A Proximity Algorithm Accelerated by Gauss-Seidel Iterations for L1/TV Denoising Models*

Qia Li [†] Charles A. Micchelli[‡] Lixin Shen[§†] Yuesheng Xu [†§¶]

January 10, 2012

Abstract

Our goal in this paper is to improve the computational performance of the proximity algorithms for the L1/TV denoising model. This leads us to a new characterization of all solutions to the L1/TV model via fixed-point equations expressed in terms of proximity operators. Based upon this observation we develop an algorithm for solving the model and establish its convergence. Furthermore, we demonstrate that the proposed algorithm can be accelerated through the use of the *component-wise* Gauss-Seidel iteration so that the CPU-time consumed is significantly reduced. Numerical experiments using the proposed algorithm for impulsive noise removal are included, with a comparison to two recently developed algorithms. The numerical results show that while the proposed algorithm enjoys a high quality of the restored images, as the other two known algorithms do, it performs significantly better in terms of computational efficiency measured in the CPU-time consumed.

1 Introduction

The L1/TV image denoising model has attracted considerable attention due to its various attractive mathematical properties such as contrast invariant, data driven parameter selection, multiscale image decomposition, and morphological invariance [5, 13, 18]. Moreover, it has successful applications in impulsive noise removal [1, 13], computer vision [6], biomedical imaging [17], optical flow and object tracking [19], and shape denoising [19]. The L1/TV image denoising model is an optimization problem with an objective function being the sum of a data fidelity term and a regularization term together with a regularization parameter balancing these terms. The data fidelity term is defined as the distance, measured by the ℓ^1-norm, between a given noisy image and the ideal one to be sought. The regularization term is the total-variation of the ideal image [14]. A noticeable feature of the model is that both terms are not differentiable. This makes the model particularly useful in the above applications in spite of the fact that finding its solutions is a challenging task both mathematically and numerically.

*This research is supported in part by the US National Science Foundation under grant DMS-0712827 and DMS-1115523, by Guangdong Provincial Government of China through the "Computational Science Innovative Research Team" program.

[†]Guangdong Province Key Laboratory of Computational Science, School of Mathematics and Computational Sciences, Sun Yat-sen University, Guangzhou 510275, P. R. China.

[‡]Department of Mathematics and Statistics, SUNY at Albany, Albany, NY 12222

[§]Department of Mathematics, Syracuse University, Syracuse, NY 13244.

[¶]All correspondence should be sent to this author. E-mail address: yxu06@syr.edu.

In this paper, we are interested in developing a numerically efficient algorithm for the L1/TV model. Depending on how the non-differentiable fidelity and regularization terms of the image denoising model are treated, algorithms proposed in the recent literature for the L1/TV model can be grouped into two categories. In category one, prior to developing a numerical scheme for the model, additional quadratic terms with some auxiliary variables are added to the objective function of the model to take care of either the non-differentiability of its fidelity or regularization term, or both. For instance, in [2, 7], a quadratic term involving an auxiliary variable was introduced to overcome the non-differentiability of the fidelity term. Whether or not we view the total-variation as a non-differentiable function or as a composition of a non-differentiable function with a first-order difference matrix, a quadratic term was added to the model to serve the purpose of remedying the regularization term [11]. The non-differentiability of both the fidelity and regularization terms was tackled simultaneously by introducing two extra quadratic terms to the primal formulation of the model in [16] and to its dual formulation in [8]. An essential idea for these treatments, as pointed out in [11], is to replace the non-differentiable functions in the model by their corresponding differentiable Moreau envelopes which were originally introduced in [12]. This modification will certainly simplify the numerical treatment of the model, but, unfortunately it introduces additional parameters whose fine-tuning complicates the use of the resulting algorithms. Moreover, replacing the ℓ^1-norm in the original model by its Moreau envelope may reduce the efficiency of the original model in denoising and preserving edges of the images being treated. Unlike this category, algorithms in category two are faithfully solving the L1/TV model directly without introducing additional terms to the model. These algorithms were developed by using either the proximity operator [11], the primal-dual formulation [4], or the augmented Lagrangian method [15]. In contrast to category one, the algorithms in category two are particularly suitable for cartoon-like images. This observation will be verified in the numerical experiment section of this paper via numerical examples. It is our goal of this paper to develop a mathematically rigorous and computationally efficient algorithm for the L1/TV model.

To motivate our work for the L1/TV model, we provide additional comments on some aforementioned algorithms in category two. The algorithm presented in [11] is the Picard iteration of the fixed-point equations arising from a characterization of the solutions to the L1/TV model. This characterization was derived by exploiting the proximity operator. The algorithm described in [4] originated from the primal-dual formulation of the model in a way that the primal and dual variables of the formulation are alternatively updated. Both of these algorithms can be viewed as a block-wise Gauss-Seidel iteration in the sense that once one variable in either the fixed-point equations in [11] or the primal-dual formulation in [4] is updated via the Jacobi iteration, it is immediately used to generate the other associated variables which is also updated via the Jacobi iteration. Interestingly, for the algorithms in [4, 11], updating each variable in the algorithms via the Jacobi iteration is identical to updating this variable via the Gauss-Seidel iteration. In other words, the formulation of these algorithms will not lead to real component-wise Gauss-Seidel iteration schemes for solving the model. As we know, the component-wise Gauss-Seidel iteration usually can speed up convergence of an algorithm. We are therefore interested in a proximity algorithm for the L1/TV model accelerated via the component-wise Gauss-Seidel iteration. To this end, we formulate an alternative characterization for solutions of the model by introducing a new variable for the characterization given in [11]. We then propose an algorithm based on the new characterization. The new algorithm will lead to a component-wise Gauss-Seidel iteration scheme which improves the convergence of the algorithm. We shall confirm this by numerical examples in Section 5.

We organize this paper in six sections. In Section 2 we present an alternative characterization of the solutions of the L1/TV model. Based on this new characterization we develop a new algorithm

which allows component-wise Gauss-Seidel iteration. A convergence result for the algorithm is proved in Section 3. We provide the resulting component-wise Gauss-Seidel iteration scheme in Section 4. Numerical experiments are presented in Section 5 to demonstrate the quality of the restored images and the computational efficiency of the proposed algorithm. Our conclusion is presented in Section 6.

2 A Characterization and an Algorithm

In this section, we first describe the L1/TV model for image denoising and briefly review a characterization of its solutions, which was originally established in [11]. Following this discussion, we provide a new characterization for solutions of the model which will lead to a more computationally efficient algorithm.

We begin with a description of notation used in this paper. An image is viewed as a vector in $\mathbb{R}^m$ by sequentially concatenating the columns of the image. We use $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$, respectively, to denote the inner product and the corresponding ℓ^2-norm of an Euclidean space $\mathbb{H}$ while $\|\cdot\|_1$ is used to denote the ℓ^1-norm.

For a proper convex function ψ defined on an Euclidean space $\mathbb{H}$, that is, $\psi : \mathbb{H} \to \mathbb{R} \cup \{+\infty\}$, having a non-empty domain (the set on which ψ is finite), the proximity operator of ψ, denoted by prox_ψ, is a mapping from $\mathbb{H}$ to itself, defined for a given vector $x \in \mathbb{H}$ by

$$\text{prox}_\psi(x) := \arg\min\left\{\psi(u) + \frac{1}{2}\|u - x\|^2 : u \in \mathbb{H}\right\}.$$

The subdifferential of this function ψ at a given vector $x \in \mathbb{H}$ is the set defined by

$$\partial\psi(x) := \{y : \ y \in \mathbb{H} \quad \text{and} \quad \psi(z) \geq \psi(x) + \langle y, z - x\rangle \quad \text{for all} \quad z \in \mathbb{H}\}.$$

The subdifferential and the proximity operator of the function ψ are intimately related. Specifically, for x in the domain of ψ and $y \in \mathbb{H}$ we have that

$$y \in \partial\psi(x) \quad \text{if and only if} \quad x = \text{prox}_\psi(x + y), \tag{1}$$

for a discussion of this relation, please see, e.g., [10] or [3, Proposition 16.34].

The variational problem that we consider in this paper is

$$\min\{\lambda\|u - x\|_1 + (\varphi \circ B)(u) : u \in \mathbb{R}^m\}, \tag{2}$$

where x is a given vector in $\mathbb{R}^m$, λ is a positive number, φ is a proper convex function on $\mathbb{R}^n$, and B is an $n \times m$ matrix. In the L1/TV model, B is a $2m \times m$ first-order difference matrix and φ the ℓ^1-norm for the anisotropic total-variation or the ℓ^2-norm for the isotropic total-variation on $\mathbb{R}^{2m}$. Consequently, the function $\varphi \circ B$ is a generic form of the total-variation regularization term appearing in the image denoising model (see, e.g., [10] or Section 4 for more detail). For this reason, we call the above variational problem the L1/TV model. Certainly, the solutions of the L1/TV model exist but may not be unique.

A characterization of a solution of the L1/TV model was proved in [11] in terms of a coupled system of fixed-point equations using the proximity operator. This characterization leads to a Picard iteration algorithm for finding a solution of the L1/TV model. Our principal goal is to provide an alternative characterization of a solution of the L1/TV model which will allow us to accelerate the resulting proximity algorithm by using a component-wise Gauss-Seidel methodology. To this end, we recall the characterization theorem established in [11].

Proposition 2.1. *Let $x \in \mathbb{R}^m$ and let λ be a positive number. If $u \in \mathbb{R}^m$ is a solution of the variational problem (2), then for any $\alpha, \beta > 0$, there exists a vector $b \in \mathbb{R}^n$ such that*

$$\begin{aligned} u &= x + \operatorname{prox}_{\frac{1}{\alpha}\|\cdot\|_1}\left(u - x - \rho B^\top b\right), && (3)\\ b &= (I - \operatorname{prox}_{\frac{1}{\beta}\varphi})(Bu + b), && (4) \end{aligned}$$

where

$$\rho := \frac{\beta}{\lambda\alpha}. \tag{5}$$

Conversely, if there exist $\alpha > 0$, $\beta > 0$, $b \in \mathbb{R}^n$, and $u \in \mathbb{R}^m$ satisfying equations (3) and (4), then u is a solution of the variational problem (2).

The fixed-point equations (3) and (4) immediately yield the iterative algorithm proposed in [11] for a solution of the variational problem. Specifically, for given vectors $u^0 \in \mathbb{R}^m$ and $b^0 \in \mathbb{R}^n$, and for $k \in \mathbb{N}_0 := \{0\} \cup \mathbb{N}$, where $\mathbb{N}$ is the set of all natural numbers, we compute

$$\begin{cases} u^{k+1} = x + \operatorname{prox}_{\frac{1}{\alpha}\|\cdot\|_1}(u^k - x - \rho B^\top b^k), \\ b^{k+1} = (I - \operatorname{prox}_{\frac{1}{\beta}\varphi})(Bu^{k+1} + b^k). \end{cases} \tag{6}$$

The implementation of algorithm (6) requires the availability of a convenient way of computing $\operatorname{prox}_{\frac{1}{\alpha}\|\cdot\|_1}$ and $\operatorname{prox}_{\frac{1}{\beta}\varphi}$. Reference [10] provides an explicit formula for computing $\operatorname{prox}_{\frac{1}{\alpha}\|\cdot\|_1}$ which can be identified as the well-known shrinkage operator with the threshold $\frac{1}{\alpha}$ [9]. Although it is not easy to compute $\operatorname{prox}_{\frac{1}{\beta}\varphi}$ for a general function φ, in the image processing context, the function φ is either $\|\cdot\|_1$ or a variant of the ℓ^1 norm, and thus, $\operatorname{prox}_{\frac{1}{\beta}\varphi}$ can be easily computed. This will be seen in Section 4.

This iterative algorithm generates a sequence $\mathcal{D} := (u^k : k \in \mathbb{N})$ in $\mathbb{R}^m$ and a sequence $\mathcal{B} := (b^k : k \in \mathbb{N})$ in $\mathbb{R}^n$, which we couple together as the sequence $\mathcal{U} \times \mathcal{B} := ((u^k, b^k) : k \in \mathbb{N})$ in $\mathbb{R}^m \times \mathbb{R}^n$. Further details, along comments on how the parameters α and β in this algorithm are iteratively updated, will be provided in Section 4, where the corresponding algorithm is referred to as the MSXZ.

A careful examination of the MSXZ shows that the component-wise Gauss-Seidel iteration applied to the algorithm is just the same as the original algorithm. In other words, the component-wise Gauss-Seidel iteration applied to the algorithm will not change the algorithm at all. This motivates us to reformulate the characterization of a solution of the L1/TV model presented in Proposition 2.1 in order to be able to accelerate the algorithm using a Gauss-Seidel iteration strategy. To this end, for a positive number γ and an $n \times m$ matrix, we define

$$B_\gamma := I - \gamma B^\top B.$$

We establish an alternative characterization of a solution of the L1/TV model in the following proposition.

Proposition 2.2. *Let x be in $\mathbb{R}^m$ and let λ be a positive number. If $u \in \mathbb{R}^m$ is a solution of the variational problem (2), then for any positive numbers α and $\beta > 0$, there exist two vectors d, b in $\mathbb{R}^n$ such that*

$$\begin{aligned} u &= x + \operatorname{prox}_{\frac{1}{\alpha}\|\cdot\|_1}(B_\rho u - x - \rho B^\top(b - d)), && (7)\\ d &= \operatorname{prox}_{\frac{1}{\beta}\varphi}(Bu + b), && (8)\\ b &= Bu + b - d, && (9) \end{aligned}$$

where ρ is the positive real number defined in (5). Conversely, if there exist $\alpha > 0$, $\beta > 0$, $d \in \mathbb{R}^n$, $b \in \mathbb{R}^n$, and $u \in \mathbb{R}^m$ satisfying equations (7), (8) and (9), then u is a solution to the variational problem (2).

Proof. We reformulate equations (3) and (4) in Proposition 2.1 to their equivalent equations (7), (8) and (9). To this end, we introduce a vector $d \in \mathbb{R}^d$ such that

$$d = \operatorname{prox}_{\frac{1}{\beta}\varphi}(Bu + b).$$

By equation (4), it holds that $d = Bu$ which is equation (9). With this relation, equation (3) can be rewritten as

$$u = x + \operatorname{prox}_{\frac{1}{\alpha}\|\cdot\|_1}(B_\rho u - x - \rho B^\top(b - d)).$$

Conversely, equations (7), (8) and (9) yield, after d is eliminated, equations (3) and (4). □

Based on the fixed-point equations (7), (8) and (9) with respect to the triple (u, d, b), we generate a sequence $\mathcal{U}$ in $\mathbb{R}^m$ and two sequences $\mathcal{D}$ and $\mathcal{B}$ in $\mathbb{R}^n$ from arbitrary initial vectors $u^0 \in \mathbb{R}^m$, $d^0 \in \mathbb{R}^n$, and $b^0 \in \mathbb{R}^n$. The iteration is stated as follows:

$$\begin{cases} u^{k+1} = x + \operatorname{prox}_{\frac{1}{\alpha}\|\cdot\|_1}(B_\rho u^k - x - \rho B^\top(b^k - d^k)), \\ d^{k+1} = \operatorname{prox}_{\frac{1}{\beta}\varphi}(Bu^{k+1} + b^k), \\ b^{k+1} = Bu^{k+1} + b^k - d^{k+1}. \end{cases} \tag{10}$$

In the next section we shall analyze the convergence of this Picard iteration.

We point out a major difference of the characterizations given in Proposition 2.1 and Proposition 2.2 in terms of their corresponding iterative schemes (6) and (10). In viewing the first equation of (6), the i-th entry of u^k contributes only to the i-th entry of u^{k+1} due to the fact that the proximity operator $\operatorname{prox}_{\frac{1}{\alpha}\|\cdot\|_1}$ is performed component-wise. In contrast, in the first equation of (10), since the vector u^k is multiplied by the matrix B_ρ from the left, the i-th entry and those from its neighbor have an impact to the i-th entry of u^{k+1}. From the matrix point of view, the term u^k in the first equation of (6) is changed to the term $B_\rho u^k$ in the first equation of (10), that is, the matrix I is changed to B_ρ. Hence, the special structure of B_ρ can vary by adjusting the parameters α and β. For example, when B is the first order difference matrix (see e.g., [10]), we know that the eigenvalues of $B^\top B$ lie in the interval $[0, 8)$. As a result, if we choose α and β such that $\rho \leq 1/4$, the eigenvalues of B_ρ will be in the interval $[-1, 1)$. Therefore, for the later case, we shall be able to design a component-wise Gauss-Seidel iteration based on the iterative scheme (10) and the resulting component-wise Gauss-Seidel iteration scheme may accelerate the convergence of the scheme. We shall verify this point numerically in Section 5.

To close this section, we remark that u^{k+1}, d^{k+1}, and b^{k+1} in (10) can be viewed as the unique solutions to the following minimization problems

$$\begin{cases} u^{k+1} = \arg\min\left\{\|u - x\|_1 + \frac{\alpha}{2}\|u - B_\rho u^k + \rho B^\top(b^k - d^k)\|^2 : u \in \mathbb{R}^m\right\}, \\ d^{k+1} = \arg\min\left\{\varphi(d) + \frac{\beta}{2}\|d - (Bu^{k+1} + b^k)\|^2 : d \in \mathbb{R}^n\right\}, \\ b^{k+1} = \arg\min\left\{\frac{\beta}{2}\|b - (Bu^{k+1} + b^k - d^{k+1})\|^2 : b \in \mathbb{R}^n\right\}, \end{cases} \tag{11}$$

where the first two minimization problems are obtained directly from the first two equations in (10) by using the definition of the proximity operator. This point of view is helpful for proving the convergence of iteration (10) which will be done in the next section.

3 Convergence Analysis

This section is devoted to the convergence analysis of the iterative scheme (10). For the proof of this property we rely upon the equivalent form (11).

We first introduce notation which will be frequently used in this section. Let α, β, and λ be given positive numbers. For any two vectors (y_1, z_1) and (y_2, z_2) in $\mathbb{R}^m \times \mathbb{R}^n$, we define

$$r((y_1, z_1), (y_2, z_2)) \quad := \quad \frac{\lambda\alpha}{2}\|y_2 - y_1\|^2 + \frac{\beta}{2}\|z_2 - z_1\|^2 \tag{12}$$

and

$$a((y_1, z_1), (y_2, z_2)) \quad := \quad \langle B(y_2 - y_1), z_2 - z_1\rangle. \tag{13}$$

In this notation we do not explicitly show the dependence of r and a on the real numbers α, β, λ since they are fixed throughout our discussion.

Clearly, we have the symmetry properties that

$$r((y_1, z_1), (y_2, z_2)) = r((y_2, z_2), (y_1, z_1))$$

and

$$a((y_1, z_1), (y_2, z_2)) = a((y_2, z_2), (y_1, z_1)).$$

Furthermore, we define a function L on $\mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^n$ at (u, d, b) as

$$L(u, d, b) := \lambda\|u - x\|_1 + \varphi(d) + \beta\langle Bu - d, b\rangle. \tag{14}$$

The following lemmas will be used in establishing the convergence result of the iterative scheme (11). The first lemma gives information about the functions r and a defined above. To state our observation we introduce the constant

$$\tau := \|B\|\sqrt{\rho}, \tag{15}$$

where $\|B\|$ is the operator norm $\max\{\|Bx\| : \|x\| \le 1\}$ of the matrix B.

Lemma 3.1. *If (y_1, z_1) and (y_2, z_2) are two vectors in $\mathbb{R}^m \times \mathbb{R}^n$ then*

$$|a((y_1, z_1), (y_2, z_2))| \le \frac{\tau}{\beta} r((y_1, z_1), (y_2, z_2)).$$

Proof. By the Cauchy-Schwarz inequality, we have that

$$|a((y_1, z_1), (y_2, z_2))| \le \|B\|\|y_2 - y_1\|\|z_2 - z_1\|.$$

It follows by the geometric-arithmetic inequality, for any $\gamma > 0$, that

$$|a((y_1, z_1), (y_2, z_2))| \le \|B\| \left(\frac{\gamma}{2}\|y_2 - y_1\|^2 + \frac{1}{2\gamma}\|z_2 - z_1\|^2\right).$$

In particular, when we choose $\gamma := 1/\sqrt{\rho}$, the result of this lemma follows from the above inequality, (12), and (15). □

We next present a useful property of a proper convex function.

Lemma 3.2. *Let q be a proper convex function on an Euclidean space $\mathbb{H}$, let u_0 and $u_\star$ be two points in $\mathbb{H}$, and let α be a positive number. Then $u_\star = \mathrm{prox}_{\frac{1}{\alpha}q}(u_0)$ if and only if for all $u \in \mathbb{H}$*

$$q(u) \geq q(u_\star) + \alpha\langle u_0 - u_\star, u - u_\star\rangle. \tag{16}$$

Moreover, if we define $Q := q + \frac{\alpha}{2}\|\cdot - u_0\|^2$, then (16) is equivalent to the following inequality

$$Q(u) \geq Q(u_\star) + \frac{\alpha}{2}\|u - u_\star\|^2. \tag{17}$$

Proof. First, by (1), $u_\star = \mathrm{prox}_{\frac{1}{\alpha}q}(u_0)$ if and only if $(u_0 - u_\star) \in \partial(\frac{1}{\alpha}q)(u_\star)$, i.e., $\alpha(u_0 - u_\star) \in \partial q(u_\star)$. This is equivalent to, by the definition of the subdifferential, inequality (16).

In the remaining part of the proof, we just need to show that inequalities (16) are (17) are equivalent. To this end, by adding $\frac{\alpha}{2}\|u - u_0\|^2$ to the both sides of inequality (16), we find that

$$Q(u) \geq Q(u_\star) + \frac{\alpha}{2}\|u - u_0\|^2 - \frac{\alpha}{2}\|u_\star - u_0\|^2 + \alpha\langle u_0 - u_\star, u - u_\star\rangle.$$

By the law of cosine, the last three terms in the right of the above inequality is equal to $\frac{\alpha}{2}\|u - u_\star\|^2$. Thus, the above inequality is the exactly same as (17). Therefore, the equivalence of (16) and (17) follows. □

We remark that inequality (16) appeared in [3, Proposition 12.26]. Here, we give an alternative proof via Fermat's rule.

We now return to the minimization problem (11). For a given triple $(u_0, d_0, b_0) \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^n$, we define a triple $(u_\star, d_\star, b_\star) \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^n$ as follows:

$$\begin{cases} u_\star = \arg\min\left\{\|u - x\|_1 + \frac{\alpha}{2}\|u - B_\rho u_0 + \rho B^\top(b_0 - d_0)\|^2 : u \in \mathbb{R}^m\right\}, \\ d_\star = \arg\min\left\{\varphi(d) + \frac{\beta}{2}\|d - (Bu_\star + b_0)\|^2 : d \in \mathbb{R}^n\right\}, \\ b_\star = \arg\min\left\{\frac{\beta}{2}\|b - (Bu_\star + b_0 - d_\star)\|^2 : b \in \mathbb{R}^n\right\}. \end{cases} \tag{18}$$

Clearly, referring back to (11) we see that when the vector $(u_0, d_0, b_0) \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^n$ is chosen to be the vector (u^k, d^k, b^k), the vector $(u_\star, d_\star, b_\star)$ defined by (18) is $(u^{k+1}, d^{k+1}, b^{k+1})$, which is generated by the iteration (11).

For each $u \in \mathbb{R}^m$, $d \in \mathbb{R}^n$, and $b \in \mathbb{R}^n$, we define

$$\begin{aligned} F(u) &:= \lambda\|u - x\|_1 + \langle \beta B^\top(b_0 + Bu_0 - d_0), u\rangle + \frac{\lambda\alpha}{2}\|u - u_0\|^2, \\ G(d) &:= \varphi(d) + \frac{\beta}{2}\|d - (Bu_\star + b_0)\|^2, \\ H(b) &:= -\beta\langle Bu_\star - d_\star, b\rangle + \frac{\beta}{2}\|b - b_0\|^2. \end{aligned}$$

By expanding the quadratic terms in the first and third equations of (18) followed by dropping off the constant terms with respect to the variables to be minimized, we obtain that

$$\begin{aligned} u_\star &= \arg\min\{F(u) : u \in \mathbb{R}^m\}, \\ d_\star &= \arg\min\{G(d) : d \in \mathbb{R}^n\}, \\ b_\star &= \arg\min\{H(b) : b \in \mathbb{R}^n\}, \end{aligned}$$

Applying Lemma 3.2 to the functions F, G, and H, we derive several inequalities which will be useful later in the proof of the convergence result. First by using inequality (16), we have that

$$\varphi(d) \geq \varphi(d_\star) + \beta\langle d - d_\star, Bu_\star + b_0 - d_\star\rangle.$$

Replacing d in the above inequality by Bu and using the relation of $b_\star + d_\star = Bu_\star + b_0$, which is a direct consequence of the optimality of $b_\star$, we obtain

$$0 \geq \varphi(d_\star) - \varphi(Bu) + \beta\langle Bu - d_\star, b_\star\rangle \tag{19}$$

Next, by the definitions of F and H, we have that

$$F(u) + H(b) = r((u_0, b_0), (u, b)) + \lambda\|u - x\|_1 + \langle\beta B^\top(b_0 + Bu_0 - d_0), u\rangle - \beta\langle Bu_\star - d_\star, b\rangle. \tag{20}$$

By the optimality of $u_\star$ and $b_\star$ and inequality (17) in Lemma 3.2, we have for any pair $(u, b) \in \mathbb{R}^m \times \mathbb{R}^n$ that

$$F(u) \geq F(u_\star) + \frac{\lambda\alpha}{2}\|u - u_\star\|^2 \quad \text{and} \quad H(b) \geq H(b_\star) + \frac{\beta}{2}\|b - b_\star\|^2,$$

which yield

$$F(u) + H(b) \geq F(u_\star) + H(b_\star) + r((u_\star, b_\star), (u, b)).$$

Applying (20) to the above inequality, we have that

$$\mathcal{E} \geq \lambda\|u_\star - x\|_1 - \lambda\|u - x\|_1 + \langle\beta B^\top(b_0 + Bu_0 - d_0), u_\star - u\rangle - \beta\langle Bu_\star - d_\star, b_\star - b\rangle \tag{21}$$

where

$$\mathcal{E} := r((u_0, b_0), (u, b)) - r((u_\star, b_\star), (u, b)) - r((u_0, b_0), (u_\star, b_\star)).$$

By simply adding inequalities (19) and (21) together and using the definition of the function L, we get that

$$\mathcal{E} \geq L(u_\star, d_\star, b) - L(u, Bu, b) - \beta\langle B(u_\star - u), b_\star - b_0 - (Bu_0 - d_0)\rangle. \tag{22}$$

This technical inequality will be used later in the proof of the convergence result.

The next lemma shows that if a triple (u, d, b) is a fixed-point of (7), (8), (9) for some positive numbers α and $\beta > 0$, then the function $L(\cdot, \cdot, b)$ achieves its minimum value at the (u, d).

Lemma 3.3. *Let x be a vector in $\mathbb{R}^m$, B be an $n \times m$ matrix, and φ be a proper convex function on $\mathbb{R}^n$. If (u, d, b) is a solution of the fixed-point equations (7), (8), (9) for some positive numbers α, $\beta > 0$, then*

$$(u, d) = \arg\min\{L(\hat{u}, \hat{d}, b) : (\hat{u}, \hat{d}) \in \mathbb{R}^m \times \mathbb{R}^n).$$

Proof. Since $L(\hat{u}, \hat{d}, b)$ is convex with respect to $(\hat{u}, \hat{d})$, by Fermat's rule, it suffices to show that

$$0 \in \lambda\partial\|\cdot\|_1(u - x) + \beta B^\top b, \tag{23}$$

and

$$0 \in \partial\varphi(d) - \beta b. \tag{24}$$

Indeed, by (9), we have that $Bu = d$. Hence, applying (1) to equations (7) and (8) yields immediately (23) and (24), respectively. The proof is complete. □

The next lemma shows that the sequences $\mathcal{U}$, $\mathcal{D}$, $\mathcal{B}$ generated by (10) converge to a solution of the fixed-point equations (7), (8), (9) provided that certain conditions imposed on the sequences $\mathcal{U}$ and $\mathcal{D}$ are satisfied.

Lemma 3.4. *If the sequences $\mathcal{U}$, $\mathcal{D}$, $\mathcal{B}$ generated by the iterative scheme* (10) *satisfy*

(a) $\mathcal{U}$ and $\mathcal{B}$ are the bounded sequences in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively.

(b) $\lim_{k\to\infty} \|u^{k+1} - u^k\| = 0$ and $\lim_{k\to\infty} \|b^{k+1} - b^k\| = 0$,

then, there exists a subsequence of $\mathcal{U}\times\mathcal{D}\times\mathcal{B}$ that converges to a solution of the fixed-point equations (7), (8), (9).

Proof. By the boundedness of the sequences $\mathcal{U}$ and $\mathcal{B}$, there exists a subsequence $(\mathcal{U}\times\mathcal{B})' := ((u^{k_j}, b^{k_j}) : j \in \mathbb{N})$ of $\mathcal{U}\times\mathcal{B}$ that converges to a point (u, b) in $\mathbb{R}^m \times \mathbb{R}^n$. By item (b), both $(\mathcal{U}\times\mathcal{B})'_+ := ((u^{k_j+1}, b^{k_j+1}) : j \in \mathbb{N})$ and $(\mathcal{U}\times\mathcal{B})'_- := ((u^{k_j-1}, b^{k_j-1}) : j \in \mathbb{N})$ converge to (u, b) too. From the third step of the iteration (10), we know that the subsequences $\mathcal{D}'_+$ and $\mathcal{D}'$ also converge to a point $d = Bu$. Therefore, both $(\mathcal{U}\times\mathcal{B}\times\mathcal{D})'_+$ and $(\mathcal{U}\times\mathcal{B}\times\mathcal{D})'$ converge to the same point (u, d, b). It follows from the iterative scheme (10) that the triple (u, d, b) satisfies equations (7), (8), (9). □

We are now ready to prove the convergence theorem of the iterative scheme (10).

Theorem 3.5. *If x is a vector in $\mathbb{R}^m$, B an $n\times m$ matrix, φ a proper convex function on $\mathbb{R}^n$, and α, β, λ are positive numbers such that*

$$\rho := \frac{\beta}{\lambda\alpha} < \frac{1}{\|B\|^2}, \tag{25}$$

then the sequence $\mathcal{U}$ generated by the iterative scheme (10) *converges to a solution of the variational problem* (2).

Proof. By Proposition 2.2, finding a solution of model (2) amounts to finding a solution of the fixed-point equations (7), (8), and (9). It suffices to verify that the sequences $\mathcal{U}$, $\mathcal{D}$, $\mathcal{B}$, generated by the iterative scheme (10), satisfy conditions (a) and (b) in Lemma 3.4.

The equivalence between the iterative schemes (10) and (11) allows us to identify the triples (u^k, d^k, b^k) and $(u^{k+1}, d^{k+1}, b^{k+1})$ as (u_0, d_0, b_0) and $(u_\star, d_\star, b_\star)$ in (18). To proceed, we introduce the quantities

$$\mathcal{E}_k := r((u^k, b^k), (u, b)) - r((u^{k+1}, b^{k+1}), (u, b)) - r((u^{k+1}, b^{k+1}), (u^k, b^k))$$

and

$$\mathcal{I}_k := \langle B(u^{k+1} - u), b^{k+1} - b^k - (Bu^k - d^k)\rangle.$$

By (22), we have for any pair $(u, b) \in \mathbb{R}^m \times \mathbb{R}^n$ that

$$\mathcal{E}_k \geq L(u^{k+1}, d^{k+1}, b) - L(u, Bu, b) - \beta\mathcal{I}_k. \tag{26}$$

In particular, when we choose the triple (u, d, b) to be a solution of the fixed point equations (7), (8), and (9), by Lemma 3.3, we have for all $k \in \mathbb{N}$ that

$$L(u^{k+1}, d^{k+1}, b) - L(u, d, b) \geq 0.$$

Moreover, since $Bu = d$, we conclude for all $k \in \mathbb{N}$ that

$$L(u^{k+1}, d^{k+1}, b) - L(u, Bu, b) \geq 0 \tag{27}$$

and so $\mathcal{E}_k \geq -\beta\mathcal{I}_k$. Noting that $Bu^k - d^k = b^k - b^{k-1}$, for the last term of (26) we have that

$$\mathcal{I}_k = \langle B(u^{k+1} - u), b^{k+1} - b^k \rangle - \langle B(u^{k+1} - u), b^k - b^{k-1} \rangle.$$

By splitting $B(u^{k+1} - u)$ into the sum of $B(u^{k+1} - u^k)$ and $B(u^k - u)$ for the last term of the above equation and using definition (13), we get the equation

$$\mathcal{I}_k = a((u, b^k), (u^{k+1}, b^{k+1})) - a((u, b^{k-1}), (u^k, b^k)) - a((u^k, b^{k-1}), (u^{k+1}, b^k)). \tag{28}$$

Furthermore, applying Lemma 3.1 to the last term of (28) leads to

$$a((u^k, b^{k-1}), (u^{k+1}, b^k)) \geq -\frac{\tau}{\beta} r((u^k, b^{k-1}), (u^{k+1}, b^k)), \tag{29}$$

where τ is given by (15). Using (27), (28) and (29) together with (26), we obtain that

$$\mathcal{E}_k \geq -\beta(a((u, b^k), (u^{k+1}, b^{k+1})) - a((u, b^{k-1}), (u^k, b^k))) - \tau r((u^k, b^{k-1}), (u^{k+1}, b^k)). \tag{30}$$

Summing both sides of inequality (30) on k from p to $q-1$. We conclude that

$$\sum_{k=p}^{q-1} \mathcal{E}_k \geq -\beta(a((u, b^{q-1}), (u^q, b^q)) - a((u, b^{p-1}), (u^p, b^p))) - \tau \sum_{k=p}^{q-1} r((u^k, b^{k-1}), (u^{k+1}, b^k)). \tag{31}$$

Using the formula

$$\sum_{k=p}^{q-1} \mathcal{E}_k = r((u^p, b^p), (u, b)) - r((u^q, b^q), (u, b)) - \sum_{k=p}^{q-1} r((u^{k+1}, b^{k+1}), (u^k, b^k)),$$

and the inequality

$$a((u, b^{q-1}), (u^q, b^q)) \leq \frac{\tau}{\beta} r((u, b^{p-1}), (u^p, b^p)),$$

which is due to Lemma 3.1, and the fact that

$$r((u^k, b^{k-1}), (u^{k+1}, b^k)) = r((u^{k+1}, b^{k+1}), (u^k, b^k)) + \frac{\beta}{2}(\|b^k - b^{k-1}\|^2 - \|b^{k+1} - b^k\|^2),$$

inequality (31) can be further simplified to

$$r((u^p, b^p), (u, b)) \geq (1 - \tau) \left(r((u^q, b^q), (u, b)) + \sum_{k=p}^{q-1} r((u^{k+1}, b^{k+1}), (u^k, b^k)) \right) + c_p, \tag{32}$$

where

$$c_p := \beta a((u, b^{p-1}), (u^p, b^p)) - \frac{\tau\beta}{2} \|b^p - b^{p-1}\|^2.$$

In light of (25) and (15), we know that $\tau < 1$, that is, $1 - \tau$ is positive. Hence, taking $p = 0$ in (32), we conclude that the sequences $\mathcal{U}$ and $\mathcal{B}$ are bounded and that

$$\lim_{k \to \infty} \|u^{k+1} - u^k\| = 0 \quad \text{and} \quad \lim_{k \to \infty} \|b^{k+1} - b^k\| = 0.$$

By Lemma 3.4, there exists a subsequence $(\mathcal{U} \times \mathcal{D} \times \mathcal{B})'$ of $\mathcal{U} \times \mathcal{D} \times \mathcal{B}$ that converges to a solution $(\hat{u}, \hat{d}, \hat{b})$ of the fixed-point equations (7), (8), and (9). We denote by $(k_j : j \in \mathbb{N})$ the indexes corresponding to this subsequence.

Finally, we show that $\lim_{q\to\infty}(u^q, b^q) = (\hat{u}, \hat{b})$. To this end, we let $p = k_j < q$ and take $(u, b) = (\hat{u}, \hat{b})$ in (32). We then obtain that

$$r((u^{k_j}, b^{k_j}), (\hat{u}, \hat{b})) - c_{k_j} \geq (1-\tau)\left(r((u^q, b^q), (\hat{u}, \hat{b})) + \sum_{k=k_j}^{q-1} r((u^{k+1}, b^{k+1}), (u^k, b^k))\right). \tag{33}$$

Letting $k_j \to \infty$, we see that $q \to \infty$ and $\lim_{q\to\infty}(u^q, b^q) = (\hat{u}, \hat{b})$. In particular, u^q converges to a solution of model (2). □

4 Denoising Algorithms

In this section, we apply the fixed-point algorithm developed in Section 2 to the L1/TV denoising model. First, as a corollary of Theorem 3.5, we show the convergence of the resulting denoising algorithm. By exploiting the specific structure of the denoising algorithm, we propose an algorithm accelerated by using a Gauss-Seidel strategy. Finally, we compare our algorithm with two algorithms recently developed in the literature for the model and point out that unlike our algorithm, Gauss-Seidel iterations do not accelerate convergence of these algorithms.

In our image denoising model, $\varphi \circ B$ is taken to be the total-variation. For convenience of exposition, we consider a digital image as a $q \times q$ square matrix for a positive integer q and treat this image matrix as a vector in $\mathbb{R}^{q^2}$ in such a way that the ij-th entry of the image matrix corresponds to the $((i-1)q + j)$-th entry of the vector in $\mathbb{R}^{q^2}$. In this section, we shall reserve m for q^2, that is $m := q^2$. For an image $u \in \mathbb{R}^m$, the total-variation of the image has the form $\varphi(Bu)$. Here B is a $2m \times m$ matrix defined, through a $q \times q$ matrix D and the notion of matrix Kronecker product $\otimes$, as

$$B := \begin{bmatrix} I_q \otimes D \\ D \otimes I_q \end{bmatrix} \quad \text{with} \quad D := \begin{bmatrix} 0 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}. \tag{34}$$

The convex function $\varphi : \mathbb{R}^{2m} \to \mathbb{R}$ is defined at $z \in \mathbb{R}^{2m}$ as

$$\varphi(z) := \|z\|_1 \quad \text{or} \quad \varphi(z) := \sum_{i=1}^{m} \left\| [z_i, z_{m+i}]^\top \right\|, \tag{35}$$

with the corresponding $\varphi(Bu)$ being referred to as, respectively, the anisotropic or isotropic total-variation of u.

With these preparations, the following result is a direct consequence of Theorem 3.5 applied to the L1/TV denoising model.

Theorem 4.1. *If x is a square image in $\mathbb{R}^m$, B the $2m \times m$ matrix defined by* (34)*, φ the function on $\mathbb{R}^{2m}$ given by* (35)*, and α, β, λ are positive numbers such that*

$$\rho := \frac{\beta}{\lambda\alpha} < \left(8\sin^2\frac{(q-1)\pi}{2q}\right)^{-1}, \tag{36}$$

then the sequence $\mathcal{U}$ generated by the iterative scheme (10) *converges to a solution of model* (2) *where the function $\varphi \circ B$ is defined as in* (34) *and* (35)*.*

Proof. This theorem follows directly from Theorem 3.5. In fact, according to [10], for the matrix B specifically defined as in (34), we have that

$$\|B\|^2 = 8\sin^2\frac{(q-1)\pi}{2q}.$$

Hence, when inequality (36) holds, the hypotheses of Theorem 3.5 are all satisfied, and thus, the sequence $\mathcal{U}$ generated by the iterative scheme (10) converges to a solution of model (2) where the function $\varphi \circ B$ is defined as in (34) and (35). □

We now review specific formulas for the proximity operator of the function φ defined in (35). Notice that for a positive number λ, one easily establishes for $x \in \mathbb{R}$

$$\text{prox}_{\frac{1}{\lambda}|\cdot|}(x) = \max\left\{|x| - \frac{1}{\lambda}, 0\right\}\text{sign}(x) \tag{37}$$

and for $x \in \mathbb{R}^k$

$$\text{prox}_{\frac{1}{\lambda}\|\cdot\|}(x) = \text{prox}_{\frac{1}{\lambda}|\cdot|}(\|x\|)\frac{x}{\|x\|}. \tag{38}$$

By employing the definition of the proximity operator and the separability of the functions $\|\cdot\|_1$ and $\|\cdot\|^2$, for $z = (z_1, z_2, \ldots, z_{2m}) \in \mathbb{R}^{2m}$ and $\varphi(z) = \|z\|_1$, we have that

$$\text{prox}_{\frac{1}{\lambda}\varphi}(z) = \left(\text{prox}_{\frac{1}{\lambda}|\cdot|}(z_1), \text{prox}_{\frac{1}{\lambda}|\cdot|}(z_2), \ldots, \text{prox}_{\frac{1}{\lambda}|\cdot|}(z_{2m})\right).$$

Thus, $\text{prox}_{\frac{1}{\lambda}\varphi}(z)$ can be computed by combining the above equation with formula (37).

For $z = (z_1, z_2, \ldots, z_{2m}) \in \mathbb{R}^{2m}$ and $\varphi(z) = \sum_{i=1}^{m}\|(z_i, z_{m+i})\|$, the pair of the i-th and $(m+i)$-th entries of the vector $\text{prox}_{\frac{1}{\lambda}\varphi}(z)$ is the two-dimensional vector

$$\text{prox}_{\frac{1}{\lambda}\|\cdot\|}((z_i, z_{m+i})),$$

which can be evaluated by using (38).

The following algorithm describes the procedure for finding a solution of the above L1/TV image denoising model according to Theorem 4.1.

Algorithm 1 (Proposed L1/TV Denoising Algorithm)

1: Given: noisy image x in $\mathbb{R}^m$; $\lambda > 0$, $\alpha > 0$, $\beta > 0$ such that $\rho < \frac{1}{8}$
2: Initialization: $u^0 = x$, $d^0 = 0$, and $b^0 = 0$
3: **repeat**
4: Step 1: $u^{k+1} \longleftarrow x + \text{prox}_{\frac{1}{\alpha}\|\cdot\|_1}(B_\rho u^k - x - \rho B^\top(b^k - d^k))$
5: Step 2: $d^{k+1} \longleftarrow \text{prox}_{\frac{1}{\beta}\varphi}(Bu^{k+1} + b^k)$
6: Step 3: $b^{k+1} \longleftarrow b^k + Bu^{k+1} - d^{k+1}$
7: **until** "convergence"
8: Write the output of $u^{(k)}$ from the above loop as $u^\star$.

Next, we propose an acceleration for Algorithm 1 by using a Gauss-Seidel iteration strategy. To this end, let us take a closer look at Step 1 in the algorithm. By b_U^k we denote the vector formed with the first m entries of the vector b^k and b_L^k that formed with the last m entries of b^k. Likewise, d_U^k and d_L^k are defined in the same way from the vector d^k. For convenience of exposition, we shall

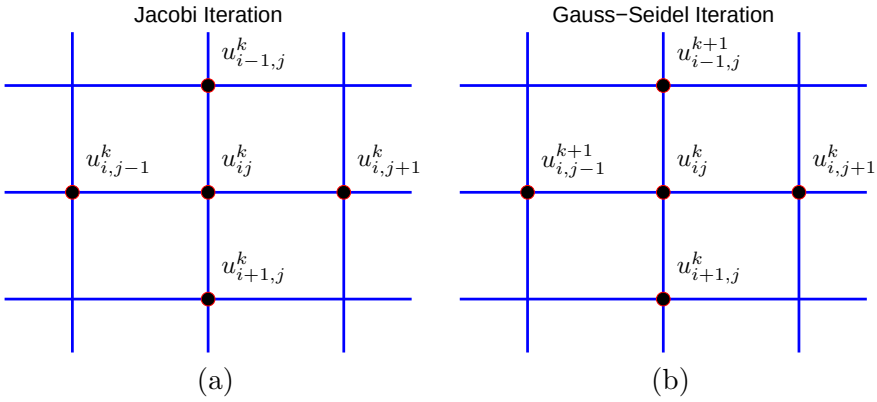


Figure 1: (a) Jabobi iteration; and (b) Gauss-Seidel iteration.

view all vectors in Algorithm 1 as matrices for the time being. The component-wise expression of Step 1 in Algorithm 1 for interior pixel at the (i,j)-th location may be written as

$$\begin{aligned} u^{k+1}_{ij} = x_{ij} + \text{prox}_{\frac{1}{\alpha}|\cdot|}((1-4\rho)u^k_{ij} + \rho(u^k_{i-1,j} + u^k_{i+1,j} + u^k_{i,j-1} + u^k_{i,j+1}) - x_{ij} \\ -\rho((b^k_U)_{ij} - (b^k_U)_{i+1,j} + (b^k_L)_{ij} - (b^k_L)_{i,j+1}) \\ +\rho((d^k_U)_{ij} + (d^k_U)_{i+1,j} - (d^k_L)_{ij} + (d^k_L)_{i,j+1})). \end{aligned} \tag{39}$$

Similar expressions can be derived for boundary pixels. In viewing (39), the pixel u^k_{ij} and its four neighboring pixel values $u^k_{i-1,j}$, $u^k_{i+1,j}$, $u^k_{i,j-1}$, and $u^k_{i,j+1}$ all contribute to the updated pixel u^{k+1}_{ij}. This is illustrated in Figure 1(a). Notice that when an order is preassigned for updating pixels at the k-th iteration to obtain u^{k+1} at the $(k+1)$-th iteration, prior to having the pixel u^{k+1}_{ij}, some of pixels $u^k_{i-1,j}$, $u^k_{i+1,j}$, $u^k_{i,j-1}$, and $u^k_{i,j+1}$ have already been updated. Therefore, following the idea of the Gauss-Seidel iteration these updated pixels can be used to update the pixel u^{k+1}_{ij}. A particular scenario is illustrated in Figure 1(b). In this way, we can expect that the resulting scheme be able to accelerate the convergence of the resulting algorithm. For example, if the components of u^{k+1} are updated from its top to bottom and column by column, then, instead of using (39), the pixel u^{k+1}_{ij} can be computed by the following formula:

$$\begin{aligned} u^{k+1}_{ij} = x_{ij} + \text{prox}_{\frac{1}{\alpha}|\cdot|}((1-4\rho)u^k_{ij} + \rho(u^{k+1}_{i-1,j} + u^k_{i+1,j} + u^{k+1}_{i,j-1} + u^k_{i,j+1}) - x_{ij} \\ -\rho((b^k_U)_{ij} - (b^k_U)_{i+1,j} + (b^k_L)_{i,j} - (b^k_L)_{i,j+1}) \\ +\rho((d^k_U)_{ij} + (d^k_U)_{i+1,j} - (d^k_L)_{ij} + (d^k_L)_{i,j+1})). \end{aligned} \tag{40}$$

With this modification in Step 1 of Algorithm 1, we have Algorithm 2 below accelerated by a Gauss-Seidel strategy.

Next, we compare Algorithm 1 to two most related algorithms for the L1/TV model, in the context of acceleration by using Gauss-Seidel iterations. The first one is the MSXZ which was originally introduced in [11] and has been reviewed in Section 2. This algorithm begins with an initial pair $(u^0, b^0) \in \mathbb{R}^m \times \mathbb{R}^{2m}$ and then iteratively generates two sequences $\mathcal{U}$ and $\mathcal{B}$ via the iterative scheme

$$\begin{cases} u^{k+1} = x + \text{prox}_{\frac{1}{\alpha}\|\cdot\|_1}(u^k - x - \rho B^t b^k), \\ b^{k+1} = (\mathrm{I} - \text{prox}_{\frac{1}{\beta}\varphi})(Bu^{k+1} + b^k). \end{cases} \tag{41}$$

The iteration scheme (41) shows that generating u^{k+1}_{ij} at the $(k+1)$-th iteration involves only the pixel u^k_{ij} at the k-th iteration. Hence, the iterative scheme using Jacobi iterations for the first

Algorithm 2 (Proposed Algorithm Accelerated by Gauss-Seidel Iterations)

1: Given: noisy image x in $\mathbb{R}^m$; $\lambda > 0$, $\alpha > 0$, $\beta > 0$ such that $\rho < \frac{1}{8}$
2: Initialization: $u^{(0)} = x$, $d^{(0)} = 0$ and $b^{(0)} = 0$
3: **repeat**
4: Step 1: update the components of u^{k+1} according to (40)
5: Step 2: $d^{k+1} \longleftarrow \text{prox}_{\frac{1}{\beta}\varphi}(Bu^{k+1} + b^k)$
6: Step 3: $b^{k+1} \longleftarrow b^k + Bu^{k+1} - d^{k+1}$
7: **until** "convergence"
8: Write the output of $u^{(k)}$ from the above loop as $u^\star$.

equation of (41) is identical to that using Gauss-Seidel iterations. In other words, the component-wise Gauss-Seidel iteration applied to (41) will not result in acceleration of convergence.

The second algorithm to compare with is the one proposed in [4], which will be referred to as the CP. For any initial guess $(u^{-1}, u^0, b^0) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^{2m}$, this algorithm generates the sequences $\mathcal{U}$ and $\mathcal{B}$ via the following scheme

$$\begin{cases} b^{k+1} = (I - \text{prox}_{\frac{1}{\beta}\varphi})(B\bar{u}^k + b^k), \\ u^{k+1} = x + \text{prox}_{\frac{1}{\alpha}\|\cdot\|_1}(u^k - x - \rho B^\top b^{k+1}), \\ \bar{u}^{k+1} = 2u^k - u^{k-1}. \end{cases} \tag{42}$$

This algorithm is similar to the MSXZ in the way that the component-wise Gauss-Seidel iteration applied to it will not result in acceleration of convergence.

Finally, it is interesting to point it out that the CP (that is, equation (42)) may be derived from our original characterization result (Proposition 2.1). By letting $\bar{u} = u$ in the right hand side of equation (4), changing the order of these two equations and applying the block Gauss-Seidel iteration to the resulting fixed-point equations with the third iteration equation $\bar{u}^{k+1} = 2u^k - u^{k-1}$ based on equation $\bar{u} = u$, we obtain the CP.

5 Numerical Experiments

In this section, we demonstrate the computational performance of Algorithm 2 for the L1/TV denoising model in the setting of impulsive noise removal. We shall mainly compare numerical results of the proposed algorithm to those of the MSXZ and the CP which were reviewed in the previous section.

All the experiments are performed under Windows 7 and MATLAB R2008a running on a PC equipped with an Intel Core 2 Quad CPU at 2.12 GHz and 2G RAM memory. All algorithms tested in the experiments are carried out until the stopping condition

$$\|u^{k+1} - u^k\|^2 / \|u^k\|^2 < 10^{-3}$$

is satisfied. The quality of a denoised image $u^\star$ is evaluated in terms of the peak signal-to-noise ratio (PSNR) defined by

$$\text{PSNR} = 10 \log_{10} \frac{255^2 m}{\|u^\star - u\|^2} (\text{dB}),$$

where u is the original image with a total of m pixels.

Two experiments are conducted for different purposes. In the first experiment, we illustrate that the L1/TV model is more suitable for cartoon-like images in the context of impulsive noise removal than smoothed L1/TV models, as commented in Section 1. For a comparison, the smoothed L1/TV model in [16] is used in this experiment. To introduce this smoothed L1/TV model, let us review the notion of the Moreau envelope. Let $f : \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous convex function. The Moreau envelope of the function f at $v \in \mathbb{R}^p$ is defined through the formula

$$\mathrm{env}_f(v) := \min\left\{\frac{1}{2}\|y - v\|^2 + f(y) : y \in \mathbb{R}^p\right\}.$$

Then the smoothed L1/TV model in [16] can be restated in terms of Moreau envelope as

$$\min\left\{\lambda \mathrm{env}_{\alpha\|\cdot\|_1}(u - x) + \mathrm{env}_{\beta\varphi}(Bu) : u \in \mathbb{R}^m\right\}, \tag{43}$$

where α and β are two positive parameters. Here we refer the algorithm developed in [16] for the above smoothed L1/TV model to as "FTVd".

Our experiment is conducted with a test image of size 256×256 with a square in its center, shown in Figure 2(a). The intensity of this image at the ij-th locations is 200 if $65 \le i, j \le 192$; 50 otherwise. Row 65 (the boundary of the square) and Row 128 (the line passing through the center of the square) of the image are shown in Figure 2(b) and (c), respectively. We add the salt-pepper noise of level 40% to the image. The resulting noisy image and its row 65 and row 128 are displayed in Figure 2(d), (e), and (f), respectively. We apply the FTVd and Algorithm 2 to the noisy image. The denoised image with the FTVd has a PSNR value 36.37 dB while the denoised image with Algorithm 2 has a PSNR value 37.56 dB. Both denoised images are shown in the left column of Figure 3. It can be observed that the visual quality of the image restored with model (2) via Algorithm 2 is better than the one with model (43) via the FTVd, especially around the edges of the square. This shows that the use of the non-differentiable total variation in the L1-TV model helps us preserve sharp edges in the restored images.

The second experiment is designed to compare the performance of Algorithm 2 with that of the MSXZ and the CP for impulsive noise removal. Three test images displaying in Figure 4 are the images of "Cameraman" and "Lena" of size 256×256 and the image "Plane" of size 320×400. We explore how the algorithms perform when the level of impulsive noise added to the test images changes.

The numerical results obtained from these algorithms for this experiment with noise level 10%, 20%, 30%, 40%, and 50% are reported in Table 1. We find that the PSNR-values of the restored images by Algorithm 2 are higher for most cases than those of the MSXZ and the CP. Table 1 also shows that Algorithm 2 uses less CPU-time than the CP or the MSXZ does. The restored images obtained from these algorithms for the noise level at 30% are displayed in Figure 5. As it can be seen, the denoised images by these methods have similar visual quality as they are all based on the same L1/TV model which ultimately determines the quality of restoration. Furthermore, for the image of "Cameraman" corrupted by impulsive noise at level 30%, the changes of the PSNR values of u^k and the objective function values at u^k as k increases are illustrated in Figure 6 for the CP and Algorithm 2. It can be seen from the plots that Algorithm 2 converges significantly faster than the CP.

6 Conclusions

We develop a proximity algorithm for solutions of the L1/TV image denoising model, which can be accelerated by using the component-wise Gauss-Seidel iteration. The algorithm is based on

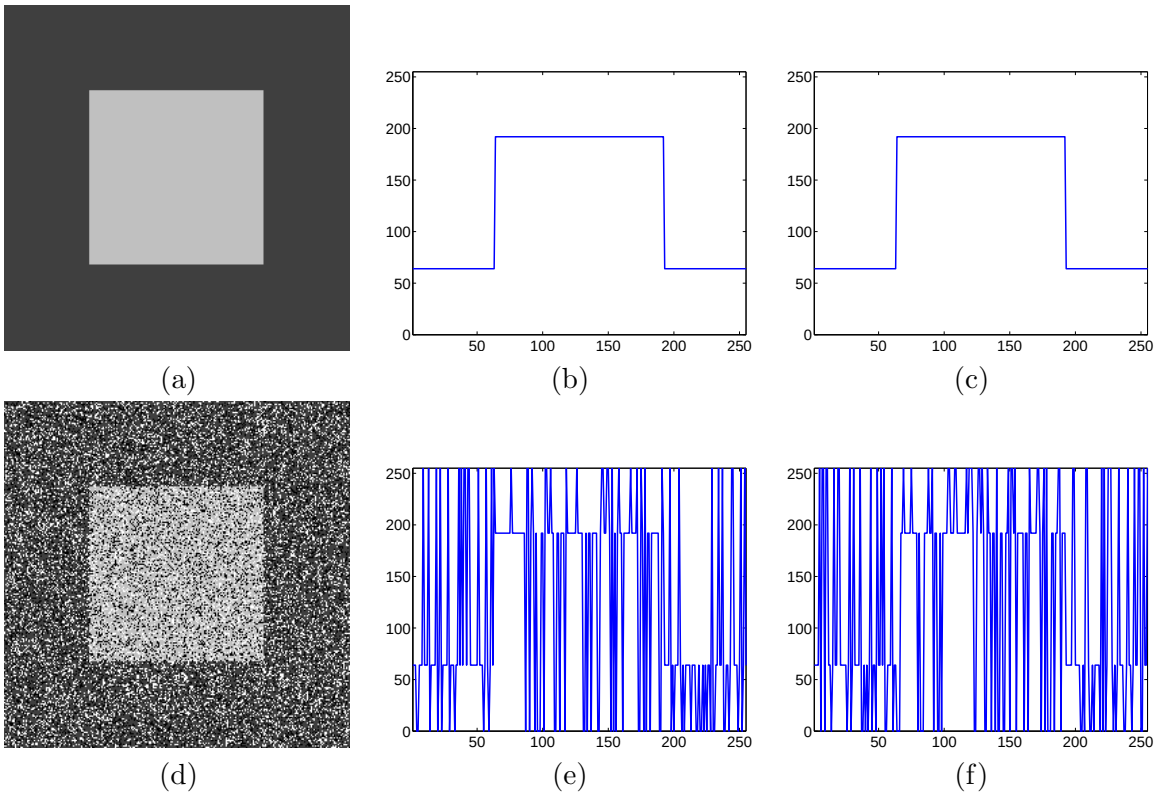


Figure 2: First row (from left to right): The original geometric test image and its Row 65 and Row 128. Second row (from left to right): The image corrupted with 40% salt-pepper noise and its profiles at row 65 and row 128.

Table 1: The summary of the restoration results of the MSXZ, the CP, and Algorithm 2.

		Cameraman		Lena		Airplane	
Level	Method	PSNR (dB)	Time (seconds)	PSNR (dB)	Time (seconds)	PSNR (dB)	Time (seconds)
10%	MSXZ	29.49	2.09	31.31	2.22	38.40	4.21
	CP	29.06	1.97	30.88	1.79	38.25	1.56
	Algorithm 2	29.39	1.06	31.34	0.81	38.56	0.88
20%	MSXZ	26.60	2.41	28.83	2.44	36.26	4.86
	CP	26.40	2.07	28.63	1.70	36.20	1.65
	Algorithm 2	26.58	1.08	28.82	0.87	36.25	1.06
30%	MSXZ	24.94	2.57	27.25	2.64	34.38	5.31
	CP	24.77	2.04	27.11	1.77	34.31	2.06
	Algorithm 2	24.89	1.13	27.22	1.01	34.33	1.26
40%	MSXZ	23.44	2.88	25.82	2.91	32.33	5.61
	CP	23.33	2.18	25.74	1.91	32.35	2.44
	Algorithm 2	23.37	1.38	25.78	1.18	32.42	1.60
50%	MSXZ	21.51	2.96	24.19	2.94	28.89	5.72
	CP	21.46	2.36	24.15	2.06	29.37	3.55
	Algorithm 2	21.51	1.61	24.16	1.36	29.59	2.50

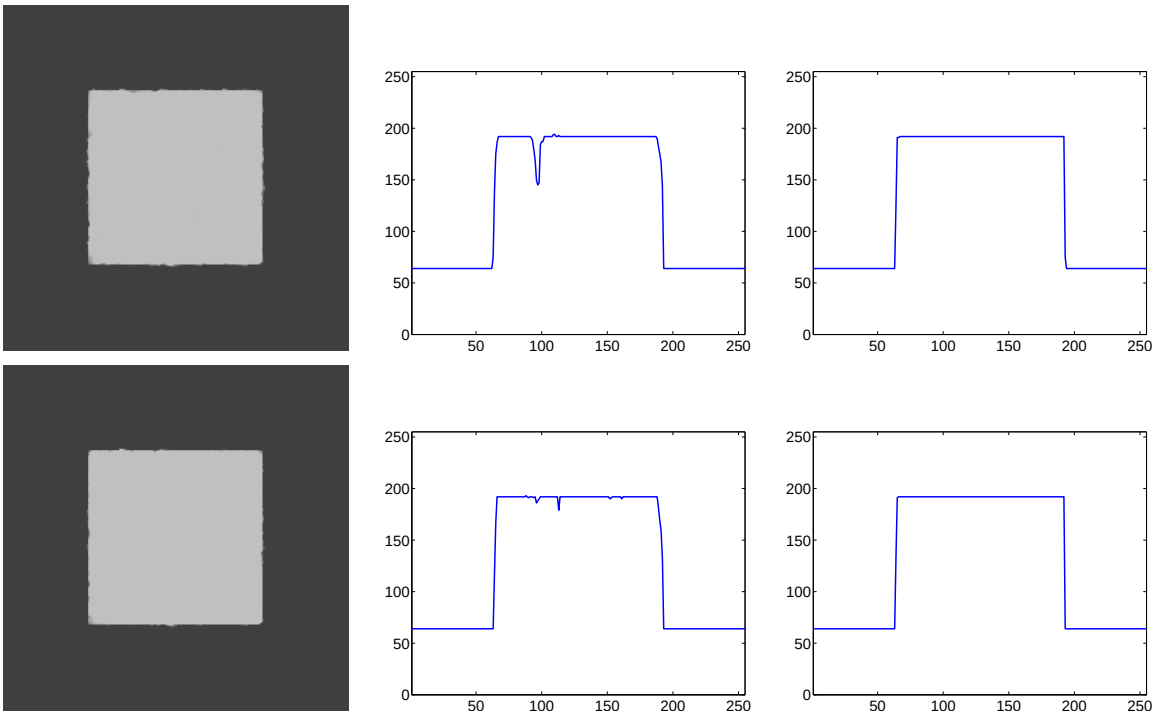
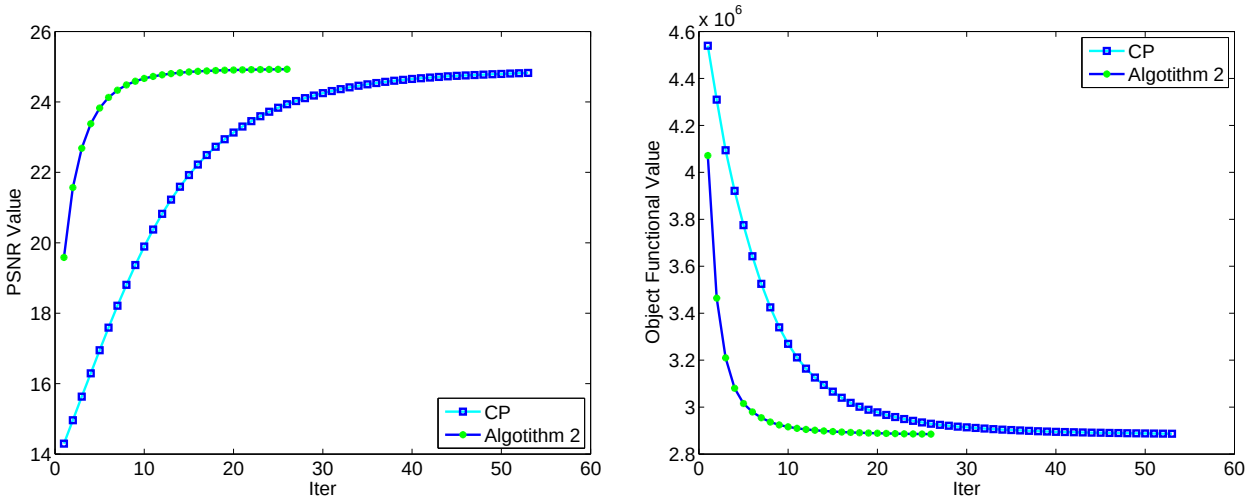


Figure 3: Restored images and its cross sections at row 65 and row 128 with the smoothed L1/TV model (43) (Row 1) and the L1/TV model (2) (Row 2).

(a) (b) (c)

Figure 4: Original images. (a) "Cameraman"; (b) "Lena"; and (c) "Airplane".

Figure 5: (Row 1) The noisy images. The restored images by the MSXZ (Row 2); the CP (Row 3); and Algorithm 2 (Row 4).



(a) (b)

Figure 6: The plots of (a) PSNR values versus the number of iterations and (b) the values of the objective function of the L1/TV model versus the number of iterations for the image of "Cameraman" corrupted by 30% impulsive noise.

an alternative characterization of solutions of the L1/TV model. A convergence result for this algorithm is established. A comparison of the proposed Gauss-Seidel iterative algorithm with two known algorithms shows that our algorithm outperforms them in terms of the convergence speed.

References

[1] S. Alliney, *A property of the minimum vectors of a regularizing functional defined by means of the absolute norm*, IEEE Transactions on Signal Processing, 45 (1997), pp. 913–917.

[2] J. E. Aujol, G. Gilboa, T. Chan, and S. Osher, *Structure-texture image decomposition-modeling, algorithms, and parameter selection*, International Journal of Computer Vision, 67 (2006), pp. 111–136.

[3] H. L. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, AMS Books in Mathematics, Springer, New York, 2011.

[4] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision, 40 (2011), pp. 120–145.

[5] T. Chan and S. Esedoglu, *Aspects of total variation regularized l^1 function approximation*, SIAM Journal on Applied Mathematics, 65 (2005), pp. 1817–1837.

[6] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. Huang, *Total variation models for variable lighting face recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 28 (2006), pp. 1519–1524.

[7] C. Clason, B. Jin, and K. Kunisch, *A duality-based splitting method for L1-TV image restoration with automatic regularization parameter choice*, SIAM Journal on Scientific Computing, 32 (2010), pp. 1484–1505.

[8] Y. Dong, M. Hintermuller, and M. Neri, *A primal-dual method for l^1-TV image denoising*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 1168–1189.

[9] D. Donoho and I. Johnstone, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika, 81 (1994), pp. 425–455.

[10] C. A. Micchelli, L. Shen, and Y. Xu, *Proximity algorithms for image models: Denoising*, Inverse Problems, 27 (2011), p. 045009(30pp).

[11] C. A. Micchelli, L. Shen, Y. Xu, and X. Zeng, *Proximity algorithms for image models II: L1/TV denosing*, Advances in Computational Mathematics, online version available, (2011).

[12] J.-J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.

[13] M. Nikolova, *A variational approach to remove outliers and impulse noise*, Journal of Mathematical Imaging and Vision, 20 (2004), pp. 99–120.

[14] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.

[15] C. Wu, J. Zhang, and X.-C. Tai, *Augmented lagrangian method for total variation restoration with non-quadratic fidelity*, Inverse Problems and Imaging, 5 (2011), pp. 237–261.

[16] J. Yang, Y. Zhang, and W. Yin, *An efficient TVL1 algorithm for deblurring multichannel images corrupted by impulsive noise*, SIAM Journal on Scientific Computing, 31 (2009), pp. 2842–2865.

[17] W. Yin, T. Chen, X. S. Zhou, and A. Chakraborty, *Background correction for cDNA microarray image using the TV+L1 model*, Bioinformatics, 10 (2005), pp. 2410–2416.

[18] W. Yin, D. Goldfard, and S. Osher, *The total variation regularized l^1 model for multiscale decomposition*, Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal, 6 (2007), pp. 190–211.

[19] C. Zach, T. Pock, and H. Bischof, *A duality based approach for realtime TV-l^1 optical flow*, in Lecture Notes in Computer Sciences, vol. 4713, Berlin Heidelberg, 2007, Springer-Verlag, pp. 214–223.