J Math Imaging Vis (2009) 35: 143–154
DOI 10.1007/s10851-009-0160-3

On Nonmonotone Chambolle Gradient Projection Algorithms for Total Variation Image Restoration

Gaohang Yu · Liqun Qi · Yuhong Dai

Published online: 16 June 2009

Abstract The main aim of this paper is to accelerate the Chambolle gradient projection method for total variation image restoration. In the proposed minimization method model, we use the well known Barzilai-Borwein stepsize instead of the constant time stepsize in Chambolle's method. Further, we adopt the adaptive nonmonotone line search scheme proposed by Dai and Fletcher to guarantee the global convergence of the proposed method. Numerical results illustrate the efficiency of this method and indicate that such a nonmonotone method is more suitable to solve some large-scale inverse problems.

Keywords Image restoration · Total variation · Gradient projection · Constrained optimization

This work was partly supported by the Research Grant Council of Hong Kong, a postdoctoral fellowship from the Department of Applied Mathematics at the Hong Kong Polytechnic University (1-ZV0K), a grant from the Ph.D. Programs Foundation of Ministry of Education of China (No.200805581022) and the National Natural Science Foundation of China (No.10571171, 10831006 and 60804008), and the CAS grant kjcx-yw-s7-03.

G. Yu (✉)
Key Laboratory of Numerical Simulation Technology, School of Mathematics and Computer Sciences, GanNan Normal University, Ganzhou 341000, Jiangxi, China
e-mail: maghyu@163.com

Present address:
G. Yu
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

L. Qi
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
e-mail: maqilq@inet.polyu.edu.hk

Y. Dai
LSEC, Institute of Computational Mathematics, Chinese Academy of Sciences, P.O. Box 2719, Beijing 100190, China
e-mail: dyh@lsec.cc.ac.cn

1 Introduction

Total variation (TV) models have achieved a great success in image processing. They have been used in many applications such as image restoration, image deblurring, image inpainting (see [2, 13] and references therein). Image restoration, especially for image denoising, forms a significant preliminary step in many machine vision tasks such as object detection and recognition. The typical image restoration model was first introduced by Rudin, Osher and Fatemi (ROF) in [24]. The ROF model could preserve sharp discontinuities (edges) in an image while removing noise by the following minimization of a functional:

$$\min_u P(u) := \int_\Omega |\nabla u| dx + \frac{\lambda}{2} \|u - f\|_2^2. \tag{1}$$

Here ∇ is the gradient operator and $\int_\Omega |\nabla u| dx$ stands for the total variation of u. Ω is the domain of definition of the image, which is assumed to bounded in $\Re^n$ with Lipschitz boundary. Usually Ω is simply a rectangle in $\Re^2$. $f : \Omega \to \Re$ is the degraded image to restore. The minimizer u of (1) is the restored image we want to compute (see for instance [11] for a thorough mathematical analysis of this problem). λ is a weighting parameter which controls the amount of denoising. $|\cdot|$ represents the Euclidean (ℓ_2) norm on $\Re^2$. The total variation of u has the following equivalent dual form:

$$\int_\Omega |\nabla u| dx = \max_{w\in C_c^1(\Omega),|w|\le 1} \int_\Omega \nabla u \cdot w$$
$$= \max_{|w|\le 1} \int_\Omega -u\nabla \cdot w, \tag{2}$$

where $w : \Omega \to \Re^2$ is the dual variable, and $\nabla\cdot$ is the divergence operator. The idea of duality has been proposed first by Chan, Golub and Mulet [14], later by Carter [8] and Chambolle in [9, 10], and then by Zhu et al. [27–29]. In [14], the authors applied Newton's method to solve the primal-dual system of the ROF model. So, their method was shown to have a locally quadratic convergence rate. In [22], a non-smooth Newton method is considered and the achieved convergence is superlinear. In [12], the authors presented some fast multigrid methods to solve the dual ROF model. In [29], the authors developed duality-based gradient projection algorithms and sequential quadratic programming algorithms for total variation image restoration problems. In [9, 10], Chambolle proposed some gradient descent algorithms, which become popular for their simplicity and fast convergence to medium-accurate visually satisfactory solutions. There are so many total variation minimization algorithms based on Chambolle's method (e.g. [1, 3, 21]). In practice, a relatively large number of iterations is still required. In some case, the iterates of Chambolle gradient descent algorithm slowly approach the minimum, especial for some very ill conditioned problems. Maybe, the most disadvantage of Chambolle method is that it requires the functional value to decrease monotonically at each iteration.

In this paper, we extend Chambolle's method to nonmonotone descent method for total variation minimization. We use the well known Barzilai-Borwein stepsize [4] instead of the constant stepsize in Chambolle's projection algorithm and adopt the adaptive nonmonotone line search scheme proposed by Dai and Fletcher [16] to guarantee the global convergence. Numerical results illustrate the efficiency of this scheme and indicate that such a nonmonotone method is more suitable to solve some large-scale inverse problems. In Sect. 2, we review Chambolle's projection method. Our nonmonotone descent gradient method is presented in Sect. 3. In Sect. 4, numerical experiments are given to illustrate the convergence and efficiency of the proposed method. Finally, we have a conclusion section.

2 The Chambolle Projection Method

In this section we give a review on the Chambolle projection method for total variation minimization [9]. We first take some notation and definitions in [9]. We assume that our images are matrices of size $N \times N$. We denote by $\mathcal{X}$ the Euclidean space $\Re^{(N\times N)}$, and $\mathcal{Y} = \mathcal{X} \times \mathcal{X}$. The space $\mathcal{X}$ and $\mathcal{Y}$ will be endowed with the scalar product, defined in the standard way respectively by

$$\langle s,t\rangle_{\mathcal{X}} = \sum_{1\le i,j\le N} s_{i,j} t_{i,j}, \quad s,t \in \mathcal{X}$$

and

$$\langle p,q\rangle_{\mathcal{Y}} = \sum_{1\le i,j\le N} p^1_{i,j} q^1_{i,j} + p^2_{i,j} q^2_{i,j},$$
$$p = (p^1, p^2),\ q = (q^1, q^2) \in \mathcal{Y}.$$

To define a discrete total variation, we introduce a discrete version of the gradient operator $\nabla : \mathcal{X} \to \mathcal{Y}$, which is defined by

$$(\nabla u)_{i,j} = \left((\nabla u)^1_{i,j}, (\nabla u)^2_{i,j}\right)$$

with

$$(\nabla u)^1_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j} & \text{if } i < N, \\ 0 & \text{if } i = N, \end{cases}$$

$$(\nabla u)^2_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j} & \text{if } j < N, \\ 0 & \text{if } j = N \end{cases}$$

for $i, j = 1, \ldots, N$. The discrete total variation of u is then defined by

$$TV(u) := \sum_{1\le i,j\le N} |(\nabla u)_{i,j}|.$$

Therefore, the discretization of minimization problem (1) is given by

$$\min_{u\in\mathcal{X}} TV(u) + \frac{\lambda}{2}\|u - f\|^2_{\mathcal{X}}. \tag{3}$$

Here $u, f \in \mathcal{X}$ are discretization vectors of related continuous variables, and $\|\cdot\|_{\mathcal{X}}$ is the Euclidean norm in $\mathcal{X}$, given by $\|u\|^2_{\mathcal{X}} = \langle u,u\rangle_{\mathcal{X}}$. Chambolle showed in [9] that the solution u of problem (3) should be given by

$$u = f + \pi_{\frac{1}{\lambda}\mathcal{K}}(-f), \tag{4}$$

where $\pi_{\frac{1}{\lambda}\mathcal{K}}(\cdot)$ is the orthogonal projection onto the convex set $\frac{1}{\lambda}\mathcal{K}$ with

$$\mathcal{K} := \{\nabla \cdot w : w \in \mathcal{Y}, |w_{i,j}| \le 1,\ \forall i,j = 1,\ldots,N\}$$

and the discrete divergence operator $\nabla\cdot : \mathcal{Y} \to \mathcal{X}$ is defined by $\nabla\cdot = -\nabla^*$, that is, for any $w \in \mathcal{Y}$ and $u \in \mathcal{X}$, $\langle -\nabla \cdot$

$w, u\rangle_{\mathcal{X}} = \langle w, \nabla u\rangle_{\mathcal{Y}}$. It is easy to check that the divergence operator can be defined explicitly as follows:

$$(\nabla \cdot w)_{i,j} = \begin{cases} w^1_{i,j} - w^1_{i-1,j} & \text{if } 1 < i < N, \\ w^1_{i,j} & \text{if } i = 1, \\ -w^1_{i-1,j} & \text{if } i = N, \end{cases} + \begin{cases} w^2_{i,j} - w^2_{i,j-1} & \text{if } 1 < j < N, \\ w^2_{i,j} & \text{if } j = 1, \\ -w^2_{i,j-1} & \text{if } j = N. \end{cases}$$

Hence, computing the solution of (3) hinges on computing the nonlinear projection $\pi_{\frac{1}{\lambda}\mathcal{K}}(-f)$, which amounts to solving the following constrained minimization problem:

$$\min_{w\in K} \left\| f + \frac{1}{\lambda}\nabla \cdot w \right\|^2_{\mathcal{X}} \quad \text{with}$$
$$K := \{w : w \in \mathcal{Y}, |w_{i,j}|^2 \leq 1, \ \forall i, j = 1, \ldots, N\}. \tag{5}$$

In [8, 27], Carter and Zhu et al derived the constrained optimization problem (5) by a different way. With the definition of total variation (2), the primal ROF model (1) becomes

$$\min_u \max_{w\in C^1_c(\Omega), |w|\leq 1} \int_\Omega -u\nabla \cdot w + \frac{\lambda}{2}\|u - f\|^2.$$

Interchanging the min and max (see e.g., Proposition 2.4 in [18]), we obtain

$$\max_{w\in C^1_c(\Omega), |w|\leq 1} \min_u \int_\Omega -u\nabla \cdot w + \frac{\lambda}{2}\|u - f\|^2.$$

The inner minimization problem can be solved exactly as follows

$$u = f + \frac{1}{\lambda}\nabla \cdot w. \tag{6}$$

We eliminate the primal variable u by the above formula. Then we are left with the dual problem

$$\max_{w\in C^1_c(\Omega), |w|\leq 1} D(w) := \frac{\lambda}{2}\left[\|f\|^2 - \left\| f + \frac{1}{\lambda}\nabla \cdot w \right\|^2\right], \tag{7}$$

or, equivalently, $\min_{|w|\leq 1} \frac{1}{2}\|\lambda f + \nabla \cdot w\|^2$. The main advantage of the dual formulation (5) or (7) is that the objective function is nicely quadratic and hence there is no issue with non-differentiability as in the primal formulation. However, the dual problem needs to deal with the constraints on the dual variable w.

In [9], Chambolle analyzed the Karush-Kuhn-Tucker conditions for the dual problem (5) and he derived that the Lagrange multiplier $\alpha \in \mathcal{X}$ satisfies

$$\alpha_{i,j} = |(\nabla(\nabla \cdot w + \lambda f))_{i,j}| \quad \forall i, j = 1, \ldots, N.$$

Then the Euler-Lagrange equation of the problem (5) can be presented as follows

$$-\big(\nabla(\nabla \cdot w + \lambda f)\big)_{i,j} + |(\nabla(\nabla \cdot w + \lambda f))_{i,j}| w_{i,j} = 0 \quad \forall i, j. \tag{8}$$

He then proposed a semi-implicit gradient descent algorithm:

$$w^{k+1} = w^k - \tau\big(-\nabla(\nabla \cdot w^k + \lambda f) + |\nabla(\nabla \cdot w^k + \lambda f)| w^{k+1}\big) \tag{9}$$

or

$$w^{k+1}_{i,j} = \frac{w^k_{i,j} + \tau\big(\nabla(\nabla \cdot w^k + \lambda f)\big)_{i,j}}{1 + \tau\big|\big(\nabla(\nabla \cdot w^k + \lambda f)\big)_{i,j}\big|}, \quad \forall i, j, \tag{10}$$

where $\tau > 0$ is a time step chosen suitably small for convergence. It is clear that the constraints $|w| \leq 1$ are automatically handled in Chambolle's projection algorithm provided that the initial guess satisfies so. In [9], Chambolle gave a sufficient condition which ensuring the convergence of the algorithm.

Theorem 1 *Suppose that the parameter* $0 < \tau \leq 1/8$. *Then* $\frac{1}{\lambda}\nabla \cdot w^k$ *converges to* $\pi_{\frac{1}{\lambda}\mathcal{K}}(-f)$ *as* $k \to \infty$.

In practice, convergence of the Chambolle projection algorithm is generally observed as long as $\tau < 1/4$. An extension of this algorithm to color images has been proposed in [7]. Instead of using (10), Chambolle suggested in [10] to use a simple projected gradient descent method:

$$w^{k+1}_{i,j} = \frac{w^k_{i,j} + \tau\big(\nabla(\nabla \cdot w^k + \lambda f)\big)_{i,j}}{\max\{1, |w^k_{i,j} + \tau(\nabla(\nabla \cdot w^k + \lambda f))_{i,j}|\}} \quad \forall i, j. \tag{11}$$

In [10], Chambolle proved the stability of (11). Application of basic results about the projected gradient algorithm [5] shows that in fact (11) is convergent provided $0 < \tau < 1/4$. Experiments show that $\tau = 1/4$ seems not optimal, and a better convergence is obtained for $\tau = 0.248$. Both Chambolle's semi-implicit gradient descent algorithm (9) and Chambolle's projected gradient algorithm (11) require the functional value to decrease monotonically at each iteration, which made the iterates of Chambolle's methods slowly approach the minimum when the minimization problem is very ill conditioned.

3 The Nonmonotone Chambolle Projection Algorithms

Motivated to improve Chambolle projection method, we replace the constant time step τ by the well-known Barzilai-Borwein stepsize [4]. In the rest of this paper, we report on the development and implementation of some nonmonotone gradient projection algorithms for the image restoration problem.

Given a starting point w^0 and using the notation $g^k = \nabla F(w^k)$, the gradient methods for $\min_{w\in\Re^n} F(w)$ are defined by the iteration $w^{k+1} = w^k - t^k g^k, \quad k = 0, 1, \ldots,$ where the stepsize $t^k > 0$ is determined through an appropriate selection rule. In the classical steepest descent (SD) method, the stepsize $t^k > 0$ is obtained by minimizing the function $F(w)$ along the ray $\{w^k - tg^k : t > 0\}$. It is well-known that the SD method can be very slow when the Hessian of $F(w)$ is ill-conditioned at a local minimum. In this case, the iterates slowly approach the minimum in a zigzag fashion. Early efforts to improve the SD method gave rise to the development of the conjugate gradient (CG) method. In 1988, Barzilai and Borwein [4] developed an ingenious gradient method in which stepsize $t^k (k > 0)$ is determined by:

$$t^k_{BB} = \frac{\langle s^{k-1}, s^{k-1}\rangle}{\langle y^{k-1}, s^{k-1}\rangle}, \tag{12}$$

where $s^{k-1} = w^k - w^{k-1}$ and $y^{k-1} = g^k - g^{k-1}$. In fact, t^k is derived from an approximately secant equation: $t^k_{BB} = \arg\min_{t\in\Re} \|\frac{1}{t}s^{k-1} - y^{k-1}\|$. The BB method performs much better than the SD method in practice. Especially, when the objective function is a convex quadratic function and $n = 2$, a sequence generated by the BB method converges R-superlinearly to the global minimizer [4]. For any dimension convex quadratic function, it is still globally convergent [23] but the convergence is R-linear [17].

Let $w \in \mathcal{Y}$ and $F(w) := \frac{1}{2}\|\lambda f + \nabla\cdot w\|^2_{\mathcal{X}}$. Then $g^k := \nabla F(w^k) = -\nabla(\nabla\cdot w^k + \lambda f)$. So, we have $y^{k-1} = -\nabla(\nabla\cdot s^{k-1})$ and $\langle y^{k-1}, s^{k-1}\rangle_{\mathcal{Y}} = \langle -\nabla(\nabla\cdot s^{k-1}), s^{k-1}\rangle_{\mathcal{Y}} = \langle -(\nabla\cdot s^{k-1}), -(\nabla\cdot s^{k-1})\rangle_{\mathcal{X}}$. And then (12) becomes as follows

$$t^k_{BB} = \frac{\langle s^{k-1}, s^{k-1}\rangle}{\langle y^{k-1}, s^{k-1}\rangle} = \frac{\|s^{k-1}\|^2_{\mathcal{Y}}}{\|\nabla\cdot s^{k-1}\|^2_{\mathcal{X}}}. \tag{13}$$

Using the above formula to compute the time step τ in the Chambolle projection method, we obtain two corresponding nonmonotone Chambolle projection algorithms as follows.

Nonmonotone Chambolle semi-implicit gradient projection algorithm:

$$w^{k+1}_{i,j}(w^k, t^k, g^k) := \frac{w^k_{i,j} - t^k g^k_{i,j}}{1 + t^k|g^k_{i,j}|}, \quad \forall i, j = 1, \ldots, N \quad \text{with } t^k = t^k_{BB}. \tag{14}$$

Nonmonotone Chambolle projected gradient algorithm:

$$w^{k+1}_{i,j}(w^k, t^k, g^k) := \frac{w^k_{i,j} - t^k g^k_{i,j}}{\max\{1, |w^k_{i,j} - t^k g^k_{i,j}|\}}, \quad \forall i, j = 1, \ldots, N \quad \text{with } t^k = t^k_{BB}. \tag{15}$$

The above two nonmonotone Chambolle gradient projection algorithms cannot ensure the objective functional value decrease monotonically at each iteration. Therefore, in order to ensure global convergence, it is necessary to modify the above two algorithms by incorporating some sort of nonmonotone line search [16, 20]. In [16], Dai and Fletcher proposed an adaptive nonmonotone line search. The numerical results reported in [16] show that this kind of line search is particularly suitable for BB-like methods in the non-quadratic case. The method has a reference function value F_r, and each iteration must improve on the reference value such that:

$$F\big(w^{k+1}(w^k, \beta t^k_{BB}, g^k)\big) \le F_r + \theta\beta\langle g^k, d^k\rangle_{\mathcal{Y}}, \tag{16}$$

where $d^k := w^{k+1}(w^k, t^k_{BB}, g^k) - w^k$ denotes the current search direction, $\theta \in (0, 1)$ is a given constant and $t > 0$ is the tried stepsize. During the line search procedure we need to try a decreasing sequence of values of β, starting with $\beta = 1$, until the test condition (16) is satisfied. Let us denote by F_{best} the current least value of the objective function over all past iterates, that is, at the k-th iteration $F_{best} = \min_{1\le i\le k} F(w^i)$. The number of iterations since the value of F_{best} was obtained is denoted by l. Also we define the candidate function value F_c to be the maximum value of the objective function since the value of F_{best} was found.

Now let us describe how to determine the reference function F_r. Suppose that L is a preset positive integer. Initially, we can set $F_r = +\infty$. This choice of F_r allows $F(w^k) \ge F(w^0)$ on early iterations. If the method can find a better function value in L iterations, then the value of F_r remains unchanged. Otherwise, if $l = L$, we reset the reference function value F_r to F_c and reset F_c to the current value $F(w^k)$. A more precise statement is presented at the Step 6 of the NTVM algorithm.

Let $P(g^k)$ defined by

$$P(g^k) = w^{k+1}(w^k, 1, g^k) - w^k, \tag{17}$$

where $w^{k+1}(w^k, 1, g^k)$ was defined by (14) or (15) with $t^k := 1$. Then if $||P(g^k)|| = 0$, we declare that w^k is a constrained stationary point. We can show the following property.

Proposition 1 *If $||P(g^k)|| = 0$, then w^k satisfies the Euler-Lagrange equation (8).*

Proof In the case where $w^{k+1}(w^k, 1, g^k)$ is computed by (14), we have $\frac{w_{ij}^k - g_{ij}^k}{1+|g_{ij}^k|} - w_{ij}^k = 0,\ \forall i, j$. So, Euler-Lagrange equation (8) holds. On the other hand, If $w^{k+1}(w^k, 1, g^k)$ is computed by (15), then we have

$$w_{ij}^k - g_{ij}^k - \max\{1, |w_{ij}^k - g_{ij}^k|\} w_{ij}^k = 0, \quad \forall i, j.$$

When $|w_{ij}^k - g_{ij}^k| \le 1$, we can obtain $g_{ij}^k = 0$. Otherwise, if $|w_{ij}^k - g_{ij}^k| > 1$ while $g_{ij}^k \neq 0$, we can derive that

$$w_{ij}^k = -\frac{1}{|w_{ij}^k - g_{ij}^k| - 1} g_{ij}^k.$$

Set $\frac{1}{|w_{ij}^k - g_{ij}^k|-1} = t$, and then we have $w_{ij}^k = -t g_{ij}^k$. Combining $\frac{1}{|w_{ij}^k - g_{ij}^k|-1} = t$ with $w_{ij}^k = -t g_{ij}^k$, it follows that $t(t+1)|g_{ij}^k| - t = 1$. Since $g_{ij}^k \neq 0$ and $t > 0$, we can get that $t = \frac{1}{|g_{ij}^k|}$. So, $-g_{ij}^k + |g_{ij}^k| w_{ij}^k = 0$. Therefore, the Euler-Lagrange equation (8) holds for either case. □

Then our proposed method model can be presented as follows.

NTVM algorithm—nonmonotone total variation minimization algorithms

Step 1: Given $w^0 \in \mathcal{Y}$, $t^0 > 0$, integer $L \ge 0$, $\theta \in (0,1)$, $0 < \rho_{min} < \rho_{max}$, $0 < \lambda_1 < \lambda_2 < 1$. Set $F_r = +\infty$, $F_{best} = F_c = F(w^0)$. Set $k := 0$.

Step 2: If $||P(g^k)|| = 0$, then stop.

Step 3: Impose t^k such that $t^k \in [\rho_{min}, \rho_{max}]$; Compute $w^{k+1}(w^k, t^k, g^k)$ by (14) (or (15)). Set $\beta_k = 1$ and $d^k := w^{k+1}(w^k, t^k, g^k) - w^k$.

Step 4: If $F(w^{k+1}(w^k, \beta_k t^k, g^k)) \le F_r + \theta \beta_k \langle g^k, d^k \rangle_{\mathcal{Y}}$, then define $t^k := \beta_k t^k$, update $w^{k+1} := w^{k+1}(w^k, t^k, g^k)$, and go to Step 6.

Step 5: Choose $\sigma \in [\lambda_1, \lambda_2]$, set $\beta_k = \sigma \beta_k$, and go to Step 4.

Step 6: Update the reference function F_r:
If $F(w^{k+1}) \le F_{best}$, **then**
$F_{best} = F^{k+1}$, $F_c = F^{k+1}$, $l = 0$.
Else $F_c = \max\{F_c, F^{k+1}\}$, $l = l + 1$,
if $l = L$, **then** $F_r = F_c$, $F_c = F^{k+1}$, $l = 0$.
End If

Step 7: Set $k := k + 1$. Compute t^k by the formula (13). Go to Step 2.

Let $F(w^i) = \max\{F(w^1), F(w^2), \ldots, F(w^{k_0})\}$, where k_0 denotes the first iteration on which $l = L$. From the algorithm, we know that the iteration sequence w^k remains in the level set $S = \{w \in \mathcal{Y} | F(w) < F(w^i)\}$ for all $k > 0$. We can show the global convergence for the NTVM algorithm as discussed in [16]. It can be seen that if F_{best} is updated an infinite number of times then global convergence occurs. Assume the contrary that F_{best} is unchanged for all k sufficiently large. In this case there exists an infinite subsequence of iterations k_i on which $l = L$ and F_r is reset to F_c. Now $F_c < F_r$ because F_c is a recent value of F^k for which $F^k < F_r$. Thus values of F_r that are reset on iteration k_i are strictly monotonically decreasing. Hence there exists a subsequence on which F^k decreases without bound, which contradicts the fact that F_{best} is unchanged. Recall that the objective function $F(w)$ is bounded below and $g(w)$ is Lipschitz continuous (the Lipschitz constant is bounded by 8, see [9], p. 92), we have the following theorem. For completeness, we provide its rigorous proof in the appendix.

Theorem 2 *Let $\{w^k\}$ be a sequence generated by the NTVM algorithm. Then any accumulation point of the sequence $\{w^k\}$ is a constrained stationary point.*

The following property can be found in [29].

Proposition 2 *Let $\{w^k\}$ be any sequence with $w^k \in K$ for all $k = 1, 2, \ldots$ such that all accumulation points of $\{w^k\}$ are stationary points of (5). Then the sequence $u^k := f + \frac{1}{\lambda}\nabla \cdot w^k$ converges to the unique solution u^* of (3) as $k \to \infty$.*

So, combining Theorem 2 with Proposition 2, we have the following global convergence result.

Corollary 1 *Let $\{w^k\}$ be a sequence generated by the NTVM algorithm. Then there exists a subsequence $\{w^{k_n}\}$ such that $u^{k_n} := f + \frac{1}{\lambda}\nabla \cdot w^{k_n}$ converges to the unique solution u^* of (3) as $n \to \infty$.*

4 Numerical Experiments

4.1 Comparisons to Chambolle Gradient Methods

In this subsection, the numerical results are presented to demonstrate the comparison results of our proposed nonmonotone total variation minimization (NTVM) algorithms with Chambolle gradient projection algorithms for image denoising. The simulations are preformed in Matlab 7.4 (R2007a) on a PC with an Intel Core 2 Duo CPU at 3.0 GHz and 2 GB of memory. The noisy images are generated by adding Gaussian noise to the clean images using the MATLAB function *imnoise*, with variance parameter set to 0.01. The fidelity parameter λ is taken to be 0.053 throughout the experiments. In the NTVM algorithm, we fix $t^0 = 1/\|g(w^0)\|$, $L = 5$, $\theta = 10^{-4}$, $\rho_{max} = 1/\rho_{min} = 10^{10}$, $\sigma = 0.5$. To assess the restoration performance qualitatively, we use the PSNR (peak signal to noise ratio) defined as

$$PSNR = 10 \log_{10} \frac{255^2}{\frac{1}{MN}\sum_{i,j}(u_{i,j}^r - u_{i,j}^*)^2},$$

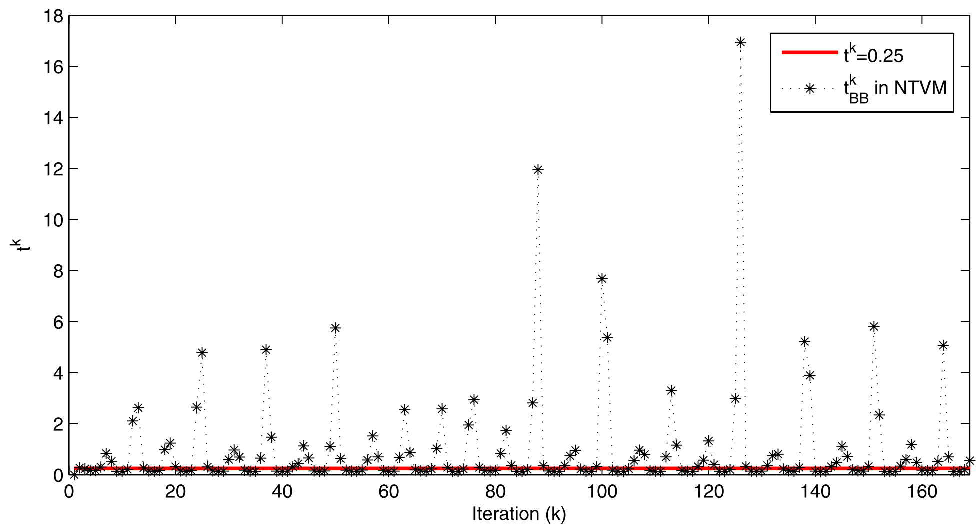


Fig. 1 The value of t_{BB}^k in the denoising for "shape" by NTVM algorithm

where $u_{i,j}^r$ and $u_{i,j}^*$ denote the pixel values of the restored image and the original image, respectively.

We test the following four algorithms:

- **Chambolle:** Chambolle's semi-implicit gradient descent method (10) with $\tau = 0.248$.
- **PGCS:** Chambolle's **p**rojected **g**radient method (11) with the **c**onstant **s**tep $\tau = 0.248$.
- **NChambolle:** Nonmonotone Chambolle's semi-implicit gradient projection method with the adaptive nonmonotone line search (16), i.e. the NTVM algorithm with $w^{k+1}(\cdot)$ computed by (14).
- **NTVM:** Nonmonotone Chambolle's projected gradient method with the adaptive nonmonotone line search (16), i.e. the NTVM algorithm with $w^{k+1}(\cdot)$ computed by (15).

The stopping criterion of these four methods are

$$||P(g^k)|| \le 10^{-6}||P(g^0)||.$$

First we check the sequence of $\{t_{BB}^k\}$ generated by the NTVM algorithm for denoising "shape" image and we observe from the Fig. 1 that most of the values of t_{BB}^k less than 0.25 and there does not exist four sequential points with $t_{BB}^k > 0.25$.

In order to test the speed of the algorithms more fairly, the experiments are repeated for 10 different random noise samples of each image and the average of the 10 results is given in the Table 1. We report the number of iterations (Niter) and the CPU time (in second) required for the whole denoising process and the PSNR of the recovered image.

Notice from the Table 1 that the NTVM method is about three times faster than the Chambolle method while the PSNR values attained by these four methods are very similar. Figures 2 and 3 display the nonmonotone descent behavior for the NChambolle and NTVM method which indicate that such a kind of nonmonotone methods is a great improvement over the original method of Chambolle, especially for some large-size images. Figure 4 displays the restoration results by the NTVM method. The numerical experiments show that the NTVM method can restore corrupted image quite well in an efficient manner.

Table 1 Numerical results

Image	Algorithm	Niter	PSNR	CPU time
shape	Chambolle	709.7	23.854	2.627
128 × 128	NChambolle	368.5	23.854	1.641
	PGCS	565.5	23.854	2.180
	NTVM	176.7	23.854	0.816
lena	Chambolle	355.8	27.867	7.938
256 × 256	NChambolle	190.9	27.866	5.431
	PGCS	274.2	27.864	6.328
	NTVM	98.4	27.864	2.895
brain MRI	Chambolle	585.1	29.410	65.537
512 × 512	NChambolle	303.5	29.409	44.006
	PGCS	465.9	29.408	52.934
	NTVM	154.9	29.408	22.775
man	Chambolle	425.1	28.288	207.227
1024 × 1024	NChambolle	217.9	28.287	138.836
	PGCS	333.6	28.285	164.595
	NTVM	121.1	28.285	78.105
earth	Chambolle	425.8	26.876	882.497
2048 × 2048	NChambolle	220.1	26.875	608.292
	PGCS	341.0	26.873	719.828
	NTVM	108.9	26.873	306.289

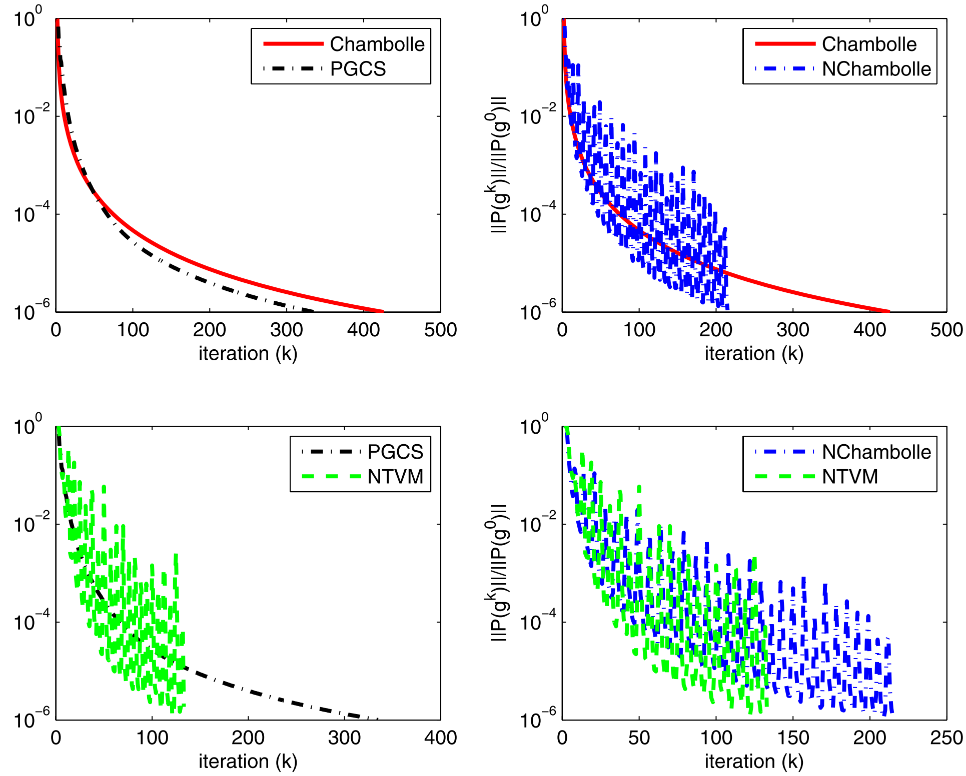


Fig. 2 Relative norm of projected gradient v.s. Iteration for denoising "man"

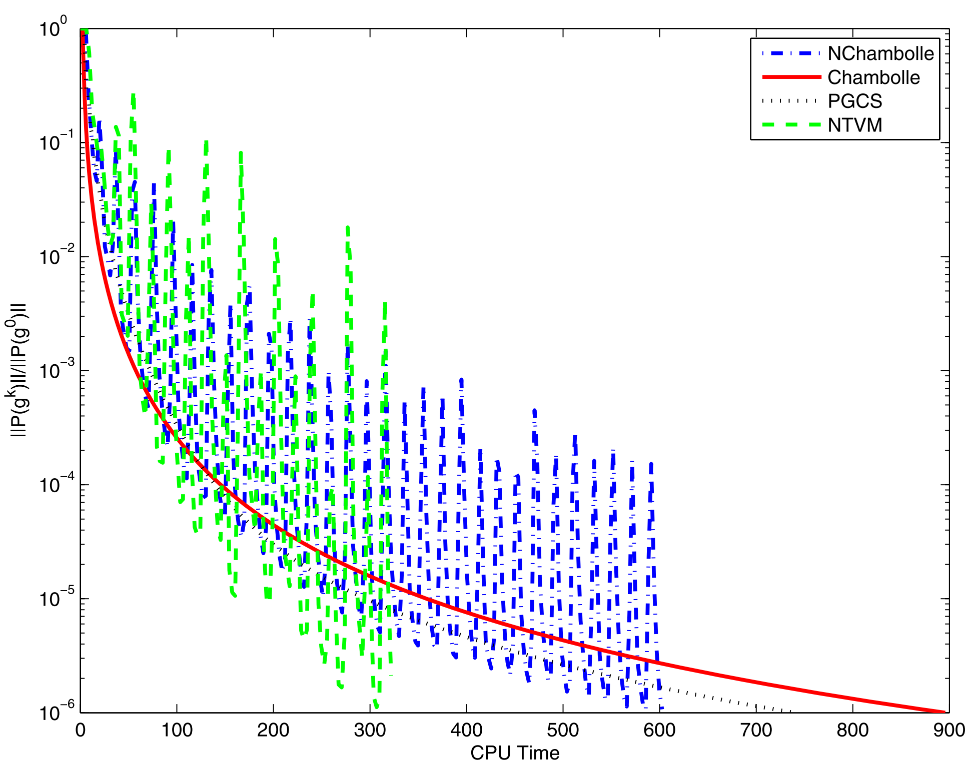


Fig. 3 Relative norm of projected gradient v.s. CPU time for denoising "earth"

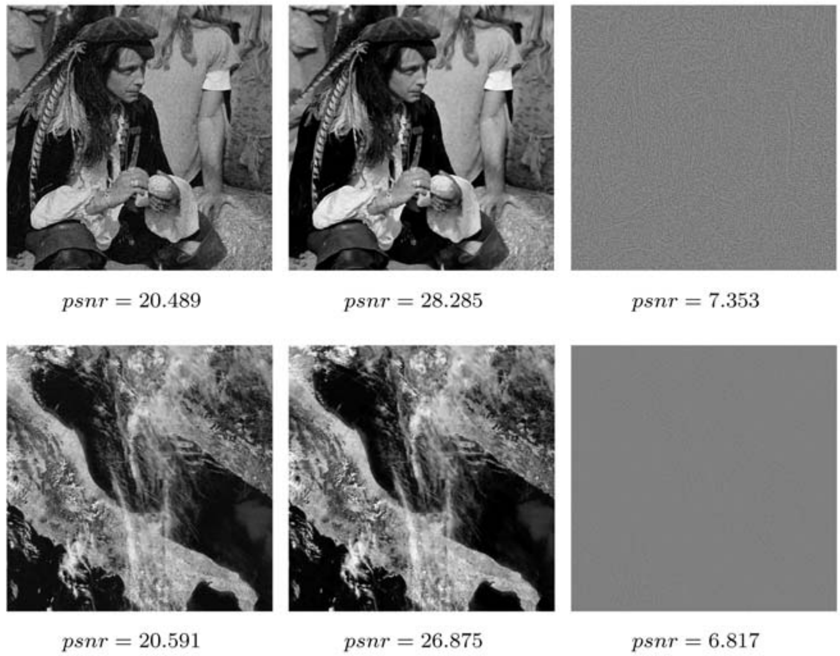


Fig. 4 The noisy image (f), the restored image (u), and the removed noise ($-\frac{1}{\lambda}\nabla \cdot w$) via NTVM. The *first row* is the results for "brain MRI" image, the *second row* is for "man" image and the *third row* is for "earth" image

4.2 Comparisons to Some Recent Efficient Algorithms

Recently, there are several methods regarded as particularly efficient for total variation image restoration. One is split Bregman method [19], which uses functional splitting and Bregman iteration for constrained optimization. They provided a freely available C++ code (SplitBregman) with Matlab mex interface at the web.[1] Similar to split Bregman iteration, an alternating minimization algorithm was proposed in [25], where the authors developed a fast algorithm for total variation based deconvolution (FTVd). Their Matlab code is available on-line.[2] The other related approach is Nesterov algorithm [26], which is an efficient scheme for convex optimization that allows to obtain a solution of precision ϵ in $O(\frac{1}{\epsilon})$ iterations. Very recently, a public-domain software for total variation image reconstruction via Nesterov algorithm was developed by Dahl, Hansen, Jensen et al. [15]. Their code is available at the web.[3] In this subsection, we would like to compare NTVM method with the above three efficient algorithms for solving total variational image denoising.

We test the following four algorithms:

- **NTVM** with the stopping condition $||P(g^k)|| \leq 10^{-3}||P(g^0)||$.
- **Nesterov:** Nesterov TVdenoise algorithm with $\tau = 0.8$.
- **SplitBregman** with $\mu = 0.053$ and the stopping tolerance parameter $tol = 10^{-3}$.
- **FTVdG:** modified FTVd algorithm in which the linear blurring matrix reduces to the identity matrix and $\mu = 25$.

The noisy images are generated by adding white Gaussian noise to the clean images with standard deviation $\sigma = 25$. The restoration results for "lena" and "Brain-MRI" are shown in Figs. 6 and 7. The detailed computational times (Sec) and PSNR values (dB) are listed in Table 2. Although these four algorithms use deferent stopping criterion, we can see that they attain some similar PSNR values. As can be seen from Table 2 and Fig. 5, SplitBregman algorithm has the best computational performance and NTVM algorithm is very competitive to FTVdG algorithm, both of which are faster than Nesterov algorithm.

5 Conclusion

In this paper we have proposed some nonmonotone total variation minimization (NTVM) algorithms. The main contribution of this paper is the development of two nonmonotone Chambolle algorithms that use well known Barzilai-Borwein (BB) formula for calculating time stepsize coupled with the adaptive nonmonotone line search proposed by Dai and Fletcher. The NTVM method is shown to be globally convergent. Numerical results indicate that the NTVM method is an improvement over the original method. For some large-scale test images, the NTVM algorithm runs about 3 times faster than the Chambolle projection

[1] http://www.math.ucla.edu/~tagoldst/code.html.

[2] http://www.caam.rice.edu/~optimization/L1/ftvd/.

[3] http://www.netlib.org/numeralgo in the file na28 or http://www2.imm.dtu.dk/~pch/mxTV/index.html.

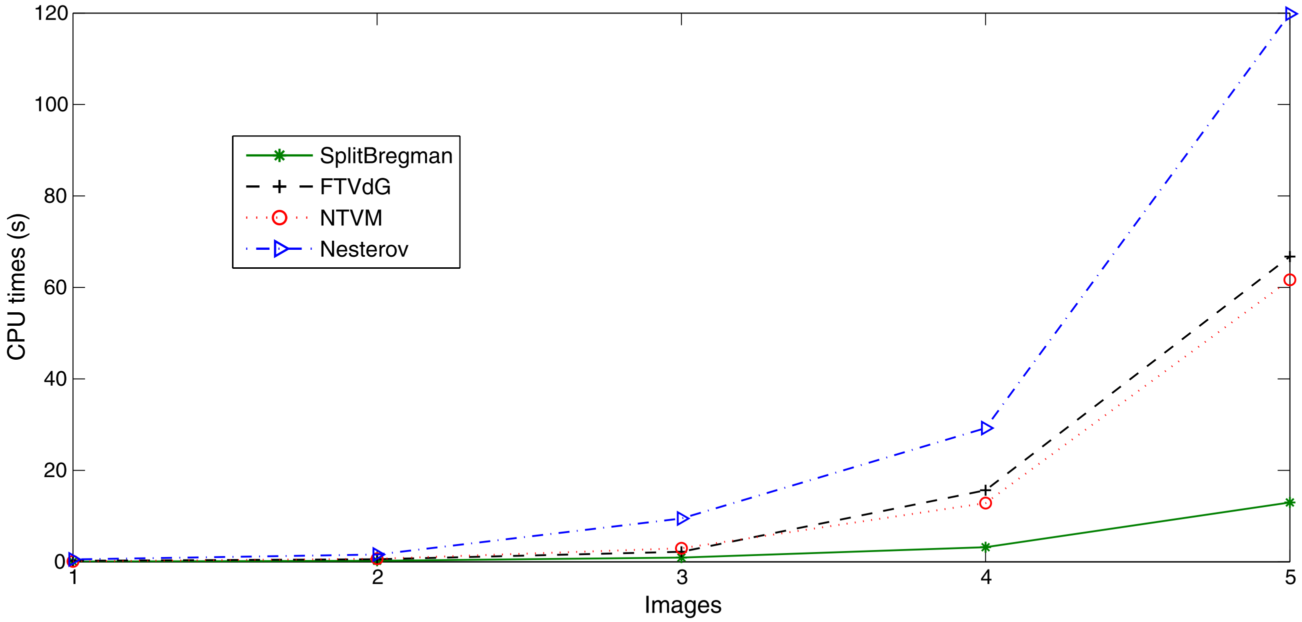
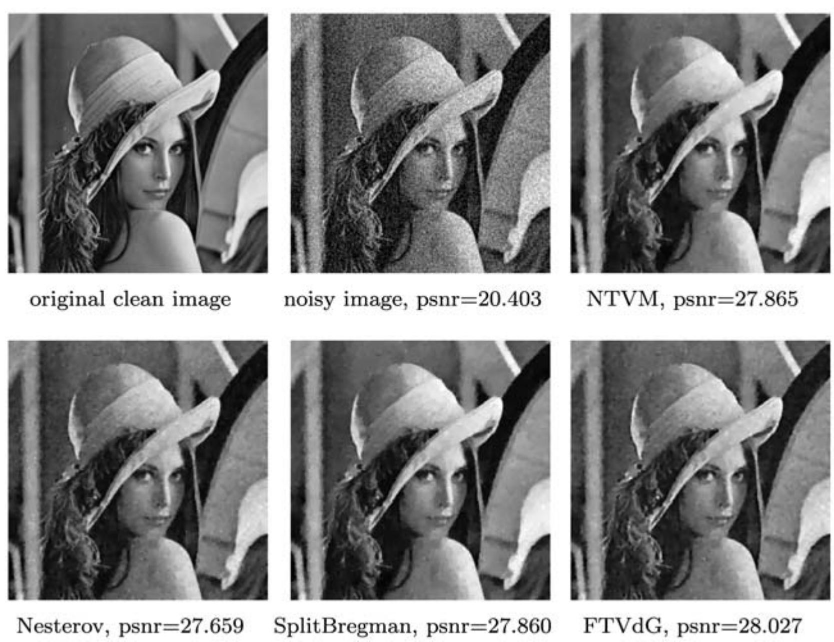


Fig. 5 Computational times of the test algorithms for denoising 5 test images in Table 2

Fig. 6 Comparisons of image denoising for "Lena" via NTVM/Nesterov/SplitBregman/FTVdG algorithm, respectively



method, i.e. it saves about 60% CPU time than the Chambolle method, which indicates, such kind of nonmonotone method is more suitable than the monotone descent method to solve some large-scale inverse problems. We also give some comparisons to illustrate that NTVM is very competitive to some efficient algorithms for total variation restoration. Further applications to image reconstruction, such as deconvolution, inpainting and multiscale decompositions could be our topics in the future work.

Acknowledgements The authors would like to thank two anonymous referees for their comments and suggestions on the first version of the article, which lead to significant improvements of the presentation.

Appendix: Proof of Convergence of NTVM Algorithm

The aim of this section is to give the sketch of the proof of Theorem 2. We denote $\mathcal{P}$ as the projection operator onto K and define the scaled projected gradient $g_t(w) = \mathcal{P}[w -$

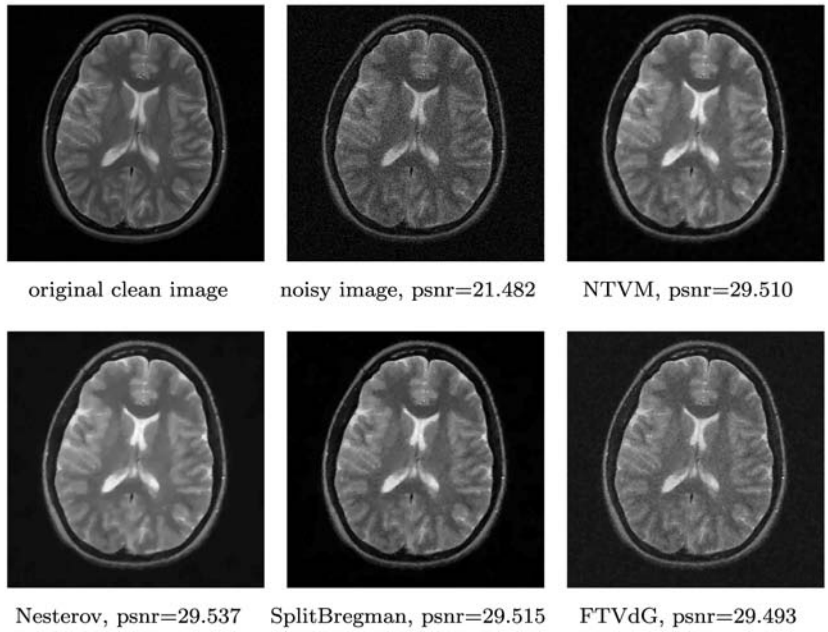


Fig. 7 Comparisons of image denoising for "BrainMRI" via NTVM/Nesterov/SplitBregman/FTVdG algorithm, respectively

Table 2 Computational times and PSNR values

Image	Algorithm	PSNR	CPU time
shape	SplitBregman	23.846	0.047
128 × 128	FTVdG	22.332	0.203
	Nesterov	22.562	0.516
	NTVM	23.883	0.078
lena	SplitBregman	27.860	0.203
256 × 256	FTVdG	28.027	0.547
	Nesterov	27.659	1.594
	NTVM	27.865	0.641
brain MRI	SplitBregman	29.515	0.906
512 × 512	FTVdG	29.493	2.203
	Nesterov	29.537	9.469
	NTVM	29.510	2.953
man	SplitBregman	28.376	3.188
1024 × 1024	FTVdG	28.676	15.656
	Nesterov	28.297	29.234
	NTVM	28.387	12.844
earth	SplitBregman	26.938	12.969
2048 × 2048	FTVdG	27.462	66.734
	Nesterov	27.134	119.859
	NTVM	26.948	61.672

$tg(w)] - w$ for all $w \in K$ and $t \in (0, \rho_{\max}]$. We will use the following lemma about the property of the scaled projected gradient, which can be found in [6].

Lemma 1 *For all $w \in K$ and $t \in (0, \rho_{\max}]$, it holds that*

$$\langle g(w), g_t(w)\rangle \le -\frac{1}{t}\|g_t(w)\|^2 \le -\frac{1}{\rho_{\max}}\|g_t(w)\|^2. \qquad (18)$$

If w^k is not a constrained stationary point, then by Lemma 1, $\langle g(w^k), d^k\rangle \le -\frac{1}{\rho_{\max}}\|g_{t^k}(w)\|^2 < 0$, i.e., the search direction is a descent direction. Hence, a stepsize satisfying (16) will be found after a finite number of trials. So, NTVM algorithm is well defined. Recall that $g(w)$ is Lipschitz continuous (the Lipschitz constant L is bounded by 8, see [9], p. 92), we have the following useful lemma.

Lemma 2 *Let β_k satisfies condition (16) in NTVM algorithm, then for all $k > 0$*

$$\beta_k \ge \min\left\{1, \frac{2(1-\theta)\lambda_1}{L}\frac{|\langle g(w^k), d^k\rangle|}{\|d^k\|^2}\right\}. \qquad (19)$$

Proof If $\beta = 1$ satisfies condition (16), then we have $\beta_k = 1$. Otherwise, there exists $\sigma \in [\lambda_1, \lambda_2]$ for which $\frac{\beta_k}{\sigma} > 0$ fails to satisfy condition (16), it follows that

$$\begin{aligned} F\left(w^k + \frac{\beta_k}{\sigma}d^k\right) &> F_r + \theta\frac{\beta_k}{\sigma}\langle g(w^k), d^k\rangle \\ &> F(w^k) + \theta\frac{\beta_k}{\sigma}\langle g(w^k), d^k\rangle. \end{aligned}$$

On the other hand, by the mean-value theorem and the Lipschitz condition, we have

$$F\left(w^k + \frac{\beta_k}{\sigma}d^k\right) - F(w^k)$$

$$
\begin{aligned}
&= \int_0^{\frac{\beta_k}{\sigma}} \langle (g(w^k + td^k) - g(w^k)), d^k \rangle dt + \frac{\beta_k}{\sigma} \langle g(w^k), d^k \rangle \\
&\le \frac{L}{2}\left(\frac{\beta_k}{\sigma}\right)^2 \|d^k\|^2 + \frac{\beta_k}{\sigma} \langle g(w^k), d^k \rangle.
\end{aligned}
$$

Therefore, combining the above two inequalities, we can obtain that (19) holds. □

We are now in position to prove Theorem 2.

Proof By contradiction. If the conclusion of Theorem 2 does not hold, we denote $\bar{w}$ an accumulation point of $\{w_k\}$, and relabel $\{w_k\}$ a subsequence converging to $\bar{w}$. For all $k > 0$, we have $\langle g(w^k), d^k \rangle \le -\varepsilon$ for some $\varepsilon > 0$. As shown in Sect. 3, in this case, there exists an infinite subsequence I such that for $k_i \in I$, the values of F_r on iterations k_i are strictly monotonically decreasing. Let $F_r^{k_i}$ denote the value F_r on iterations k_i. Then we can derive straightforward that

$$
\begin{aligned}
F(w^{k_i+1}) &\le F_r^{k_i} + \theta\beta_{k_i} \langle g(w^{k_i}), d^{k_i} \rangle \\
&\le F_r^{k_0} + \theta \sum_{j=k_0, j\in I}^{k_i} \beta_j \langle g(w^j), d^j \rangle.
\end{aligned}
$$

From Lemmas 1 and 2, we have $\langle g(w^k), d^k \rangle \le -\frac{1}{\rho_{\max}} \|d^k\|^2$ and $\beta_k \ge \min\{1, \frac{2(1-\theta)\lambda_1}{L\rho_{\max}}\}$. So, we can derive that

$$
\begin{aligned}
F_r^{k_0} - F(w^{k_i+1}) &\ge \theta \sum_{j=k_0, j\in I}^{k_i} \beta_j |\langle g(w^j), d^j \rangle| \\
&\ge \theta\varepsilon \sum_{j=k_0, j\in I}^{k_i} \min\left\{1, \frac{2(1-\theta)\lambda_1}{L\rho_{\max}}\right\}.
\end{aligned}
$$

Since $F(w)$ is bounded below, let $i \to \infty$, we get $\infty > F_r^{k_0} - F(w^{k_i+1}) \to \infty$. This is a contradiction. Hence the result of Theorem 2 holds. The proof is completed. □

References

1. Aubert, G., Aujol, J.F.: A variational approach to removing multiplicative noise. SIAM J. Appl. Math. **68**, 925–946 (2008)
2. Aubert, G., Kornprobst, P.: Mathematical Problems in Image Processing. Applied Mathematical Sciences, vol. 147. Springer, Berlin (2002)
3. Aujol, J.F.: Some algorithms for total variation based image restoration. Report CLMA 2008-05 (2008)
4. Barzilai, J., Borwein, J.M.: Two-point step size gradient methods. IMA J. Numer. Anal. **8**, 141–148 (1988)
5. Bertsekas, D.P.: Nonlinear Programming. Athena Scientific, Nashua (1999)
6. Birgin, E.G., Martinez, J.M., Raydan, M.: Nonmonotone spectral projected gradient methods on convex sets. SIAM J. Optim. **10**, 1196–1211 (2000)
7. Bresson, X., Chan, T.: Fast dual minimization of the vectorial total variation norm and applications to color image processing. Technical report, UCLA CAM Report 07-25 (2007)
8. Carter, J.L.: Dual methods for total variation-based image restoration. Ph.D. thesis, University of California at Los Angeles (2001) (Advisor: T.F. Chan)
9. Chambolle, A.: An algorithm for total variation minimization and applications. J. Math. Imaging Vis. **20**, 89–97 (2004)
10. Chambolle, A.: Total variation minimization and a class of binary MRF models. In: EMMCVPR 05. Lecture Notes in Computer Sciences, vol. 3757, pp. 136–152. Springer, Berlin (2005)
11. Chambolle, A., Lions, P.L.: Image recovery via total variation minimization and related problems. Numer. Math. **76**, 167–188 (1997)
12. Chan, T., Chen, K., Carter, J.L.: Iterative methods for solving the dual formulation arising from image restoration. Electron. Trans. Numer. Anal. **26**, 299–311 (2007)
13. Chan, T., Shen, J.: Image Processing and Analysis—Variational, PDE, Wavelet, and Stochastic Methods. Philadelphia, SIAM (2005)
14. Chan, T., Golub, G., Mulet, P.: A nonlinear primal-dual method for total variation-based image restoration. SIAM J. Sci. Comput. **20**, 1964–1977 (1999)
15. Dahl, J., Hansen, P., Jensen, S., Jensen, T.: Algorithms and software for total variation image reconstruction via first-order methods. Numer. Algorithms (2009, to appear)
16. Dai, Y.H., Fletcher, R.: Projected Barzilai-Borwein methods for large-scale box-constrained quadratic programming. Numer. Math. **100**, 21–47 (2005)
17. Dai, Y.H., Liao, L.Z.: R-linear convergence of the Barzilai and Borwein gradient method. IMA J. Numer. Anal. **22**, 1–10 (2002)
18. Ekeland, I., Témam, R.: Convex Analysis and Variational Problems. SIAM Classics in Applied Mathematics. SIAM, Philadelphia (1999)
19. Goldstein, T., Osher, S.: The split Bregman method for L_1 regularized problems. UCLA CAM Report 08-29 (2008)
20. Grippo, L., Sciandrone, M.: Nonmonotone globalization techniques for the Barzilai-Borwein gradient method. Comput. Optim. Appl. **23**, 143–169 (2002)
21. Huang, Y.M., Ng, M.K., Wen, Y.W.: A fast total variation method for multiplicative noise removal. SIAM J. Imaging Sci. **2**, 20–40 (2009)
22. Ng, M.K., Qi, L., Yang, Y.F., Huang, Y.M.: On semismooth Newton methods for total variation minimization. J. Math. Imaging Vis. **27**, 265–276 (2007)
23. Raydan, M.: On the Barzilai and Borwein choice of steplength for the gradient method. IMA J. Numer. Anal. **13**, 321–326 (1993)
24. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**, 259–268 (1992)
25. Wang, Y., Yang, J., Yin, W., Zhang, Y.: A new alternating minimization algorithm for total variation image reconstruction. SIAM J. Imaging Sci. **1**, 248–272 (2008)
26. Weiss, P., Blanc-Féraud, L., Aubert, G.: Efficient schemes for total variation minimization under constraints in image processing. SIAM J. Sci. Comput. **31**, 2047–2080 (2009)
27. Zhu, M.: Fast numerical algorithms for total variation based image restoration. Ph.D. thesis, University of California at Los Angeles, July 2008 (Advisor: T.F. Chan)
28. Zhu, M., Chan, T.F.: An efficient primal-dual hybrid gradient algorithm for total variation image restoration. Technical report, UCLA CAM Report 08-34 (2008)
29. Zhu, M., Wright, S.J., Chan, T.F.: Duality-based algorithms for total variation image restoration. Technical report, UCLA CAM Report 08-33 (2008)

Gaohang Yu received his Ph.D. degree in Computational Mathematics at Sun Yat-Sen University in 2007. From 2007 to 2009, he taught in Sun Yat-Sen University. Then he joined the School of Mathematics and Computer Science at the Gan-Nan Normal University as an Associate Professor. He is currently a postdoctoral fellow at the Department of Applied Mathematics, the Hong Kong Polytechnic University. His research interests include optimization theory and methods, and image processing.

Liqun Qi received his B.S. in Computational Mathematics at Tsinghua University in 1968, his M.S. and Ph.D. degree in Computer Sciences at University of Wisconsin-Madison in 1981 and 1984, respectively. Professor Qi has taught in Tsinghua University, China, University of Wisconsin-Madison, USA, University of New South Wales, Australia, and The City University of Hong Kong. He is now Chair Professor of Applied Mathematics and Head of Department of Applied Mathematics at The Hong Kong Polytechnic University. Professor Qi has published more than 170 research papers in international journals. He established the superlinear and quadratic convergence theory of the generalized Newton method, and played a principal role in the development of reformulation methods in optimization. Professor Qi's research work has been cited by the researchers around the world. According to the authoritative citation database www.isihighlycited.com, he is one of the world's most highly cited 300 mathematicians during the period from 1981 to 1999. Professor Liqun Qi was elected as a foreign member of the Peterovskaya Academy of Arts and Sciences, Russia in 2003. He received the Hong Kong Polytechnic University President's Awards for Excellence Performance/Achievements, based upon Research and Scholarly Activities in 2004. Professor Qi is the editor or an associate editor of eight international journals. He has chaired more than ten international conferences and workshops held at Australia, Italy, Hong Kong and the Mainland China. In 2005, Professor Qi introduced the concept of eigenvalues for higher order tensors, which now has applications in medical engineering, statistical data analysis and solid mechanics.

Yuhong Dai is Professor at Academy of Mathematics and Systems Science, Chinese Academy of Sciences. His major research interest is theory and numerical method for nonlinear programming. Recently, he also studies some special optimization problems arising from some practical fields.