processing topics such as image deblurring, image zooming and image inpainting. However, as we can see from examples, the restored images have some edges blurring effect for the high-order model. So for the future work we mainly look for more suitable models to improve this drawback.

Acknowledgement

The authors acknowledge the financial support by NNSF of China (11071060) and the university research fund of Henan University (2011YBZR003). The author is grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

References

- [1] G. Aubert and P. Kornprobst. Mathematical problems in image processing: partial differential equations and the calculus of variations. Springer, 2002.
- [2] J. Aujol. Some first-order algorithms for total variation based image restoration. *Journal of Mathematical Imaging and Vision*, 34(3)(2009), 307-327.
- [3] A. Bermudez and C. Moreno. Duality methods for solving variational inequalities. *Computers & Mathematics with Applications*, 7(1)(1981), 43-58.
- [4] D. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Athena Scientific, 1996.
- [5] A. Chambolle and P. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2)(1997), 167-188.
- [6] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1)(2004), 89-97.
- [7] A. Chambolle. Total variation minimization and a class of binary MRF models. In *EMMCVPR 05*, volume 3757 of *Lecture Notes in Computer Sciences*, (2005), 136-152.
- [8] H. Chen, J. Song and X. Tai. A dual algorithm for minimization of the LLT model. *Advances in Computational Mathematics*, 31(1-3)(2009), 115-130.
- [9] T. Chan and J. Shen. *Image Processing and Analysis-Variational, PDE, Wavelet, and Stochastic Methods*. SIAM, 2005.
- [10] T. Chan, A. Marquina, and P. Mulet. High-order total variation-based image restoration. *SIAM Journal on Scientific Computing*, 22(2)(2000), 503-516.
- [11] I. Ekeland and T. Turbull. *Infinite dimensional optimization and convexity*. The university of Chicago press, Chicago, 1983.
- [12] R. Gao, J.-P. Song, and X.-C. Tai. Image zooming algorithm based on partial differential equations technique. *International Journal of Numerical Analysis and Modeling*, 6(2)(2009), 284-292.
- [13] R. Glowinski and P. Le Tallec. *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*. SIAM, 1989.
- [14] K. Ito and K. Kunisch. Augmented Lagrangian methods for nonsmooth, convex optimization in Hilbert space. *Nonlinear Analysis*, 41(5-6)(2000), 591-616.
- [15] K. Ito and K. Kunisch. An active set strategy based on the augmented Lagrangian formulation for image restoration. *ESAIM: Mathematical Modelling and Numerical Analysis*, 33(1)(1999), 1-21.
- [16] K. Jonas and J.-B. Stéphanie. An augmented Lagrangian method for $TV_g + L^1$ -norm minimization. *Journal of Mathematical Imaging and Vision*, 38(3)(2010), 182-196.
- [17] 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 Transactions on Image Processing*, 12(12)(2003), 1579-1590.
- [18] N. Paragios, Y. Chen, and O. Faugeras. *Handbook of Mathematical Models in Computer Vision*. Springer, 2005.
- [19] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1992), 259-268.
- [20] Z.-F. Pang and Y.-F. Yang. Semismooth Newtons methods for minimization of the LLT model. *Inverse problems and Imaging*, 3(4)2009, 677-691.
- [21] G. Steidl. A note on the dual treatment of higher order regularization functionals. *Computing*, 76(2006) 135-148.
- [22] O. Scherzer. Denoising with higher order derivatives of bounded variation and an application to parameter estimation. *Computing*, 60(1998), 1-27.
- [23] X.-C. Tai and C. Wu. Augmented Lagrangian Method, Dual Methods and Split Bregman Iteration for ROF Model. *The second SSVM*, (2009), 502-513.
- [24] C. Wu, and X.-C. Tai. Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models. *SIAM Journal on Imaging Sciences*, 3(3)(2010), 300-339.
- [25] Y. You and M. Kaveh. Fourth-order partial differential equation for noise removal. *IEEE Transactions on Image Processing*, 9(10)(2000), 1723-1730.
- [26] Y. You and M. Kaveh. Blind image restoration by anisotropic regularization. *IEEE Transactions on Image Processing*, 8(3)(1999), 296-407.
- [27] F. Zhang. *Matrix Theory: Basic Results and Techniques*. Springer, 1999.



Baoli Shi received her PhD in mathematics from Hunan University in 2012, Changsha, China. She is now an associate professor in the College of Mathematics and Information Science at Henan University, Kaifeng, China. Her research interests include the inverse problems, the numerical method of PDE, fast numerical methods in im-

age processing.



Zhi-Feng Pang received his PhD in mathematics from Hunan University in 2010, Changsha, China. He is now an associate professor in the College of Mathematics and Information Science at Henan University, Kaifeng, China. Between 2010 and 2012, he was a Postdoctoral Fellow with Nanyang Technological University, Singapore and City University of Hongkong, Hongkong, respectively. His research interests include the mathematical theory and fast numerical methods in image processing and machine learning.



Lihong Huang received his PhD in mathematics from Hunan University in 1994, Changsha, China. He is now a professor in the College of Mathematics and Econometrics at Hunan University and Hunan Women's University, China. He is the author or co-author of more than 100 journal papers, six edited books. His research interests are in the areas of dynamics of neural networks, and qualitative theory of differential equations and difference equations.