Appl. Comput. Harmon. Anal. 27 (2009) 367–379

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Letter to the Editor

Convergence analysis of the Bregman method for the variational model of image denoising ☆

Rong-Qing Jia *, Hanqing Zhao, Wei Zhao

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Canada T6G 2G1

ARTICLE INFO

Article history:
Available online 13 May 2009
Communicated by Charles K. Chui
on 3 October 2008

Keywords:
Bregman method
Optimization
Total variation
Image denoising
Wavelet packets

ABSTRACT

The total variation model of Rudin, Osher, and Fatemi for image denoising is considered to be one of the best denoising models. Recently, by using the Bregman method, Goldstein and Osher obtained a very efficient algorithm for the solution of the ROF model. In this paper, we give a rigorous proof for the convergence of the Bregman method. We also indicate that a combination of the Bregman method with wavelet packet decomposition often enhances performance for certain texture rich images.

1. Introduction

The TV (total variation) model of Rudin, Osher, and Fatemi [14] for image denoising is considered to be one of the best denoising models. Recently, using the Bregman method, Goldstein and Osher [9] provided a very efficient algorithm for the solution of the ROF model. In this paper, we give a rigorous proof for the convergence of the Bregman method. Moreover, for certain texture rich images, we also discuss some improvements of the ROF model by using a combination of the Bregman method with wavelet packet decomposition.

An image is regarded as a function

$$u : \{1, \ldots, N\} \times \{1, \ldots, N\} \to \mathbb{R},$$

where $N \geqslant 2$. Suppose $u \in \mathbb{R}^{N^2} = \mathbb{R}^{\{1,\ldots,N\}\times\{1,\ldots,N\}}$. For $1 \leqslant p < \infty$, let

$$\|u\|_p := \Bigg(\sum_{1\leqslant i,j\leqslant N} |u(i,j)|^p \Bigg)^{1/p},$$

and let $\|u\|_\infty := \max_{1\leqslant i,j\leqslant N} |u(i,j)|$. We use ∇_x to denote the difference operator given by $\nabla_x u(1, j) = 0$ for $j = 1, \ldots, N$ and

$$\nabla_x u(i, j) = u(i, j) - u(i-1, j), \quad i = 2, \ldots, N, \ \ j = 1, \ldots, N.$$

☆ Supported in part by NSERC Canada under Grant OGP 121336.

* Corresponding author.
E-mail addresses: rjia@ualberta.ca (R.-Q. Jia), hzhao@math.ualberta.ca (H. Zhao), wazhao@math.ualberta.ca (W. Zhao).

1063-5203/$ – see front matter
doi:10.1016/j.acha.2009.05.002

Then ∇_x is a linear mapping from $\mathbb{R}^{N^2}$ to $\mathbb{R}^{N^2}$. Similarly, ∇_y is the difference operator from $\mathbb{R}^{N^2}$ to $\mathbb{R}^{N^2}$ given by $\nabla_y u(i,1)=0$ for $i=1,\dots,N$ and

$$\nabla_y u(i,j)=u(i,j)-u(i,j-1), \quad i=1,\dots,N, \ j=2,\dots,N.$$

The *total variation* of u is represented by

$$\|\nabla_x u\|_1+\|\nabla_y u\|_1 \quad \text{or} \quad \left\|\sqrt{(\nabla_x u)^2+(\nabla_y u)^2}\,\right\|_1.$$

Let $f\in\mathbb{R}^{N^2}$ be an observed image with noise. We wish to recover a target image u from f by denoising. The anisotropic TV model for denoising can be formulated as the following minimization problem:

$$\min_u\left[\|\nabla_x u\|_1+\|\nabla_y u\|_1+\frac{\mu}{2}\|u-f\|_2^2\right], \tag{1.1}$$

where μ is an appropriately chosen positive parameter. Correspondingly, the isotropic TV model for denoising is the following minimization problem:

$$\min_u\left\{\left\|\sqrt{(\nabla_x u)^2+(\nabla_y u)^2}\,\right\|_1+\frac{\mu}{2}\|u-f\|_2^2\right\}. \tag{1.2}$$

This motivates us to consider the general minimization problem of a convex function on the n-dimensional Euclidean space $\mathbb{R}^n$. Let $E:\mathbb{R}^n\to\mathbb{R}$ be a continuous and convex function. A vector g in $\mathbb{R}^n$ is called a *subgradient* of E at a point $v\in\mathbb{R}^n$ if

$$E(u)-E(v)-\langle g,u-v\rangle\geqslant 0 \quad \forall u\in\mathbb{R}^n.$$

The *subdifferential* $\partial E(v)$ is the set of subgradients of E at v. It is known that the subdifferential of a convex function at any point is nonempty. Clearly, v is a minimal point of E if and only if $0\in\partial E(v)$. If this is the case, we write

$$v=\arg\min_u\{E(u)\}.$$

The *Bregman distance* associated with E at v is

$$D_E^g(u,v):=E(u)-E(v)-\langle g,u-v\rangle,$$

where g is a subgradient of E at v. Clearly, $D_E^g(u,v)\geqslant 0$. Moreover, $D_E^g(u,v)\geqslant D_E^g(w,v)$ for any w on the line segment between u and v.

Let E and H be two convex functions from $\mathbb{R}^n$ to $\mathbb{R}$. Suppose that E is continuous and H is continuously differentiable. By abuse of notation, $\partial H(v)$ is also used to denote the *gradient* of H at a point $v\in\mathbb{R}^n$. We start with initial vectors g^0 and u^0 in $\mathbb{R}^n$. For $k=0,1,2,\dots,$ let

$$u^{k+1}:=\arg\min_u\left[E(u)-\left\langle g^k,u-u^k\right\rangle+H(u)\right] \tag{1.3}$$

and

$$g^{k+1}:=g^k-\partial H\left(u^{k+1}\right). \tag{1.4}$$

This is called the *Bregman iteration*, as was suggested in [1]. In the preceding process we have assumed the existence of solutions to the minimization problem in (1.3).

Let $A=(a_{ij})_{1\leqslant i\leqslant m,1\leqslant j\leqslant n}$ be an $m\times n$ matrix with its entries in $\mathbb{R}$. We assume that the rank of A is m. The matrix A induces the linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ given by $A[u_1,\dots,u_n]^T=[v_1,\dots,v_m]^T$ with

$$v_i=\sum_{j=1}^n a_{ij}u_j, \quad i=1,\dots,m.$$

Here and in what follows, we use the superscript T to denote the transpose of a matrix. Given a vector b in the range of A, we wish to find

$$\min_u E(u) \quad \text{subject to} \quad Au=b. \tag{1.5}$$

For the case when $E(u)=\|u\|_1$, Bregman iterative algorithms for the above minimization problem was analyzed in [20] by Yin et al. Among other things, they showed that the iterative approach yields exact solutions in a finite number of steps. Note that in this case solutions are not unique. For the minimization problems considered in this paper, in general, the solutions will not be achieved in a finite number of Bregman iterations.

Suppose that the minimization problem (1.5) has a unique solution $\tilde{u}$. For $\lambda>0$, set $H(u):=(\lambda/2)\|Au-b\|_2^2$. Let $(u^{k+1})_{k=0,1,\dots}$ be the sequence given by the Bregman iteration (1.3) and (1.4). In Section 4 we will give necessary and sufficient conditions for the sequence $(u^k)_{k=1,2,\dots}$ to converge to $\tilde{u}$. In order to establish this fundamental result, we will discuss the close relationship between minimization and shrinkage in Section 2, and review basic properties of the Bregman method in Section 3.

The general principle for convergence of the Bregman iteration will be applied to the minimization problems induced by the TV models for denoising. Suppose that $\tilde{u}$ is the unique solution of the minimization problem (1.1). Following [9], we introduce new vectors $v_x \in \mathbb{R}^{N^2}$ and $v_y \in \mathbb{R}^{N^2}$ and consider the minimization problem

$$\min_{v_x, v_y, u}\left[E(v_x, v_y, u) + \frac{\lambda}{2}\|v_x - \nabla_x u\|_2^2 + \frac{\lambda}{2}\|v_y - \nabla_y u\|_2^2\right], \tag{1.6}$$

where $\lambda > 0$ and $E(v_x, v_y, u) := \|v_x\|_1 + \|v_y\|_1 + (\mu/2)\|u - f\|_2^2$. For $k = 0, 1, 2, \ldots,$ let $(v_x^{k+1}, v_y^{k+1}, u^{k+1})$ be given by the Bregman iteration scheme. In Section 5 we will prove that $\lim_{k\to\infty} u^k = \tilde{u}$. For the minimization problem (1.2) of the isotropic TV model, convergence of the Bregman method will also be established.

Cai, Osher, and Shen in their two papers [2] and [3] gave an analysis for convergence of the *linearized* Bregman iteration. Their ideas will be employed in our study. However, their results do not apply directly to the minimization problems studied in this paper.

The ROF model denoises well piecewise constant images while preserving sharp edges. However, the model does not represent well texture or oscillatory details, as it has been analyzed by Meyer [12]. In Section 6 we will employ a combination of the Bregman method with wavelet packet decomposition to enhance performance for certain texture rich images.

2. Minimization and shrinkage

In this section we discuss the close relationship between minimization and shrinkage. This study is essential for our analysis of convergence of the Bregman iteration. Our discussion is inspired by the work of Donoho and Johnstone [7] on the wavelet shrinkage method and the work of Chui and Wang [5] on the wavelet-based variational method.

Let E be the function given by $E(u) = |u|$, $u \in \mathbb{R}$. It is easily seen that $\partial E(v) = \{1\}$ for $v > 0$, $\partial E(0) = [-1, 1]$, and $\partial E(v) = \{-1\}$ for $v < 0$. Thus, if E is the function given by

$$E(u) = |u| + \frac{\lambda}{2}(u - c)^2, \quad u \in \mathbb{R},$$

where $\lambda > 0$ and $c \in \mathbb{R}$, then $0 \in \partial E(v)$ if and only if $v = \mathit{shrink}(c, 1/\lambda)$, where

$$\mathit{shrink}(c, 1/\lambda) := \begin{cases} c - 1/\lambda & \text{for } c > 1/\lambda, \\ 0 & \text{for } -1/\lambda \leqslant c \leqslant 1/\lambda, \\ c + 1/\lambda & \text{for } c < -1/\lambda. \end{cases}$$

Therefore, $v = \mathit{shrink}(c, 1/\lambda)$ is the unique point such that $E(v)$ achieves its minimum.

For $\lambda > 0$ and $c \in \mathbb{R}$ we define

$$\mathit{cut}(c, 1/\lambda) := \begin{cases} 1/\lambda & \text{for } c > 1/\lambda, \\ c & \text{for } -1/\lambda \leqslant c \leqslant 1/\lambda, \\ -1/\lambda & \text{for } c < -1/\lambda. \end{cases}$$

For two vectors $v = (v_1, \ldots, v_n)^T$ and $c = (c_1, \ldots, c_n)^T$ in $\mathbb{R}^n$, we write $v = \mathit{shrink}(c, 1/\lambda)$ if $v_i = \mathit{shrink}(c_i, 1/\lambda)$ for $i = 1, \ldots, n$. Analogously, if $v_i = \mathit{cut}(c_i, 1/\lambda)$ for $i = 1, \ldots, n$, then we write $v = \mathit{cut}(c, 1/\lambda)$.

For $1 \leqslant p < \infty$, the ℓ_p norm of a vector $u = (u_1, \ldots, u_n)^T \in \mathbb{R}^n$ is defined by

$$\|u\|_p := \left(\sum_{j=1}^n |u_j|^p\right)^{1/p},$$

and the ℓ_∞ norm of u is given by $\|u\|_\infty := \max_{1\leqslant j\leqslant n} |u_j|$.

Suppose E is the function on $\mathbb{R}^n$ given by

$$E(u) = \|u\|_1 + \frac{\lambda}{2}\|u - c\|_2^2, \quad u \in \mathbb{R}^n,$$

where $\lambda > 0$, and $c = (c_1, \ldots, c_n)^T \in \mathbb{R}^n$. Given $v = (v_1, \ldots, v_n)^T \in \mathbb{R}^n$, we see that $0 \in \partial E(v)$ if and only if $v = \mathit{shrink}(c, 1/\lambda)$.

In order to study the isotropic TV model (1.2) for denoising, we need to investigate the minimization problem

$$\min_{v_x, v_y \in \mathbb{R}}\left[\sqrt{v_x^2 + v_y^2} + \frac{\lambda}{2}|v_x - b_x|^2 + \frac{\lambda}{2}|v_y - b_y|^2\right],$$

where $b_x, b_y \in \mathbb{R}$ and $\lambda > 0$. For $(v_x, v_y) \in \mathbb{R}^2$, let $G(v_x, v_y) := \sqrt{v_x^2 + v_y^2}$, $F(v_x, v_y) := (\lambda/2)|v_x - b_x|^2 + (\lambda/2)|v_y - b_y|^2$, and $E(v_x, v_y) := G(v_x, v_y) + F(v_x, v_y)$. It is easily seen that $\partial F(0, 0) = (-\lambda b_x, -\lambda b_y)$ and

$$\partial G(0, 0) = \{(g_x, g_y) \in \mathbb{R}^2\colon\ g_x^2 + g_y^2 \leqslant 1\}.$$

Consequently, $(0,0)$ belongs to $\partial E(0,0)$ if and only if $b_x^2 + b_y^2 \leqslant 1/\lambda^2$. Suppose that $E(v_x, v_y)$ achieves the minimum. If $b_x^2 + b_y^2 \leqslant 1/\lambda^2$, then $v_x = 0$ and $v_y = 0$. Otherwise, we have $v_x^2 + v_y^2 > 0$. Moreover,

$$\frac{v_x}{\sqrt{v_x^2 + v_y^2}} + \lambda(v_x - b_x) = 0 \quad \text{and} \quad \frac{v_y}{\sqrt{v_x^2 + v_y^2}} + \lambda(v_y - b_y) = 0.$$

It follows that $\lambda v_y(v_x - b_x) = \lambda v_x(v_y - b_y)$. Hence, $v_y b_x = v_x b_y$. There exists a real number t such that $v_x = t b_x$ and $v_y = t b_y$. Consequently,

$$E(v_x, v_y) = |t|\sqrt{b_x^2 + b_y^2} + \frac{\lambda}{2}(t-1)^2\big[b_x^2 + b_y^2\big].$$

Therefore, $E(v_x, v_y)$ achieves the minimum if and only if

$$t = shrink\left(1, \frac{1}{\lambda\sqrt{b_x^2 + b_y^2}}\right) = \max\left(1 - \frac{1}{\lambda\sqrt{b_x^2 + b_y^2}}, 0\right).$$

With $s := \max(\sqrt{b_x^2 + b_y^2} - 1/\lambda, 0)$ we conclude that

$$v_x = \frac{s b_x}{\sqrt{b_x^2 + b_y^2}} \quad \text{and} \quad v_y = \frac{s b_y}{\sqrt{b_x^2 + b_y^2}}. \tag{2.1}$$

The above formula is also valid when $b_x^2 + b_y^2 \leqslant 1/\lambda^2$, provided we interpret both v_x and v_y as 0 when $s = 0$. This shrinkage formula already appeared in the paper [17] of Wang, Yin and Zhang.

3. The Bregman method

In this section we review some basic properties of the Bregman method.

Let E and H be two convex functions from $\mathbb{R}^n$ to $\mathbb{R}$. Suppose that E is continuous and H is continuously differentiable. The following theorem was proved in [13] by Osher et al. For the reader's convenience, we include the proof here.

Theorem 1. *For an initial vector $g^0 \in \mathbb{R}^n$, let u^{k+1} and g^{k+1} $(k = 0, 1, \ldots)$ be given by the Bregman iteration* (1.3) *and* (1.4)*. Then $g^k \in \partial E(u^k)$ and $H(u^{k+1}) \leqslant H(u^k)$ for $k = 1, 2, \ldots$. Moreover, if $\min_{u \in \mathbb{R}^n} H(u) = 0$, then*

$$\lim_{k\to\infty} H(u^k) = 0.$$

Proof. It follows from (1.3) that $0 \in \partial E(u^{k+1}) - g^k + \partial H(u^{k+1})$. By (1.4) we have $g^{k+1} = g^k - \partial H(u^{k+1})$. Hence, $g^{k+1} \in \partial E(u^{k+1})$. Moreover,

$$E(u^{k+1}) - \langle g^k, u^{k+1} - u^k\rangle + H(u^{k+1}) \leqslant E(u) - \langle g^k, u - u^k\rangle + H(u)$$

for all $u \in \mathbb{R}^n$. Choosing $u = u^k$ in the above inequality, we obtain

$$E(u^{k+1}) - E(u^k) - \langle g^k, u^{k+1} - u^k\rangle + H(u^{k+1}) \leqslant H(u^k). \tag{3.1}$$

But $E(u^{k+1}) - E(u^k) - \langle g^k, u^{k+1} - u^k\rangle \geqslant 0$. Therefore, $H(u^{k+1}) \leqslant H(u^k)$ for $k = 1, 2, \ldots$.

For $g \in \partial E(v)$, let $D_E^g(u, v)$ be the Bregman distance associated with E at v. For $j = 1, 2, \ldots$, we have

$$\begin{aligned} D_E^{g^{j+1}}(u, u^{j+1}) - D_E^{g^j}(u, u^j) + D_E^{g^j}(u^{j+1}, u^j) &= \big[E(u) - E(u^{j+1}) - \langle g^{j+1}, u - u^{j+1}\rangle\big] - \big[E(u) - E(u^j) - \langle g^j, u - u^j\rangle\big] \\ &\quad + \big[E(u^{j+1}) - E(u^j) - \langle g^j, u^{j+1} - u^j\rangle\big] \\ &= -\langle g^{j+1} - g^j, u - u^{j+1}\rangle = \langle h^{j+1}, u - u^{j+1}\rangle, \end{aligned}$$

where $h^{j+1} := -(g^{j+1} - g^j)$. It follows from (1.4) that $h^{j+1} = \partial H(u^{j+1})$. Hence,

$$H(u) - H(u^{j+1}) - \langle h^{j+1}, u - u^{j+1}\rangle \geqslant 0 \quad \forall u \in \mathbb{R}^n.$$

Consequently,

$$D_E^{g^{j+1}}(u, u^{j+1}) - D_E^{g^j}(u, u^j) \leqslant \langle h^{j+1}, u - u^{j+1}\rangle \leqslant H(u) - H(u^{j+1}),$$

where we have used the fact $D_E^{g^j}(u^{j+1}, u^j) \geqslant 0$. Hence,

$$\sum_{j=1}^{k}\big[D_E^{g^{j+1}}(u, u^{j+1}) - D_E^{g^j}(u, u^j)\big] \leqslant \sum_{j=1}^{k}\big[H(u) - H(u^{j+1})\big].$$

Note that $H(u) - H(u^{j+1}) \leqslant H(u) - H(u^{k+1})$ for $j \leqslant k$. Therefore,

$$D_E^{g^{k+1}}(u, u^{k+1}) - D_E^{g^1}(u, u^1) \leqslant k\big[H(u) - H(u^{k+1})\big] \quad \forall u \in \mathbb{R}^n.$$

Since $\min_{u\in\mathbb{R}^n} H(u) = 0$, there exists some $w \in \mathbb{R}^n$ such that $H(w) = 0$. We deduce from the preceding inequality that

$$0 \leqslant H(u^{k+1}) \leqslant \frac{1}{k}\big[D_E^{g^1}(w, u^1) - D_E^{g^{k+1}}(w, u^{k+1})\big] \leqslant \frac{1}{k} D_E^{g^1}(w, u^1).$$

This shows $\lim_{k\to\infty} H(u^k) = 0$. □

Let us consider the special case when

$$H(u) = \frac{\lambda}{2}\|Au - b\|_2^2, \quad u \in \mathbb{R}^n,$$

where $\lambda > 0$, A is an $m \times n$ matrix, and $b \in \mathbb{R}^m$. In this case, the following iteration scheme is equivalent to the above Bregman iteration (see [20, Theorem 3.1]). Suppose $b^0 \in \mathbb{R}^m$ is given. For $k = 0, 1, 2, \ldots,$ let

$$u^{k+1} := \arg\min_u \left[E(u) + \frac{\lambda}{2}\left\|Au - b^k\right\|_2^2\right] \tag{3.2}$$

and

$$b^{k+1} := b^k + b - Au^{k+1}. \tag{3.3}$$

Set

$$g^k := \lambda A^T(b^k - b). \tag{3.4}$$

It follows from (3.2) that

$$0 \in \partial E(u^{k+1}) + \lambda A^T(Au^{k+1} - b^k).$$

Hence, for $k = 0, 1, \ldots,$

$$g^{k+1} = \lambda A^T(b^{k+1} - b) = -\lambda A^T(Au^{k+1} - b^k) \in \partial E(u^{k+1}).$$

Moreover, (3.3) and (3.4) yield

$$g^{k+1} - g^k + \lambda A^T(Au^{k+1} - b) = 0.$$

Hence, (1.3) and (1.4) are valid.

4. Convergence of the Bregman iteration

We are in a position to establish the following basic criterion for the convergence of the Bregman iteration.

Theorem 2. *Suppose that the minimization problem*

$$\min_u E(u) \quad \textit{subject to} \quad Au = b$$

has a unique solution $\tilde{u}$. For $k = 0, 1, \ldots,$ let u^{k+1} and b^{k+1} be given as in (3.2) and (3.3). Then $\lim_{k\to\infty} u^k = \tilde{u}$, provided the following three conditions are satisfied:

(a) $\lim_{k\to\infty}(u^{k+1} - u^k) = 0$,
(b) $(u^k)_{k=1,2,\ldots}$ *is a bounded sequence in* $\mathbb{R}^n$,
(c) $(b^k)_{k=1,2,\ldots}$ *is a bounded sequence in* $\mathbb{R}^m$.

Proof. For $k = 0, 1, \ldots,$ let g^k be given as in (3.4). First, we assume that $\lim_{k\to\infty} u^k = u^*$ and $\lim_{k\to\infty} g^k = g^*$ exist. By Theorem 1, $\lim_{k\to\infty}\|Au^k - b\|_2^2 = 0$. Hence, $Au^* = b$. It follows from (1.3) that

$$E(u^{k+1}) - \langle g^k, u^{k+1} - u^k\rangle + \frac{\lambda}{2}\left\|Au^{k+1} - b\right\|_2^2 \leqslant E(u) - \langle g^k, u - u^k\rangle + \frac{\lambda}{2}\|Au - b\|_2^2$$

for all $u \in \mathbb{R}^n$. Choosing $u = \tilde{u}$ in the above inequality and letting $k \to \infty$, we obtain

$$E(u^*) \leqslant E(\tilde{u}) - \langle g^*, \tilde{u} - u^*\rangle.$$

By (3.4), for $k = 1, 2, \ldots, g^k$ lies in the range of A^T. Hence, g^* also lies in the range of A^T. In other words, $g^* = A^T w$ for some $w \in \mathbb{R}^m$. Consequently,

$$\langle g^*, \tilde{u} - u^* \rangle = \langle A^T w, \tilde{u} - u^* \rangle = \langle w, A(\tilde{u} - u^*) \rangle = 0.$$

Consequently,

$$E(u^*) \leqslant E(\tilde{u}) \quad \text{and} \quad Au^* = b.$$

Therefore, $u^* = \tilde{u}$.

Next, suppose that conditions (a), (b), and (c) are satisfied. By conditions (b) and (c) we see that there exists a sequence $(k_j)_{j=1,2,\ldots}$ of increasing positive integers such that $\lim_{j\to\infty} u^{k_j} = u^*$ and $\lim_{j\to\infty} g^{k_j} = g^*$. In light of condition (a), we also have $\lim_{j\to\infty} u^{k_j+1} = u^*$. Hence, the preceding argument is still valid and thereby $u^* = \tilde{u}$. The same argument tells us that any convergent subsequence of $(u^k)_{k=1,2,\ldots}$ must converge to $\tilde{u}$. Therefore, the sequence $(u^k)_{k=1,2,\ldots}$ itself converges to $\tilde{u}$. $\square$

The following result is often useful for verification of condition (a) in Theorem 2.

Theorem 3. *Suppose $t > 0$, $1 \leqslant s \leqslant n$ and $c_1, \ldots, c_s \in \mathbb{R}$. Let*

$$G(u) := \frac{t}{2} \sum_{j=1}^{s} (u_j - c_j)^2, \quad u = (u_1, \ldots, u_n) \in \mathbb{R}^n,$$

and let $E := F + G$, where F is a continuous convex function on $\mathbb{R}^n$. If $(u^{k+1})_{k=0,1,\ldots}$ and $(g^{k+1})_{k=0,1,\ldots}$ are the sequences given by the Bregman iteration (1.3) *and* (1.4)*, then*

$$\lim_{k\to\infty} \left(u_j^{k+1} - u_j^k\right) = 0, \quad j = 1, \ldots, s.$$

Proof. We observe that G is continuously differentiable and

$$\partial G(u) = t(u_1 - c_1, \ldots, u_s - c_s, 0, \ldots, 0), \quad u = (u_1, \ldots, u_n) \in \mathbb{R}^n.$$

For $k = 1, 2, \ldots,$ let

$$y^k := \partial G(u^k) = t\left(u_1^k - c_1, \ldots, u_s^k - c_s, 0, \ldots, 0\right).$$

Clearly, $\partial E(u^k) = \partial F(u^k) + y^k$. It follows that $h^k := g^k - y^k \in \partial F(u^k)$. Hence,

$$D_F^{h^k}\left(u^{k+1}, u^k\right) = F\left(u^{k+1}\right) - F\left(u^k\right) - \left\langle h^k, u^{k+1} - u^k \right\rangle \geqslant 0.$$

Moreover,

$$D_G^{y^k}\left(u^{k+1}, u^k\right) = G\left(u^{k+1}\right) - G\left(u^k\right) - \left\langle y^k, u^{k+1} - u^k \right\rangle = \frac{t}{2} \sum_{j=1}^{s} \left(u_j^{k+1} - u_j^k\right)^2.$$

Consequently,

$$D_E^{g^k}\left(u^{k+1}, u^k\right) = D_F^{h^k}\left(u^{k+1}, u^k\right) + D_G^{y^k}\left(u^{k+1}, u^k\right) \geqslant \frac{t}{2} \sum_{j=1}^{s} \left(u_j^{k+1} - u_j^k\right)^2.$$

On the other hand, with $H(u) = (\lambda/2)\|Au - b\|_2^2$, (3.1) tells us that

$$D_E^{g^k}\left(u^{k+1}, u^k\right) \leqslant H\left(u^k\right) = \frac{\lambda}{2} \left\| Au^k - b \right\|_2^2.$$

By Theorem 1, $\lim_{k\to\infty} H(u^k) = 0$. Therefore, $\lim_{k\to\infty} (u_j^{k+1} - u_j^k) = 0$ for $j = 1, \ldots, s$. $\square$

5. The Bregman iteration for TV denoising

In this section we apply the results of the preceding section to the analysis of convergence of the Bregman iteration for TV denoising.

Let $f \in \mathbb{R}^{N^2}$ be an observed image. The anisotropic TV model for denoising is the minimization problem in (1.1). Following [9], we introduce new vectors $v_x, v_y \in \mathbb{R}^{N^2}$ and consider the function

$$E(v_x, v_y, u) := \|v_x\|_1 + \|v_y\|_1 + (\mu/2)\|u - f\|_2^2.$$

To apply the Bregman method, we turn to the minimization problem in (1.6). Choose $b_x^0 = b_y^0 = 0$. For $k = 0, 1, 2, \ldots$, let

$$\left(v_x^{k+1}, v_y^{k+1}, u^{k+1}\right) := \arg\min_{v_x, v_y, u}\left[E(v_x, v_y, u) + H^k(v_x, v_y, u)\right], \tag{5.1}$$

where $H^k(v_x, v_y, u) := (\lambda/2)\|v_x - \nabla_x u - b_x^k\|_2^2 + (\lambda/2)\|v_y - \nabla_y u - b_y^k\|_2^2$, and let

$$b_x^{k+1} := b_x^k - \left(v_x^{k+1} - \nabla_x u^{k+1}\right), \qquad b_y^{k+1} := b_y^k - \left(v_y^{k+1} - \nabla_y u^{k+1}\right). \tag{5.2}$$

Theorem 4. *Let $\tilde{u}$ be the unique solution of the minimization problem*:

$$\min_u\left[\|\nabla_x u\|_1 + \|\nabla_y u\|_1 + \frac{\mu}{2}\|u - f\|_2^2\right].$$

For $k = 0, 1, \ldots$, let $(v_x^{k+1}, v_y^{k+1}, u^{k+1})$ be given by the iteration scheme in (5.1) *and* (5.2)*. Then $\lim_{k\to\infty} u^k = \tilde{u}$.*

Proof. With $\tilde{v}_x := \nabla_x\tilde{u}$ and $\tilde{v}_y := \nabla_y\tilde{u}$, it is easily seen that $(\tilde{v}_x, \tilde{v}_y, \tilde{u})$ is the unique solution of the minimization problem in (1.6).

By Theorem 1 we have

$$\lim_{k\to\infty}\left\|v_x^k - \nabla_x u^k\right\|_2^2 = 0 \quad \text{and} \quad \lim_{k\to\infty}\left\|v_y^k - \nabla_y u^k\right\|_2^2 = 0. \tag{5.3}$$

Since $E(v_x, v_y, u) = \|v_x\|_1 + \|v_y\|_1 + (\mu/2)\|u - f\|_2^2$, by Theorem 3 we obtain

$$\lim_{k\to\infty}\left(u^{k+1} - u^k\right) = 0.$$

It follows that

$$\lim_{k\to\infty}\left(\nabla_x u^{k+1} - \nabla_x u^k\right) = 0 \quad \text{and} \quad \lim_{k\to\infty}\left(\nabla_y u^{k+1} - \nabla_y u^k\right) = 0.$$

This in connection with (5.3) gives

$$\lim_{k\to\infty}\left(v_x^{k+1} - v_x^k\right) = 0 \quad \text{and} \quad \lim_{k\to\infty}\left(v_y^{k+1} - v_y^k\right) = 0.$$

In order to prove $\lim_{k\to\infty} u^k = \tilde{u}$, by Theorem 2, it suffices to show that $(v_x^k, v_y^k, u^k)_{k=1,2\ldots}$, $(b_x^k)_{k=1,2\ldots}$ and $(b_y^k)_{k=1,2\ldots}$ are bounded sequences.

It follows from (5.1) that

$$\left\|v_x^{k+1}\right\|_1 + \frac{\lambda}{2}\left\|v_x^{k+1} - \nabla_x u^{k+1} - b_x^k\right\|_2^2 \leqslant \|v_x\|_1 + \frac{\lambda}{2}\left\|v_x - \nabla_x u^{k+1} - b_x^k\right\|_2^2$$

for all $v_x \in \mathbb{R}^{N^2}$. Hence, in light of the discussions in Section 2, we must have

$$v_x^{k+1} = shrink\left(\nabla_x u^{k+1} + b_x^k, 1/\lambda\right). \tag{5.4}$$

An analogous argument yields

$$v_y^{k+1} = shrink\left(\nabla_y u^{k+1} + b_y^k, 1/\lambda\right). \tag{5.5}$$

These two equations together with (5.2) give

$$b_x^{k+1} = cut\left(\nabla_x u^{k+1} + b_x^k, 1/\lambda\right) \quad \text{and} \quad b_y^{k+1} = cut\left(\nabla_y u^{k+1} + b_y^k, 1/\lambda\right).$$

In particular,

$$\left\|b_x^{k+1}\right\|_\infty \leqslant \frac{1}{\lambda} \quad \text{and} \quad \left\|b_y^{k+1}\right\|_\infty \leqslant \frac{1}{\lambda}.$$

Furthermore, (5.1) tells that the inequality

$$E\left(v_x^{k+1}, v_y^{k+1}, u^{k+1}\right) + H^k\left(v_x^{k+1}, v_y^{k+1}, u^{k+1}\right) \leqslant E(v_x, v_y, u) + H^k(v_x, v_y, u)$$

holds true for all (v_x, v_y, u). Choosing $v_x = v_y = u = 0$, we obtain

$$\left\|v_x^{k+1}\right\|_1 + \left\|v_y^{k+1}\right\|_1 + \frac{\mu}{2}\left\|u^{k+1} - f\right\|_2^2 \leqslant \frac{\mu}{2}\|f\|_2^2 + \frac{\lambda}{2}\left\|b_x^k\right\|_2^2 + \frac{\lambda}{2}\left\|b_y^k\right\|_2^2.$$

Therefore, $(v_x^k, v_y^k, u^k)_{k=1,2,\ldots}$ is a bounded sequence.

By Theorem 2, we conclude that the sequence $(v_x^k, v_y^k, u^k)_{k=1,2,\ldots}$ converges and thereby $\lim_{k\to\infty} u^k = \tilde{u}$. □

Now let us consider the isotropic TV model (1.2) for denoising. Following [9], we introduce new vectors $v_x, v_y \in \mathbb{R}^{N^2}$ and consider the function

$$E(v_x, v_y, u) := \left\| \sqrt{(v_x)^2 + (v_y)^2} \right\|_1 + \frac{\mu}{2} \|u - f\|_2^2.$$

We have the following result.

Theorem 5. *Let $\tilde{u}$ be the unique solution of the minimization problem given in* (1.2)*. For $k = 0, 1, \ldots,$ let $(v_x^{k+1}, v_y^{k+1}, u^{k+1})$ be given by the iteration scheme in* (5.1) *and* (5.2) *with $E(v_x, v_y, u)$ given as above. Then $\lim_{k\to\infty} u^k = \tilde{u}$.*

Proof. It suffices to show that $(b_x^k)_{k=1,2,\ldots}$ and $(b_y^k)_{k=1,2,\ldots}$ are bounded sequences. The other parts of the proof are analogous to that of Theorem 4.

Let $w_x^{k+1} := \nabla_x u^{k+1} + b_x^k$ and $w_y^{k+1} := \nabla_y u^{k+1} + b_y^k$, $k = 0, 1, \ldots.$ It follows from (5.1) that

$$\begin{aligned} &\left\| \sqrt{\left(v_x^{k+1}\right)^2 + \left(v_y^{k+1}\right)^2} \right\|_1 + \frac{\lambda}{2} \left\| v_x^{k+1} - w_x^{k+1} \right\|_2^2 + \frac{\lambda}{2} \left\| v_y^{k+1} - w_y^{k+1} \right\|_2^2 \\ &\quad \leqslant \left\| \sqrt{(v_x)^2 + (v_y)^2} \right\|_1 + \frac{\lambda}{2} \left\| v_x - w_x^{k+1} \right\|_2^2 + \frac{\lambda}{2} \left\| v_y - w_y^{k+1} \right\|_2^2 \quad \forall v_x, v_y \in \mathbb{R}^{N^2}. \end{aligned}$$

Let

$$s^{k+1} := \max\left(\sqrt{\left(w_x^{k+1}\right)^2 + \left(w_y^{k+1}\right)^2} - 1/\lambda, 0\right).$$

By (2.1) we have

$$v_x^{k+1} = \frac{s^{k+1} w_x^{k+1}}{\sqrt{(w_x^{k+1})^2 + (w_y^{k+1})^2}} \quad \text{and} \quad v_y^{k+1} = \frac{s^{k+1} w_y^{k+1}}{\sqrt{(w_x^{k+1})^2 + (w_y^{k+1})^2}},$$

where v_x^{k+1} and v_y^{k+1} are interpreted as 0 when $s^{k+1} = 0$. This in connection with (5.2) gives

$$b_x^{k+1} = w_x^{k+1} - v_x^{k+1} = \frac{w_x^{k+1}}{\sqrt{(w_x^{k+1})^2 + (w_y^{k+1})^2}} \left(\sqrt{\left(w_x^{k+1}\right)^2 + \left(w_y^{k+1}\right)^2} - s^{k+1}\right).$$

In particular, $b_x^{k+1} = w_x^{k+1}$ when $s^{k+1} = 0$. Hence, in all cases, $\|b_x^{k+1}\|_\infty \leqslant 1/\lambda$. Similarly, $\|b_y^{k+1}\|_\infty \leqslant 1/\lambda$. The proof is complete. □

6. Modified algorithms

In this section we propose modified algorithms based on a combination of the Bregman method with wavelet packet decomposition.

Let us recall the iteration scheme of Goldstein and Osher [9]. Set $b_x^0 = b_y^0 := 0$ and $v_x^0 = v_y^0 := 0$. For $k = 0, 1, \ldots,$ let u^{k+1} be the solution of the equation

$$(\mu - \lambda\Delta) u^{k+1} = \mu f + \lambda \nabla_x^T \left(v_x^k - b_x^k\right) + \lambda \nabla_y^T \left(v_y^k - b_y^k\right),$$

where $\Delta := -\nabla_x^T \nabla_x - \nabla_y^T \nabla_y$. Then update v_x^{k+1}, v_y^{k+1}, b_x^{k+1}, and b_y^{k+1} according to (5.4), (5.5), and (5.2).

The ROF model denoises well piecewise constant images while preserving sharp edges. However, the model does not represent well texture or oscillatory details. Indeed, for images with oscillatory details like Barbara, the ROF model (1.1) with small μ removes noise effectively, but many texture details are lost. On the other hand, the ROF model (1.1) with big μ retains texture details well, but many noisy spots remain. In order to preserve some oscillatory parts of images we propose the following algorithm based on a combination of the Bregman method with wavelet packet decomposition.

1. For the observed noisy image f, use the Bregman iteration to solve the ROF model (1.1) with μ replaced by 0.8μ and get an approximate image g.
2. Perform wavelet packet decomposition on $v := f - g$ three times and decompose the residual v into 64 subimages. Use the Bregman method to solve the ROF model (1.1) for each subimage.
3. Apply the inverse wavelet transform to the denoised subimages and get w. Then $g + w$ gives the target image.

In the first step, we oversmooth the noisy image f so that the approximate image g contains little noise. However, the residual $v = f - g$ still contains some parts of the original image. In the second step, we try to separate the oscillatory parts of the original image from noise by using wavelet packet decomposition. We will give more details for the second step and explain the reason why wavelet packets are useful. For a comprehensive study on wavelet packets, see the book [18] of Wickerhauser.

For $n = 1, 2, \ldots,$ let $J_n := \{1, \ldots, 2^n\}$ and $V_n := \ell_2(J_n)$. Let S_n be a discrete wavelet transform on V_n. The transform S_n can be represented as a matrix

$$S_n = \begin{bmatrix} L_n \\ H_n \end{bmatrix}, \tag{6.1}$$

where L_n is a $2^{n-1} \times 2^n$ matrix induced by a lowpass filter, and H_n is a $2^{n-1} \times 2^n$ matrix induced by a highpass filter. Each v_n is decomposed into $v_{n-1} = L_n v_n$ and $w_{n-1} = H_n v_n$. From v_{n-1} and w_{n-1} one may use the inverse wavelet transform to recover v_n:

$$v_n = S_n^{-1} \begin{bmatrix} v_{n-1} \\ w_{n-1} \end{bmatrix}.$$

For example, the matrices L_n and H_n corresponding to the Haar wavelet transform are given by

$$L_n := \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & & & & & & & & \\ & & 1 & 1 & & & & & & \\ & & & & \cdots & & & & & \\ & & & & & \cdots & & & & \\ & & & & & & 1 & 1 & & \\ & & & & & & & & 1 & 1 \end{bmatrix}$$

and

$$H_n := \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 & & & & & & & & \\ & & 1 & -1 & & & & & & \\ & & & & \cdots & & & & & \\ & & & & & \cdots & & & & \\ & & & & & & 1 & -1 & & \\ & & & & & & & & 1 & -1 \end{bmatrix}.$$

Suppose v is the 2^n vector $[1, 0, 0, 0, 1, 0, 0, 0, \ldots, \ldots, 1, 0, 0, 0]^T$. It represents an oscillatory signal. We have

$$L_n v = \frac{1}{\sqrt{2}}[1, 0, 1, 0, \ldots, \ldots, 1, 0]^T \quad \text{and} \quad H_n v = \frac{1}{\sqrt{2}}[1, 0, 1, 0, \ldots, \ldots, 1, 0]^T.$$

We see that $L_n v$ and $H_n v$ are still oscillatory. After performing the wavelet transform once more, we obtain

$$L_{n-1} L_n v = H_{n-1} L_n v = L_{n-1} H_n v = H_{n-1} H_n v = \frac{1}{2}[1, 1, \ldots, 1]^T.$$

Now all the above four sequences are constant. It is easily seen that this statement is true for every sequence v of period 4. We point out the difference between wavelet decomposition and wavelet packet decomposition. For wavelet decomposition, $H_n v$ will not be decomposed further; hence, the oscillatory part remains. For wavelet packet decomposition, $H_n v$ will be further decomposed into $L_{n-1} H_n v$ and $H_{n-1} H_n v$. If v is the original signal and some noise is added to v, then the ROF model may not perform well for denoising, since v is oscillatory. However, after wavelet packet decomposition, all the parts $L_{n-1} L_n v$, $H_{n-1} L_n v$, $L_{n-1} H_n v$, and $H_{n-1} H_n v$ are constant. The ROF model usually performs very well for constant signals with noise.

The Haar wavelet transform often yields some undesirable artifacts. We will use the orthogonal Daubechies wavelets on intervals as described in [6]. Biorthogonal wavelets can also be used. For discrete wavelets on intervals, see [16, Chap. 8] and [11, §4].

Let S_n be the wavelet transform given in (6.1). Define $M_n^0 := L_n$ and $M_n^1 := H_n$. Suppose $1 \leqslant k \leqslant n$. For $\varepsilon_1, \ldots, \varepsilon_k \in \{0, 1\}$, let

$$M_n^{\varepsilon_k \cdots \varepsilon_1} := M_{n-k+1}^{\varepsilon_k} \cdots M_n^{\varepsilon_1}.$$

Thus, after k times of wavelet packet decomposition, $v \in \ell_2(J_n)$ will be decomposed into 2^k sequences $M_n^{\varepsilon_k \cdots \varepsilon_1} v$, $\varepsilon_1, \ldots, \varepsilon_k \in \{0, 1\}$.

Images are considered as sequences on $J_n \times J_n$. For two-dimensional wavelet transforms, it is convenient to use the Kronecker product $A \otimes B$ of two matrices A and B. See, *e.g.*, [10, Chap. 4] for the definition and properties of the Kronecker product of matrices. By applying the wavelet transform once, a sequence v on $J_n \times J_n$ is decomposed into four sequences $(L_n \otimes L_n)v$, $(L_n \otimes H_n)v$, $(H_n \otimes L_n)v$, and $(H_n \otimes H_n)v$ on $J_{n-1} \times J_{n-1}$. More generally, by applying the wavelet transform k times ($1 \leqslant k \leqslant n$), v is decomposed into 2^{2k} sequences on $J_{n-k} \times J_{n-k}$ as follows:

$$\left(M_n^{\varepsilon_k \cdots \varepsilon_1} \otimes M_n^{\eta_k \cdots \eta_1}\right)v, \quad \varepsilon_1, \ldots, \varepsilon_k, \eta_1, \ldots, \eta_k \in \{0, 1\}.$$

Now let us consider images of size 512×512. In the second step of our algorithm, by applying the wavelet transform three times ($k = 3$), the residual $v = f - g$ on $J_9 \times J_9$ is decomposed into 64 subimages $(M^{\varepsilon_3 \varepsilon_2 \varepsilon_1} \otimes M^{\eta_3 \eta_2 \eta_1})v$, $\varepsilon_1, \varepsilon_2, \varepsilon_3, \eta_1, \eta_2, \eta_3 \in \{0, 1\}$. Then the Bregman method is used to solve the ROF model (1.1) for each subimage and get

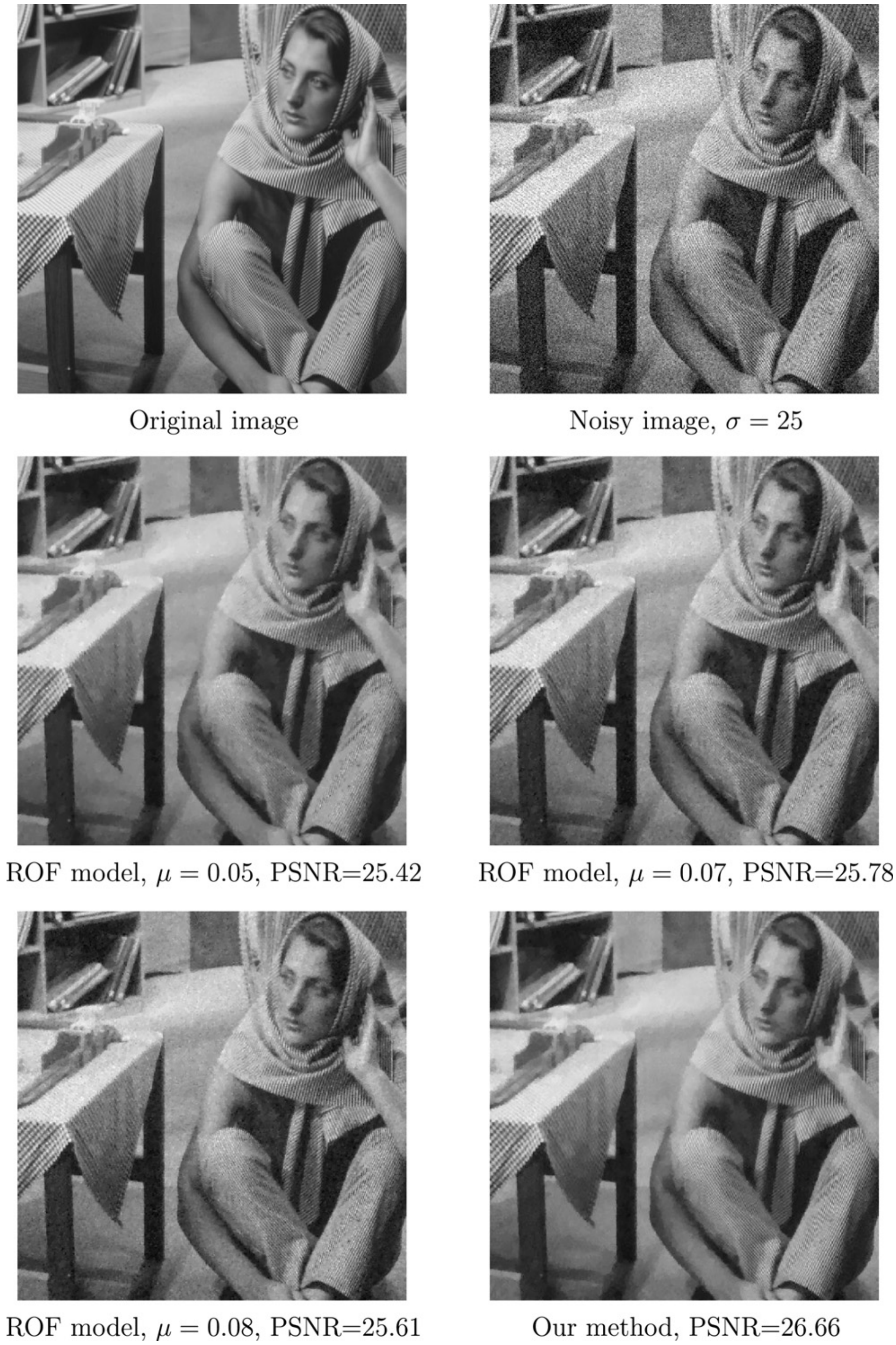
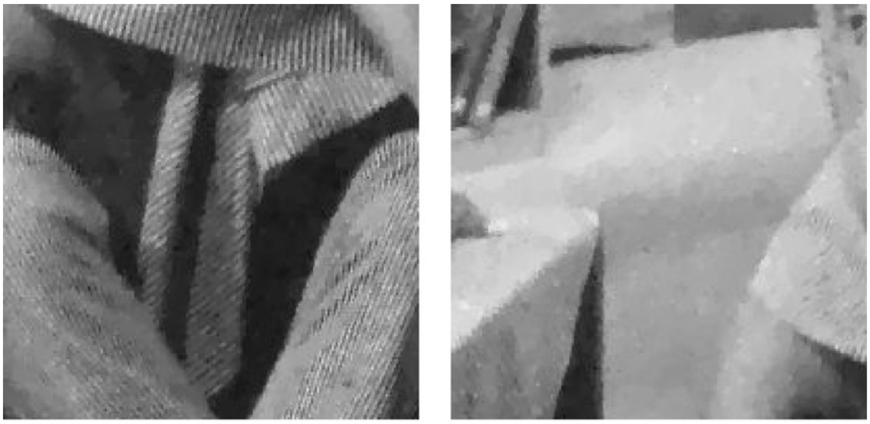


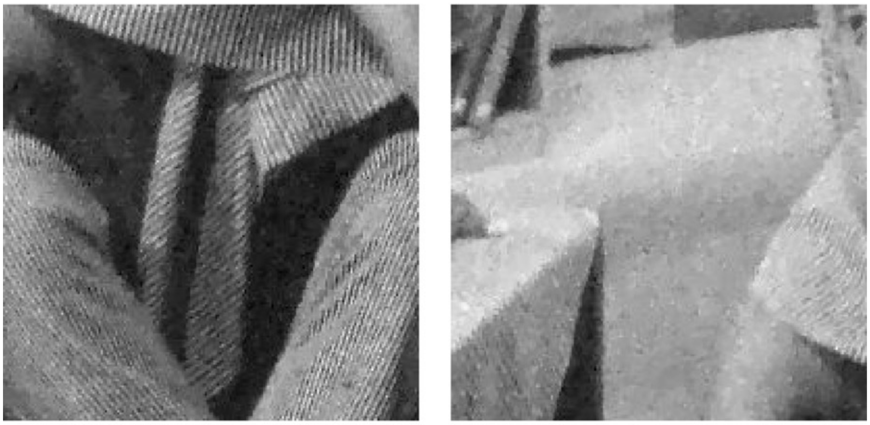
Fig. 1.

$w^{\varepsilon_3\varepsilon_2\varepsilon_1\eta_3\eta_2\eta_1}$. In the third step, the inverse wavelet transform is applied to get w from $w^{\varepsilon_3\varepsilon_2\varepsilon_1\eta_3\eta_2\eta_1}$, $\varepsilon_1, \varepsilon_2, \varepsilon_3, \eta_1, \eta_2, \eta_3 \in \{0, 1\}$. Finally, $g + w$ gives the denoised image.

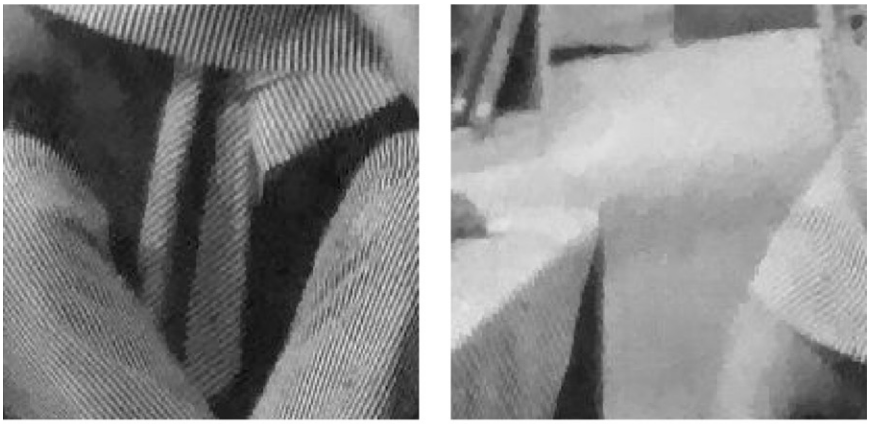
We test our algorithm on three representative images: Barbara, Fingerprint, and Dollar. Suppose that the gray-scale of an original image is in the range between 0 and 255. A Gaussian noise with normal distribution $N(0, \sigma^2)$ is added to the original image. We choose $\sigma = 25$. The peak-signal to noise ratio (PSNR) of each noised image is about 20.14. We compare our results with the pure shrinkage method via wavelet decomposition and wavelet packet decomposition, and the ROF



ROF model, $\mu = 0.05$



ROF model, $\mu = 0.07$

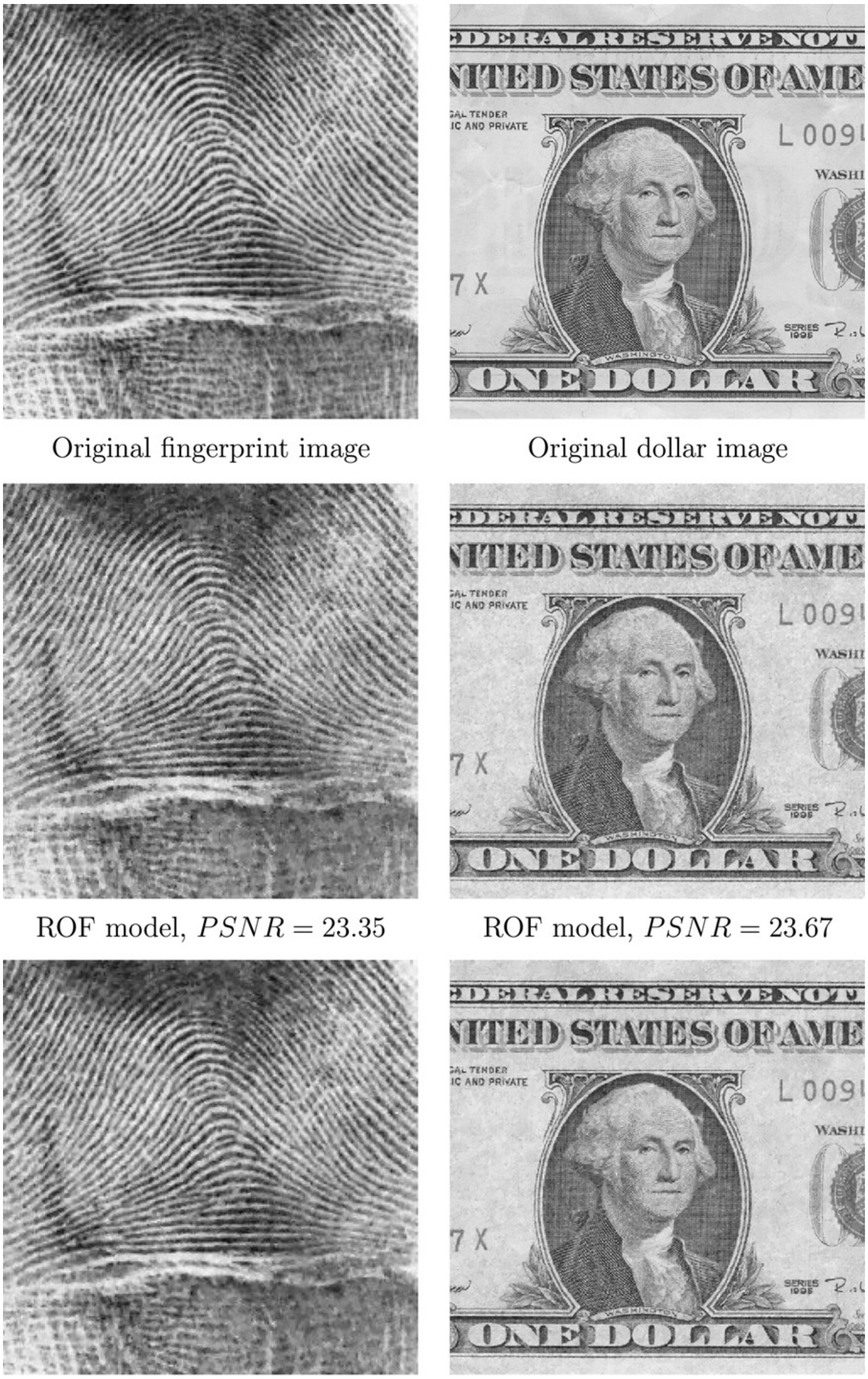


Our method

Fig. 2.

model by using the algorithm of Goldstein and Osher [9]. The following table lists the PSNR values for each image and each method. For all three images, our algorithm improves performance in terms of PSNR.

The improvement made by our algorithm is most visible for Barbara. Our algorithm has the advantage of both removing noises effectively and retaining texture details. This is demonstrated in Figs. 1 and 2. For Fingerprint and Dollar, the results are shown in Fig. 3.



Original fingerprint image

Original dollar image

ROF model, $PSNR = 23.35$

ROF model, $PSNR = 23.67$

Our method, $PSNR = 23.84$

Our method, $PSNR = 24.14$

Fig. 3.

The Bregman method can be used in many ways for image denoising. An iterative regularization method for image restoration was proposed in [13]. In its discrete form, the iteration scheme may be described as follows. Set $u^0 := 0$ and $v^0 := 0$. For $k = 0, 1, \ldots,$ compute u^{k+1} as a minimizer of the modified ROF model, *i.e.*,

$$u^{k+1} := \arg\min_u \left[\|\nabla_x u\|_1 + \|\nabla_y u\|_1 + \frac{\lambda}{2} \left\| f + v^k - u \right\|_2^2 \right],$$

Image	Wavelet shrinkage	Packet shrinkage	ROF model	Our method
Barbara	25.37	25.00	25.78	26.66
Fingerprint	22.86	22.89	23.35	23.84
Dollar	23.31	22.36	23.67	24.14

where $0 < \lambda < \mu$, and update $v^{k+1} := v^k + f - u^{k+1}$. Then the stopping criterion $\|u^k - f\|_2 \leqslant \delta$ is used, where δ represents the noise level. This scheme may be considered as a multi-step ROF model. The idea is to catch more signal than noise in each step until it is no longer possible.

Our algorithm can be viewed as a two-step ROF model. The approximate image g obtained in the first step contains most of the signal and removes most of the noise. However, the residual $v = f - g$ still contains some parts of the signal. In the second step, by using a combination of the Bregman method and wavelet packet decomposition, we can effectively catch some parts of the signal. In other words, in these two steps, we try to separate the signal from the noise. This is related to the work in [15] for separating images into texture and piecewise smooth parts.

Many researchers have studied image denoising in the wavelet domain. Chan and Zhou [4] did research on combining wavelets with variational and PDE methods. Chui and Wang [5] investigated wavelet-based minimal-energy approach to image restoration. When the total energy functional is formulated in the wavelet domain, it was proved in [5, Theorems 4 and 5] that the minimization problem reduces to soft or hard shrinkage. In [19] Xu and Osher applied the iterative regularization method in [13] to wavelet shrinkage. Clearly, the methods used in both [5] and [19] were wavelet shrinkages. In comparison, we applied the Bregman method to each subimage obtained from wavelet packet decomposition. Thus, our algorithm is different from the aforementioned schemes.

Denoising for texture rich images has been studied extensively in the literature. For example, Gilboa and Osher in [8] investigated semi-local and nonlocal variational minimizations for denoising. It would be interesting to incorporate the technique discussed in this paper to the more advanced study as in [8]. This will be topics of future research.

Acknowledgments

The authors are grateful to Dr. Zuowei Shen for his invitation to the research program on Mathematical Imaging and Digital Media held in Singapore between May 5 and Jun 27, 2008. They are pleased to acknowledge the support received from the Institute for Mathematical Sciences at the National University of Singapore during the research program including the summer school and the workshops. The authors also sincerely thank Cai, Osher, and Shen for making the two papers [2] and [3] available to them.

This research work was partially supported by NSERC Canada under Grant OGP 121336.

References

[1] L. Bregman, The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex optimization, USSR Comput. Math. Math. Phys. 7 (1967) 200–217.
[2] J.F. Cai, S. Osher, Z.W. Shen, Linearized Bregman iterations for compressed sensing, Math. Comput., in press.
[3] J.F. Cai, S. Osher, Z.W. Shen, Convergence of the linearized Bregman iteration for ℓ_1-norm minimization, Math. Comput., in press.
[4] T.F. Chan, H.M. Zhou, Total variation minimizing wavelet coefficients for image compression and denoising, in: Proceedings of International Conference on Image Processing, vol. 2, IEEE, 2001, pp. 391–394.
[5] C.K. Chui, J.Z. Wang, Wavelet-based minimal-energy approach to image restoration, Appl. Comput. Harmon. Anal. 23 (2007) 114–130.
[6] A. Cohen, I. Daubechies, P. Vial, Wavelets and fast wavelet transforms on an interval, Appl. Comput. Harmon. Anal. 1 (1993) 54–81.
[7] D.L. Donoho, I.M. Johnstone, Minimax estimation via wavelet shrinkage, Ann. Statist. 26 (1998) 879–921.
[8] G. Gilboa, S. Osher, Nonlocal linear image regularization and supervised segmentation, Multiscale Model. Simul. 6 (2007) 595–630.
[9] T. Goldstein, S. Osher, The split Bregman method for L1 regularized problems, UCLA CAM Report (08-29).
[10] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambridge University Press, Cambridge, 1991.
[11] R.Q. Jia, The projection method in the development of wavelet bases, in: Lizhen Ji, Kefeng Liu, Lo Yang, Shing-Tung Yau (Eds.), Proceedings of the Fourth International Congress of Chinese Mathematicians, vol. 1, Higher Education Press, Beijing, 2007, pp. 576–597.
[12] Y. Meyer, Oscillating Patterns in Image Processing and Nonlinear Evolution Equations, Univ. Lecture Ser., vol. 22, Amer. Math. Soc., Providence, 2001.
[13] S. Osher, M. Burger, D. Goldfarb, J.J. Xu, W.T. Yin, An iterative regularization method for total variation-based image restoration, Multiscale Model. Simul. 4 (2005) 460–489.
[14] L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. D 60 (1992) 259–268.
[15] J.-L. Starck, M. Elad, D.L. Donoho, Image Decomposition via the combination of sparse representations and a variational approach, IEEE Trans. Image Process. 14 (2005) 1570–1582.
[16] G. Strang, T. Nguyen, Wavelets and Filter Banks, Wellesley–Cambridge Press, Wellesley, MA, 1997.
[17] Y. Wang, W. Yin, Y. Zhang, A fast algorithm for image deblurring with total variation regularization, CAAM Technical Report TR07-10, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005, June 2007.
[18] M.V. Wickerhauser, Adapted Wavelet Analysis from Theory to Software, A.K. Peters, Wellesley, MA, 1994.
[19] J.J. Xu, S. Osher, Iterative regularization and nonlinear inverse scale space applied to wavelet-based denoising, IEEE Trans. Image Process. 16 (2007) 534–544.
[20] W.T. Yin, S. Osher, D. Goldfarb, J. Darbon, Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing, SIAM J. Imaging Sci. 1 (2008) 143–168.