A Fast Algorithm for the Total Variation Model of Image Denoising

Rong-Qing Jia† and Hanqing Zhao†
Department of Mathematical and Statistical Sciences
University of Alberta
Edmonton, Canada T6G 2G1
rjia@ualberta.ca and hzhao@math.ualberta.ca

Abstract

The total variation model of Rudin, Osher, and Fatemi for image denoising is considered to be one of the best denoising models. In the past, its solutions were based on nonlinear partial differential equations and the resulting algorithms were very complicated. In this paper, we propose a fast algorithm for the solution of the total variation model. Our algorithm is very simple and does not involve partial differential equations. We also provide a rigorous proof for the convergence of our algorithm.

Key words and phrases. total variation, optimization, image denoising.

† Supported in part by NSERC Canada under Grant OGP 121336

A Fast Algorithm for the Total Variation Model of Image Denoising

§1. Introduction

In this paper, we propose an efficient algorithm for the solution of the total variation model of Rudin, Osher, and Fatemi [5] for image denoising. We also provide a rigorous proof for the convergence of our algorithm.

An image is regarded as a function

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

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

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

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

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

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 $i = 1, \ldots, N$ and

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

The **total variation** of u is represented by

$$\|\nabla_x u\|_1 + \|\nabla_y u\|_1.$$

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 (Total Variation) 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.

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 convex function. A vector g in $\mathbb{R}^n$ is called a **subgradient** of E at a point $v \in \mathbb{R}^n$ if

$$E(u) - E(v) - \langle g, u - v\rangle \geq 0 \quad \forall\, u \in \mathbb{R}^n. \tag{1.2}$$

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

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

If E is given by $E(u) = |u| + \frac{\lambda}{2}(u-c)^2$, $u \in \mathbb{R}$, where $\lambda > 0$ and $c \in \mathbb{R}$, then $0 \in \partial E(v)$ if and only if $v = shrink(c, 1/\lambda)$, where

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

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

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

Clearly, $shrink(c, 1/\lambda) + cut(c, 1/\lambda) = c$. Let $v = (v_1, \ldots, v_n)$ and $c = (c_1, \ldots, c_n)$ be two vectors in $\mathbb{R}^n$. We write $v = shrink(c, 1/\lambda)$ if $v_i = shrink(c_i, 1/\lambda)$, $i = 1, \ldots, n$. Analogously, we write $v = cut(c, 1/\lambda)$ if $v_i = cut(c_i, 1/\lambda)$, $i = 1, \ldots, n$.

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

Here is an outline of the paper. In Section 2, we propose a simple algorithm for the solution of the total variation model (1.1) and demonstrate that our algorithm is very efficient. In Section 3, we give a rigorous proof for the convergence of our algorithm.

§2. A Simple Algorithm

In order to find the unique solution u^* to the minimization problem:

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

we propose the following iteration scheme: Set $b_x^0 := 0$, $b_y^0 := 0$, and $u^1 := f$. For $k = 1, 2, \ldots$, let

$$b_x^k := cut(\nabla_x u^k + b_x^{k-1}, 1/\lambda), \tag{2.1}$$

$$b_y^k := cut(\nabla_y u^k + b_y^{k-1}, 1/\lambda), \tag{2.2}$$

$$u^{k+1} := f - \frac{\lambda}{\mu}(\nabla_x^T b_x^k + \nabla_y^T b_y^k), \tag{2.3}$$

where ∇_x^T and ∇_y^T are the conjugate operators of ∇_x and ∇_y, respectively. More precisely, ∇_x^T is the linear operator on $\mathbb{R}^{N^2}$ given by

$$\nabla_x^T w(i,j) := \begin{cases} -w(2,j) & \text{if } i = 1, \\ w(i,j) - w(i+1,j) & \text{if } i = 2, \ldots, N-1, \\ w(N,j) & \text{if } i = N, \end{cases}$$

and ∇_y^T is the linear operator on $\mathbb{R}^{N^2}$ given by

$$\nabla_y^T w(i,j) := \begin{cases} -w(i,2) & \text{if } j = 1, \\ w(i,j) - w(i,j+1) & \text{if } j = 2, \ldots, N-1, \\ w(i,N) & \text{if } j = N. \end{cases}$$

Let $\Delta := -\nabla_x^T \nabla_x - \nabla_y^T \nabla_y$. Then Δ is the discrete Laplace operator. For $1 < i, j < N$,

$$-\Delta u(i,j) = 4u(i,j) - u(i+1,j) - u(i-1,j) - u(i,j+1) - u(i,j-1).$$

Moreover, $-\Delta u(1,j) = 3u(1,j) - u(2,j) - u(1,j-1) - u(1,j+1)$ for $1 < j < N$ and $-\Delta u(1,1) = 2u(1,1) - u(1,2) - u(2,1)$. When $i \in \{1, N\}$ or $j \in \{1, N\}$, similar formulas are available for $-\Delta u(i,j)$.

The main result of this paper is the following theorem, which will be proved in Section 3.

Main Theorem. *For $k = 0, 1, \ldots$, let b_x^k, b_y^k, u^{k+1} be given by the iteration scheme (2.1), (2.2) and (2.3). If $0 < \lambda/\mu < 1/8$, then $\lim_{k\to\infty} u^k = u^*$.*

In the past, solutions of the TV model were based on nonlinear partial differential equations and the resulting algorithms were very complicated. A breakthrough was made by Goldstein and Osher in [2]. Using the split Bregman method, they obtained the following iteration scheme: 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

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

Update v_x^{k+1}, v_y^{k+1}, b_x^{k+1}, and b_y^{k+1} as follows:

$$\begin{aligned} v_x^{k+1} &:= shrink(\nabla_x u^{k+1} + b_x^k, 1/\lambda), \\ v_y^{k+1} &:= shrink(\nabla_y u^{k+1} + b_y^k, 1/\lambda), \\ b_x^{k+1} &:= b_x^k - (v_x^{k+1} - \nabla_x u^{k+1}), \\ b_y^{k+1} &:= b_y^k - (v_y^{k+1} - \nabla_y u^{k+1}). \end{aligned}$$

Note that their algorithm still requires solving a partial difference equation in each iteration step. In comparison with their algorithm, our algorithm does not involve partial differential or difference equations.

Let us compare the actual implementation of our algorithm with the algorithm of Goldstein and Osher. In what follows, all the images considered have the size 512×512 and the grey-scale in the range between 0 and 255. A Gaussian noise with the normal distribution $N(0, \sigma^2)$ is added to the original image. We choose $\sigma = 25$. Let u be the original image, and let f be the noised image. By u^{k+1} we denote the result after k iterations.

For image processing, the image quality is usually measured in terms of Peak Signal-to-Noise Ratio (PSNR), which is defined by

$$PSNR = 20 \log_{10}(M/E),$$

where M is the maximum possible pixel value of the image and E is the mean squared error. In our case, $M = 255$ and $E = \|u^{k+1} - u\|_2/512$.

We tested both algorithms on four images: Lena, Barbara, Boat, and Goldhill. With $\sigma = 25$, the PSNR for each noised image is about 20.14. In the following table, the PSNR values are listed after k iterations of our algorithm (JZ_k) and the algorithm of Goldstein and Osher (GO_k). The CPU time (in seconds) needed is listed in the last column.

	Lena	*Barbara*	*Boat*	*Goldhill*	*time*
JZ_{15}	30.23	25.73	28.21	28.61	$0.095s$
GO_{15}	30.22	25.73	28.19	28.61	$0.205s$
JZ_{150}	30.23	25.73	28.18	28.61	$0.953s$
GO_{150}	30.23	25.73	28.18	28.60	$2.047s$

Clearly, 15 iterations are good enough. Moreover, our algorithm requiresless than one half of the time needed for the algorithm of Goldstein and Osher.

§3. Convergence of the Algorithm

In this section, we complete the proof of the Main Theorem. Our proof is motivated by the Bregman method (see [1]). Some basic properties of the Bregman iteration were established in [4]. A fundamental criterion for convergence of the Bregman iteration was given in [3, Theorem 2].

We observe that $\mu + \lambda\Delta$ is a real symmetric linear operator on $\mathbb{R}^{N^2}$. Suppose that η is an eigenvalue of the operator $\mu + \lambda\Delta$. Then η is a real number. We will show $\eta > 0$, provided $0 < \lambda/\mu < 1/8$. There exists a nonzero vector $u \in \mathbb{R}^{N^2}$ such that $(\mu+\lambda\Delta)u = \eta u$. It follows that $(\mu - \eta)u = -\lambda\Delta u$. Let $m := \|u\|_\infty = \max_{1\le i,j\le N} |u(i,j)|$. There exist $i_0, j_0 \in \{1, \dots, N\}$ such that $|u(i_0, j_0)| = m$. If $\eta \le 0$, then $|(\mu-\eta)u(i_0, j_0)| \ge \mu m$. On the other hand, $|-\lambda\Delta u(i_0, j_0)| \le 8\lambda m$. Consequently, $\mu m \le 8\lambda m$. Since $\mu > 8\lambda > 0$, we must have $m = 0$. In other words, $u = 0$. This shows that any eigenvalue of $\mu + \lambda\Delta$ is positive. Therefore, $\mu + \lambda\Delta$ is positive definite. Let B be the unique positive definite operator on $\mathbb{R}^{N^2}$ such that $B^2 = \mu + \lambda\Delta$.

We shall demonstrate that the algorithm given by (2.1), (2.2), and (2.3) has the alternative interpretation described as follows.

Set $v_x^0 = v_y^0 := 0$, $b_x^0 = b_y^0 := 0$, and $u^1 := f$. For $k = 1, 2, \ldots$, let

$$v_x^k := \arg\min_v \Big\{ \|v\|_1 + \frac{\lambda}{2} \|v - \nabla_x u^k - b_x^{k-1}\|_2^2 \Big\}, \tag{3.1}$$

$$v_y^k := \arg\min_v \Big\{ \|v\|_1 + \frac{\lambda}{2} \|v - \nabla_y u^k - b_y^{k-1}\|_2^2 \Big\}, \tag{3.2}$$

$$b_x^k := \nabla_x u^k + b_x^{k-1} - v_x^k, \tag{3.3}$$

$$b_y^k := \nabla_y u^k + b_y^{k-1} - v_y^k, \tag{3.4}$$

and

$$u^{k+1} := \arg\min_u \Big\{ \frac{1}{2} \|B(u-f)\|_2^2 - \langle B^2(u^k - f), u - u^k \rangle + \frac{\lambda}{2} \big\|v_x^k - \nabla_x u\big\|_2^2 + \frac{\lambda}{2} \big\|v_y^k - \nabla_y u\big\|_2^2 \Big\}. \tag{3.5}$$

In the following two lemmas we shall show that the sequences $(b_x^k)_{k=0,1,\ldots}$, $(b_y^k)_{k=0,1,\ldots}$, and $(u^{k+1})_{k=0,1,\ldots}$ satisfy (2.1), (2.2), and (2.3). These results together with Lemma 3 will enable us to prove $\lim_{k\to\infty} u^k = u^*$.

Lemma 1. *For $k = 1, 2, \ldots$, let v_x^k, v_y^k, b_x^k, b_y^k, and u^{k+1} be given by the iteration scheme (3.1) to (3.5). Then $\lim_{k\to\infty}(u^{k+1} - u^k) = 0$. Moreover,*

$$b_x^k = cut(\nabla_x u^k + b_x^{k-1}, 1/\lambda) \quad \textit{and} \quad b_y^k = cut(\nabla_y u^k + b_y^{k-1}, 1/\lambda). \tag{3.6}$$

Proof. It follows from (3.1) and (3.2) that

$$v_x^k = shrink(\nabla_x u^k + b_x^{k-1}, 1/\lambda) \quad \text{and} \quad v_y^k = shrink(\nabla_y u^k + b_y^{k-1}, 1/\lambda).$$

This in connection with (3.3) and (3.4) yields (3.6). Consequently, $\|b_x^k\|_\infty \le 1/\lambda$ and $\|b_y^k\|_\infty \le 1/\lambda$ for $k = 1, 2, \ldots$.

Write $G(v) := \|v\|_1$ for $v \in \mathbb{R}^{N^2}$. Let $g_x^k := \lambda b_x^k$ and $g_y^k := \lambda b_y^k$. It follows from (3.3) and (3.1) that

$$g_x^k = g_x^{k-1} - \lambda(v_x^k - \nabla_x u^k) = -\lambda(v_x^k - \nabla_x u^k - b_x^{k-1}) \in \partial G(v_x^k).$$

Hence, $g_x^k - \lambda(v_x^{k+1} - \nabla_x u^{k+1}) \in \partial G(v_x^{k+1})$ and thereby

$$v_x^{k+1} = \arg\min_v \Big\{ \|v\|_1 - \langle g_x^k, v - v_x^k \rangle + \frac{\lambda}{2} \|v - \nabla_x u^{k+1}\|_2^2 \Big\}. \tag{3.7}$$

Similarly,

$$v_y^{k+1} = \arg\min_v \Big\{ \|v\|_1 - \langle g_y^k, v - v_y^k \rangle + \frac{\lambda}{2} \|v - \nabla_y u^{k+1}\|_2^2 \Big\}. \tag{3.8}$$

It follows from (3.7) that

$$\|v_x^{k+1}\|_1 - \langle g_x^k, v_x^{k+1} - v_x^k \rangle + \frac{\lambda}{2} \|v_x^{k+1} - \nabla_x u^{k+1}\|_2^2 \le \|v_x^k\|_1 + \frac{\lambda}{2} \|v_x^k - \nabla_x u^{k+1}\|_2^2.$$

Since $g_x^k \in \partial G(v_x^k)$, by (1.2) we have $\|v_x^{k+1}\|_1 - \|v_x^k\|_1 - \langle g_x^k, v_x^{k+1} - v_x^k \rangle \geq 0$. Hence,

$$\frac{\lambda}{2}\left\|v_x^{k+1} - \nabla_x u^{k+1}\right\|_2^2 \leq \frac{\lambda}{2}\left\|v_x^k - \nabla_x u^{k+1}\right\|_2^2, \quad k = 1, 2, \ldots. \tag{3.9}$$

For the same reason, we deduce from (3.8) that

$$\frac{\lambda}{2}\left\|v_y^{k+1} - \nabla_y u^{k+1}\right\|_2^2 \leq \frac{\lambda}{2}\left\|v_y^k - \nabla_y u^{k+1}\right\|_2^2, \quad k = 1, 2, \ldots. \tag{3.10}$$

By (3.5) we see that the following inequality is valid for all $u \in \mathbb{R}^{N^2}$:

$$\begin{aligned}\frac{1}{2}\|B(u^{k+1} - f)\|_2^2 - \langle B^2(u^k - f), u^{k+1} - u^k \rangle + \frac{\lambda}{2}\|v_x^k - \nabla_x u^{k+1}\|_2^2 + \frac{\lambda}{2}\|v_y^k - \nabla_y u^{k+1}\|_2^2 \\ \leq \frac{1}{2}\|B(u - f)\|_2^2 - \langle B^2(u^k - f), u - u^k \rangle + \frac{\lambda}{2}\|v_x^k - \nabla_x u\|_2^2 + \frac{\lambda}{2}\|v_y^k - \nabla_y u\|_2^2.\end{aligned}$$

In particular, choosing $u = u^k$ in the above inequality, we obtain

$$\begin{aligned}\frac{1}{2}\|B(u^{k+1} - f)\|_2^2 - \langle B(u^k - f), B(u^{k+1} - u^k) \rangle + \frac{\lambda}{2}\|v_x^k - \nabla_x u^{k+1}\|_2^2 + \frac{\lambda}{2}\|v_y^k - \nabla_y u^{k+1}\|_2^2 \\ \leq \frac{1}{2}\|B(u^k - f)\|_2^2 + \frac{\lambda}{2}\|v_x^k - \nabla_x u^k\|_2^2 + \frac{\lambda}{2}\|v_y^k - \nabla_y u^k\|_2^2.\end{aligned}$$

Note that

$$\frac{1}{2}\|B(u^{k+1} - f)\|_2^2 - \frac{1}{2}\|B(u^k - f)\|_2^2 - \langle B(u^k - f), B(u^{k+1} - u^k) \rangle = \frac{1}{2}\|B(u^{k+1} - u^k)\|_2^2.$$

Hence, we deduce that

$$\frac{1}{2}\|B(u^{k+1} - u^k)\|_2^2 + \frac{\lambda}{2}\|v_x^k - \nabla_x u^{k+1}\|_2^2 + \frac{\lambda}{2}\|v_y^k - \nabla_y u^{k+1}\|_2^2 \leq \gamma_k, \quad k = 1, 2, \ldots,$$

where

$$\gamma_k := \frac{\lambda}{2}\|v_x^k - \nabla_x u^k\|_2^2 + \frac{\lambda}{2}\|v_y^k - \nabla_y u^k\|_2^2. \tag{3.11}$$

This inequality together with (3.9) and (3.10) gives

$$\frac{1}{2}\|B(u^{k+1} - u^k)\|_2^2 + \gamma_{k+1} \leq \gamma_k. \tag{3.12}$$

Thus, $\gamma_1 \geq \gamma_2 \geq \cdots$ and $\gamma_k \geq 0$ for all k. Hence, $\lim_{k\to\infty} \gamma_k$ exists. Furthermore, it follows from (3.12) that

$$\frac{1}{2}\|B(u^{k+1} - u^k)\|_2^2 \leq \gamma_k - \gamma_{k+1}.$$

Since B is positive definite, we conclude that $\lim_{k\to\infty}(u^{k+1} - u^k) = 0$. $\square$

In the following lemma we establish boundedness of relevant sequences.

Lemma 2. *For $k = 1, 2, \ldots$, let v_x^k, v_y^k, b_x^k, b_y^k, and u^{k+1} be given by the iteration scheme (3.1) to (3.5). Then all the sequences $(v_x^k)_{k=1,2,\ldots}$, $(v_y^k)_{k=1,2,\ldots}$, $(b_x^k)_{k=1,2,\ldots}$, $(b_y^k)_{k=1,2,\ldots}$, and $(u^k)_{k=1,2,\ldots}$ are bounded. Furthermore,*

$$u^{k+1} = f - \frac{\lambda}{\mu}(\nabla_x^T b_x^k + \nabla_y^T b_y^k). \tag{3.13}$$

Proof. It was proved in Lemma 1 that $\|\lambda b_x^k\|_\infty \leq 1$ and $\|\lambda b_y^k\|_\infty \leq 1$. For the other parts of the lemma, we deduce from (3.5) that

$$B^2(u^{k+1} - f) - B^2(u^k - f) - \lambda\nabla_x^T(v_x^k - \nabla_x u^{k+1}) - \lambda\nabla_y^T(v_y^k - \nabla_y u^{k+1}) = 0.$$

Consequently,

$$B^2(u^{k+1} - u^k) = \lambda(\nabla_x^T v_x^k + \nabla_y^T v_y^k) + \lambda\Delta u^{k+1}.$$

Recall that $B^2 = \mu + \lambda\Delta$. Hence,

$$\mu(u^{k+1} - u^k) + \lambda\Delta u^{k+1} - \lambda\Delta u^k = \lambda(\nabla_x^T v_x^k + \nabla_y^T v_y^k) + \lambda\Delta u^{k+1}.$$

It can be rewritten as

$$\mu(u^{k+1} - u^k) = \lambda\nabla_x^T(v_x^k - \nabla_x u^k) + \lambda\nabla_y^T(v_y^k - \nabla_y u^k).$$

By (3.3) and (3.4) we have $v_x^k - \nabla_x u^k = b_x^{k-1} - b_x^k$ and $v_y^k - \nabla_y u^k = b_y^{k-1} - b_y^k$. Hence,

$$\mu(u^{k+1} - u^k) = \lambda\nabla_x^T(b_x^{k-1} - b_x^k) + \lambda\nabla_y^T(b_y^{k-1} - b_y^k).$$

It follows that

$$\sum_{j=1}^{k} \mu(u^{j+1} - u^j) = \sum_{j=1}^{k} \left[\lambda\nabla_x^T(b_x^{j-1} - b_x^j) + \lambda\nabla_y^T(b_y^{j-1} - b_y^j)\right],$$

that is,

$$\mu(u^{k+1} - f) = -\lambda\nabla_x^T b_x^k - \lambda\nabla_y^T b_y^k.$$

This establishes (3.13). Since $\|\lambda b_x^k\|_\infty \leq 1$ and $\|\lambda b_y^k\|_\infty \leq 1$, we see that the sequence $(u^k)_{k=1,2,\ldots}$ is bounded. Moreover, by (3.3) and (3.4) we have

$$v_x^k = \nabla_x u^k + (b_x^{k-1} - b_x^k) \quad \text{and} \quad v_y^k = \nabla_y u^k + (b_y^{k-1} - b_y^k).$$

Therefore, the sequences $(v_x^k)_{k=1,2,\ldots}$ and $(v_y^k)_{k=1,2,\ldots}$ are also bounded. □

In the proof of the following lemma, we employ the technique used in [4, Prop. 3.2].

Lemma 3. *For $k = 1, 2, \ldots$, let v_x^k, v_y^k, b_x^k, b_y^k, and u^{k+1} be given by the iteration scheme (3.1) to (3.5). Then*

$$\lim_{k\to\infty}(v_x^k - \nabla_x u^k) = 0 \quad \textit{and} \quad \lim_{k\to\infty}(v_y^k - \nabla_y u^k) = 0.$$

Proof. For $G(v) := \|v\|_1$, recall that $g_x^j = \lambda b_x^j \in \partial G(v_x^j)$, $j = 1, 2, \ldots$. It is easily seen that

$$\begin{aligned}
&\big[G(v) - G(v_x^{j+1}) - \langle g_x^{j+1}, v - v_x^{j+1}\rangle\big] \\
&\quad - \big[G(v) - G(v_x^j) - \langle g_x^j, v - v_x^j\rangle\big] \\
&\quad + \big[G(v_x^{j+1}) - G(v_x^j) - \langle g_x^j, v_x^{j+1} - v_x^j\rangle\big] \\
&= \langle g_x^j - g_x^{j+1}, v - v_x^{j+1}\rangle.
\end{aligned}$$

Since $G(v_x^{j+1}) - G(v_x^j) - \langle g_x^j, v_x^{j+1} - v_x^j\rangle \geq 0$, we have

$$\langle g_x^j - g_x^{j+1}, v - v_x^{j+1}\rangle \geq -\|v_x^{j+1}\|_1 - \langle g_x^{j+1}, v - v_x^{j+1}\rangle + \|v_x^j\|_1 + \langle g_x^j, v - v_x^j\rangle. \tag{3.14}$$

But (3.3) implies $g_x^j - g_x^{j+1} = \lambda(v_x^{j+1} - \nabla_x u^{j+1})$. Hence, by (1.2) we have

$$\frac{\lambda}{2}\big\|v - \nabla_x u^{j+1}\big\|_2^2 - \frac{\lambda}{2}\big\|v_x^{j+1} - \nabla_x u^{j+1}\big\|_2^2 - \langle g_x^j - g_x^{j+1}, v - v_x^{j+1}\rangle \geq 0 \quad \forall v \in \mathbb{R}^{N^2}.$$

Combining the above inequality with (3.14), we see that for all $v \in \mathbb{R}^{N^2}$,

$$\frac{\lambda}{2}\big\|v - \nabla_x u^{j+1}\big\|_2^2 - \frac{\lambda}{2}\big\|v_x^{j+1} - \nabla_x u^{j+1}\big\|_2^2 \geq -\|v_x^{j+1}\|_1 - \langle g_x^{j+1}, v - v_x^{j+1}\rangle + \|v_x^j\|_1 + \langle g_x^j, v - v_x^j\rangle.$$

Choosing $v := \nabla_x u^{j+1}$ in the above inequality, we obtain

$$\frac{\lambda}{2}\big\|v_x^{j+1} - \nabla_x u^{j+1}\big\|_2^2 \leq \|v_x^{j+1}\|_1 + \langle g_x^{j+1}, \nabla_x u^{j+1} - v_x^{j+1}\rangle - \|v_x^j\|_1 - \langle g_x^j, \nabla_x u^{j+1} - v_x^j\rangle.$$

With $\beta_j := \langle g_x^j, \nabla_x u^j - v_x^j\rangle$, the above inequality can be rewritten as

$$\frac{\lambda}{2}\big\|v_x^{j+1} - \nabla_x u^{j+1}\big\|_2^2 \leq (\|v_x^{j+1}\|_1 - \|v_x^j\|_1) + (\beta_{j+1} - \beta_j) - \langle g_x^j, \nabla_x(u^{j+1} - u^j)\rangle.$$

Consequently, for $1 \leq m < k$, we have

$$\sum_{j=m}^{k-1}\frac{\lambda}{2}\big\|v_x^{j+1} - \nabla_x u^{j+1}\big\|_2^2 \leq \sum_{j=m}^{k-1}\big[(\|v_x^{j+1}\|_1 - \|v_x^j\|_1) + (\beta_{j+1} - \beta_j) - \langle g_x^j, \nabla_x(u^{j+1} - u^j)\rangle\big].$$

It follows that

$$\sum_{j=m}^{k-1}\frac{\lambda}{2}\big\|v_x^{j+1} - \nabla_x u^{j+1}\big\|_2^2 \leq (\|v_x^k\|_1 - \|v_x^m\|_1) + (\beta_k - \beta_m) - \sum_{j=m}^{k-1}\langle g_x^j, \nabla_x(u^{j+1} - u^j)\rangle.$$

By Lemma 2, the sequences $(u^j)_{j=1,2\ldots}$, $(v_x^j)_{j=1,2\ldots}$ and $(g_x^j)_{j=1,2\ldots}$ are bounded. Hence, there exist positive constants C_1 and C_2 independent of k and m such that

$$\sum_{j=m}^{k-1} \frac{\lambda}{2} \left\| v_x^{j+1} - \nabla_x u^{j+1} \right\|_2^2 \le C_1 + C_2(k-m)\eta_m, \tag{3.15}$$

where $\eta_m := \sup_{j\ge m} \|u^{j+1} - u^j\|_2$. By Lemma 1, we have $\lim_{m\to\infty} \eta_m = 0$. In an analogous way, we derive that

$$\sum_{j=m}^{k-1} \frac{\lambda}{2} \left\| v_y^{j+1} - \nabla_y u^{j+1} \right\|_2^2 \le C_1 + C_2(k-m)\eta_m. \tag{3.16}$$

Let γ_k $(k = 1, 2, \ldots)$ be defined as in (3.11). Adding (3.15) and (3.16) together gives

$$\sum_{j=m}^{k-1} \gamma_{j+1} \le 2C_1 + 2C_2(k-m)\eta_m.$$

By (3.12) we have $\gamma_k \le \gamma_j$ for $j \le k$. Hence, $(k-m)\gamma_k \le 2C_1 + 2C_2(k-m)\eta_m$, that is,

$$\gamma_k \le \frac{2C_1}{k-m} + 2C_2\eta_m.$$

Choosing m to be the integer part of $k/2$, we obtain $\lim_{k\to\infty} \gamma_k = 0$. This completes the proof of the lemma. □

We are in a position to prove the Main Theorem.

Let $F(u) := (\mu/2)\|u - f\|_2^2$. Then $\partial F(u) = \mu(u - f)$. For $w \in \mathbb{R}^{N^2}$ we have

$$F(u^{k+1} + w) - F(u^{k+1}) - \langle \mu(u^{k+1} - f), w \rangle \ge 0.$$

By Lemma 2, $-\mu(u^{k+1} - f) = \lambda \nabla_x^T b_x^k + \lambda \nabla_y^T b_y^k$. Hence,

$$F(u^{k+1} + w) - F(u^{k+1}) + \langle \lambda b_x^k, \nabla_x w \rangle + \langle \lambda b_y^k, \nabla_y w \rangle \ge 0. \tag{3.17}$$

Recall that $G(v) = \|v\|_1$, $\lambda b_x^k \in \partial G(v_x^k)$ and $\lambda b_y^k \in \partial G(v_y^k)$. Consequently,

$$\|v_x^k + \nabla_x w\|_1 - \|v_x^k\|_1 - \langle \lambda b_x^k, \nabla_x w \rangle \ge 0, \tag{3.18}$$

and

$$\|v_y^k + \nabla_y w\|_1 - \|v_y^k\|_1 - \langle \lambda b_y^k, \nabla_y w \rangle \ge 0. \tag{3.19}$$

Adding (3.17), (3.18), and (3.19) together gives

$$\|v_x^k + \nabla_x w\|_1 - \|v_x^k\|_1 + \|v_y^k + \nabla_y w\|_1 - \|v_y^k\|_1 + F(u^{k+1} + w) - F(u^{k+1}) \ge 0,$$

that is,

$$\|v_x^k\|_1 + \|v_y^k\|_1 + F(u^{k+1}) \le \|v_x^k + \nabla_x w\|_1 + \|v_y^k + \nabla_y w\|_1 + F(u^{k+1} + w). \tag{3.20}$$

Suppose that $(k_j)_{j=1,2,\dots}$ is an increasing sequence of positive integers such that the sequence $(u^{k_j})_{j=1,2,\dots}$ converges to the limit $\tilde{u}$. By Lemma 1, $\lim_{k\to\infty}(u^{k+1}-u^k) = 0$. Hence, $\lim_{j\to\infty} u^{k_j+1} = \tilde{u}$. Moreover, Lemma 3 tells us that

$$\lim_{j\to\infty} v_x^{k_j} = \lim_{j\to\infty} \left[(v_x^{k_j} - \nabla_x u^{k_j}) + \nabla_x u^{k_j}\right] = \nabla_x \tilde{u} \quad \text{and} \quad \lim_{j\to\infty} v_y^{k_j} = \nabla_y \tilde{u}.$$

Replacing k by k_j in (3.20) and letting $j \to \infty$, we obtain

$$\|\nabla_x \tilde{u}\|_1 + \|\nabla_y \tilde{u}\|_1 + F(\tilde{u}) \le \|\nabla_x(\tilde{u} + w)\|_1 + \|\nabla_y(\tilde{u} + w)\|_1 + F(\tilde{u} + w).$$

This is true for all $w \in \mathbb{R}^{N^2}$. On the other hand, u^* is the unique solution to the minimization problem (1.1). Therefore, we must have $\tilde{u} = u^*$. Since $(u^k)_{k=1,2,\dots}$ is a bounded sequence, we conclude

$$\lim_{k\to\infty} u^k = u^*.$$

This completes the proof of the Main Theorem.

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 Computational Mathematics and Mathematical Physics **7** (1967), 200–217.

[2] T. Goldstein and S. Osher, The split Bregman method for L1 regularized problems, UCLA CAM Report (08-29).

[3] R. Q. Jia, H. Q. Zhao, and W. Zhao, Convergence analysis of the Bregman method for the variational model of image denoising, preprint.

[4] S. Osher, M. Burger, D. Goldfarb, J. J. Xu, and W. T. Yin, An iterative regularization method for total variation-based image restoration, Multiscale Model Simulation **4** (2005), 460–489.

[5] L. Rudin, S. Osher, and E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. D **60** (1992), 259-268.