Institute for Computational Mathematics
Hong Kong Baptist University

ICM Research Report
08-13

A FAST ℓ1-TV ALGORITHM FOR IMAGE RESTORATION

XIAOXIA GUO*, FANG LI†, AND MICHAEL K. NG ‡

Abstract. Image restoration problems are often solved by finding the minimizer of a suitable objective function consisting of a data-fitting term and a regularization term. In this paper, we consider the data-fitting term measured in the ℓ1 norm to handle non-Gaussian additive noise, and the regularization term given by the total variation to restore image edges. We propose a new objective function for this image restoration problem by making use of new variables to modify the data-fitting term and the total variation regularization term. An alternating minimization algorithm based on the new objective function is employed to restore blurred and noisy images. Our experimental results show that the quality of restored images by the proposed method is competitive with those restored by the other tested methods. We also show the convergence of the alternating minimization algorithm and demonstrate that the proposed algorithm is very efficient.

Keywords: Iterative algorithm, image restoration, deblurring, denoising, l1-norm, total variation

1. Introduction. The problem of image restoration is considered. The observed image is the convolution of a shift invariant blurring function with the true image plus some additive noise. Let $f(x, y)$ be the original scene, $g(x, y)$ the observed scene, and $h(x, y)$ the blurring function. The image formation process can be modeled as follows:

$$g(x, y) = h(x, y) \star f(x, y) + n(x, y).$$

Here $n(x, y)$ is the additive noise and $\star$ denotes two-dimensional convolution. Let f, g, and n be the discretized original scene, observed scene, and additive noise, respectively. Let H be the corresponding blurring matrix of appropriate size built according the discretized point spread h. Then the discretized image formation process can be put into matrix-vector form:

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{n} \tag{1.1}$$

Assuming that the discretized scenes have $N \times N$ pixels, then $\boldsymbol{f}, \boldsymbol{g}$, and $\boldsymbol{n}$ are vectors of length N^2, and $\boldsymbol{H}$ is a matrix of $N^2 \times N^2$. We remark that H is a matrix of block Toeplitz with Toeplitz blocks (BTTB) when zero boundary conditions are applied, and block Toeplitz-plus-Hankel with Toeplitz-plus-Hankel blocks (BTHTHB) when Neumann boundary conditions are used [24].

Image restoration problems tend to be very ill-conditioned. Directly solving (1.1) will yield a solution that is extremely sensitive to noise; therefore, regularization methods are needed to stabilize the solution. For example, the linear least squares problem with Tikhonov's regularization [31] takes the following form:

$$\min_{\boldsymbol{f}} \left\{ \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|_2^2 + \alpha \|\boldsymbol{R}\boldsymbol{f}\|_2^2 \right\}. \tag{1.2}$$

In this optimization problem, the second term is a regularization term that measures the "irregularity" of the solution. we call $\boldsymbol{R}$ the regularization operator and α the

*College of Mathematical sciences, Ocean University of China, Qingdao, and Centre for Mathematical Imaging and Vision, Hong Kong Baptist University. Research supported in part by NNSF 10601050.

†Department of Mathematics, East China Normal University, Shanghai, P. R. China

‡The Corresponding Author. Centre for Mathematical Imaging and Vision, and Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong. E-mail: mng@math.hkbu.edu.hk. Research supported in part by RGC 201508 and HKBU FRGs.

regularization parameter. Very often $\boldsymbol{R}$ is chosen to be the difference operator, for example, the first order or second order finite difference operator. In [24], Ng *et al.* have designed a fast algorithm for solving (1.2) based on fast cosine transforms. We note in this case that only a linear system is required to be solved. However, when a total variation (TV) regularization [30] is employed or when $\ell 1$ norm is used in the data-fitting term, this fast cosine transform based algorithm cannot be applied directly. For total variation minimization methods for image recovery, we refer to [30, 10]. For theory and computation of variational image deblurring, we refer to [14]. In [16], Daubechies and Teschke presented a wavelet-based treatment of variational functionals that induce a decomposition of images into oscillation and cartoon components and possibly an appropriate noise component. Their approach can incorporate blur operators into the formulation so that the minimization leads to a simultaneous decomposition, deblurring and denoising. In [11], Chan *et al.* also studied wavelet deblurring algorithms for a special image restoration problem where a high-resolution image can be reconstructed from a set of low-resolution images.

In this paper, we consider that the data-fitting term between $\boldsymbol{Hf}$ and $\boldsymbol{g}$ is measured in the $\ell 1$ norm that can handle non-Guassian-type noises and the regularization term is given by the total variation that can restore edges in the restoration process, i.e., we are interested in the following image restoration model:

$$\min_{\boldsymbol{f}} \{\lambda \|\boldsymbol{g} - \boldsymbol{Hf}\|_1 + \|\boldsymbol{f}\|_{TV}\} \tag{1.3}$$

Here $\|\cdot\|_{TV}$ is the discrete TV regularization term. The discrete gradient operator $\nabla : \mathbb{R}^{N^2} \to \mathbb{R}^{N^2}$ is defined by

$$(\nabla \boldsymbol{u})_{j,k} = ((\nabla \boldsymbol{u})^x_{j,k}, (\nabla \boldsymbol{u})^y_{j,k})$$

with

$$(\nabla \boldsymbol{u})^x_{j,k} = \begin{cases} \boldsymbol{u}_{j+1,k} - \boldsymbol{u}_{j,k} & \text{if } j < n, \\ 0 & \text{if } j = n, \end{cases} \qquad (\nabla \boldsymbol{f})^y_{j,k} = \begin{cases} \boldsymbol{u}_{j,k+1} - \boldsymbol{u}_{j,k} & \text{if } k < n, \\ 0 & \text{if } k = N \end{cases} \tag{1.4}$$

for $j, k = 1, \ldots, N$. Here $\boldsymbol{u}_{j,k}$ refers to the $(jN + k)$th entry of the vector $\boldsymbol{u}$ (it is the (j, k)th pixel location of the image). The discrete total variation of $\boldsymbol{u}$ is defined by

$$\|\boldsymbol{u}\|_{TV} := \sum_{1 \le j,k \le N} |(\nabla \boldsymbol{u})_{j,k}|_2 = \sum_{1 \le j,k \le N} \sqrt{|(\nabla \boldsymbol{u})^x_{j,k}|^2 + |(\nabla \boldsymbol{u})^y_{j,k}|^2}.$$

Here $|\cdot|_2$ is the Euclidean norm in $\mathbb{R}^2$.

1.1. Image Restoration using $\ell 1$ Norm. Due to the presence of edges, the prior distribution of an image rarely satisfies the Gaussian assumption well. In many cases, the additive noise does not satisfy the Gaussian assumption either, for instance, the noise may follow a Laplace distribution [2]. In the literature, there has been a growing interest in using $\ell 1$ norm for image restoration [1, 13, 17, 18, 20, 21, 26, 27, 28, 35].

1.1.1. Total Variation Denoising Models. In [3], Aujol *et al.* studied TV-$\ell 1$ denoising model:

$$\min_{\boldsymbol{u}} \{\lambda \|\boldsymbol{g} - \boldsymbol{u}\|_1 + \|\boldsymbol{u}\|_{TV}\} \tag{1.5}$$

Instead of (1.5), Aujol *et al.* proposed and considered the following functional:

$$\min_{\boldsymbol{u},\boldsymbol{v}} \left\{ \frac{1}{2\alpha}\|\boldsymbol{g}-\boldsymbol{u}-\boldsymbol{v}\|_2^2 + \lambda\|\boldsymbol{v}\|_1 + \|\boldsymbol{u}\|_{TV} \right\} \tag{1.6}$$

They aim at splitting an image $\boldsymbol{g}$ into two components $\boldsymbol{u}$ and $\boldsymbol{v}$. The parameter α is set to be small so that $\boldsymbol{f} = \boldsymbol{u} + \boldsymbol{v}$ can be obtained. Aujol *et al.* proposed a fast algorithm to solve (1.6). Their idea is to fix $\boldsymbol{v}$ and search for $\boldsymbol{u}$ as a solution of

$$\min_{\boldsymbol{u}} \left\{ \|\boldsymbol{u}\|_{TV} + \frac{1}{2\alpha}\|\boldsymbol{g}-\boldsymbol{u}-\boldsymbol{v}\|_2^2 \right\}$$

and then fix $\boldsymbol{u}$ and search for $\boldsymbol{v}$ as a solution of

$$\min_{\boldsymbol{u}} \left\{ \frac{1}{2\alpha}\|\boldsymbol{g}-\boldsymbol{u}-\lambda\boldsymbol{v}\|_1 \right\}$$

They showed that iterating these two minimizations is a way to compute the solution of problem (1.6). This kind of splitting algorithm is shown to be very efficient.

1.1.2. The FTVd Algorithm. In [33], Yang *et al.* extended the alternating minimization algorithm recently proposed in [34, 32] to the case of recovering blurred images corrupted by impulsive rather than Gaussian noise. They perform deblurring and denoising jointly by solving a TV regularization problem with a $\ell 1$-norm data fidelity term and derive the algorithm by applying the well-known quadratic penalty function technique. They study the following approximate problem to (1.5):

$$\min_{\boldsymbol{w},\boldsymbol{z},\boldsymbol{u}} \sum_{i=1}^{N^2} \left(\alpha_i\|\boldsymbol{w}\|_2 + \frac{\beta}{2}\|\boldsymbol{w}_i - \boldsymbol{G}_i\boldsymbol{u}\|_2^2 \right) + \mu\left(\|\boldsymbol{z}\|_1 + \frac{\gamma}{2}\|\boldsymbol{z} - (\boldsymbol{H}\boldsymbol{u}-\boldsymbol{g})\|_2^2 \right), \tag{1.7}$$

where β, $\gamma \gg 0$ are penalty parameters. Here $\boldsymbol{w}_i$ are auxiliary variables that approximate $\boldsymbol{H}\boldsymbol{u}-\boldsymbol{g}$ and $\boldsymbol{G}_i\boldsymbol{u}$ in the nondifferentiable norms in the TV regularization. They introduce (1.7) because it is numerically easier to minimize by an iterative and alternating approach due to fact that with any two of the three variables $\boldsymbol{w}$, $\boldsymbol{z}$ and $\boldsymbol{u}$ fixed, the minimizer of (1.7) with respect to the third one has a closed-form formula that is easy to compute. This approach is also numerical stable for large values of β and γ. Since $\boldsymbol{w}$ and $\boldsymbol{z}$ are decoupled for given $\boldsymbol{u}$, their algorithm minimizes the objective function in (1.7) with respect to $(\boldsymbol{w},\boldsymbol{z})$ and $\boldsymbol{u}$, alternatively. They show that for any fixed $\beta,\gamma > 0$, their proposed alternating minimization scheme generates a sequence of points converging to a solution of (1.7). Their numerical results on images with different blurs and impulsive noise demonstrate the efficiency of the algorithm.

1.1.3. The LAD Method. In [18], Fu *et al.* considered the following minimization problem for image restoration:

$$\min_{\boldsymbol{f}} \|\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f}\|_1 + \alpha\|\boldsymbol{R}\boldsymbol{f}\|_1. \tag{1.8}$$

The solution to (1.8) is called the least absolute deviation (LAD) solution. The LAD solution is formulated as the solution to a linear programming problem which is solved by interior point methods. At each iteration of the interior point method, a structured linear system must be solved. The preconditioned conjugate gradient method with factorized sparse inverse preconditioners is employed to solve such structured

inner systems. The advantage of using this approach is that the constraint on non-negativity of the image solution can be included without extra cost. Experimental results are presented to demonstrate the effectiveness of their approach. They also demonstrate the quality of the restored images using the minimization of $\ell 1$-$\ell 1$ norms is better than that using only $\ell 2$ norm in the data-fitting term. In their paper, Fu *et al.* employed the $\ell 1$ norm of $\boldsymbol{Rf}$ as the regularization term instead the total variation (TV) regularization. Also the interior point method and the preconditioned conjugate gradient method still takes a lot of computational time in order to solve the corresponding minimization problem.

Recently, Krishnan *et al.* [19] also studied an optimization problem with non-negativity constraint arising from total variation deblurring problems. They employed a primal-dual active-set method for the non-negativity constrained problem. The image restoration results were presented to illustrate the effectiveness of their approach. However, the data-fitting term in their model is $\ell 2$ norm only.

1.2. Outline. In this paper, the main aim is to develop and study a fast $\ell 1$-TV minimization method for image restoration. The proposed unconstrained $\ell 1$-TV deblurring problem is given by

$$\min_{\boldsymbol{w},\boldsymbol{f},\boldsymbol{u}} \mathcal{J}(\boldsymbol{w},\boldsymbol{f},\boldsymbol{u}) = \min_{\boldsymbol{w},\boldsymbol{f},\boldsymbol{u}} \|\boldsymbol{Hf}-\boldsymbol{w}\|_2^2 + \alpha_1\|\boldsymbol{f}-\boldsymbol{u}\|_2^2 + \alpha_2\|\boldsymbol{w}-\boldsymbol{g}\|_1 + \alpha_3\|\boldsymbol{u}\|_{TV} \tag{1.9}$$

where $\alpha_1, \alpha_2, \alpha_3$ are positive regularization parameters. The proposed model is different from (1.3). It is clear that when α_1 and α_2 are sufficiently large, the new model in (1.9) is close to the model in (1.3), i.e., we force $\boldsymbol{f} \approx \boldsymbol{u}$ and $\boldsymbol{w} \approx \boldsymbol{g}$. The minimizer of the model in (1.9) is also close to the minimizer of the model in (1.3), see for instance [33].

We note that (1.9) can be rewritten as follows:

$$\begin{aligned}\min_{\boldsymbol{w},\boldsymbol{f},\boldsymbol{u}} & \mathcal{J}(\boldsymbol{w},\boldsymbol{f},\boldsymbol{u}) \\ = \min_{\boldsymbol{u}}\{\min_{\boldsymbol{w}}\{\min_{\boldsymbol{f}}\{\|\boldsymbol{Hf}-\boldsymbol{w}\|_2^2 + \alpha_1\|\boldsymbol{f}-\boldsymbol{u}\|_2^2\} + \alpha_2\|\boldsymbol{w}-\boldsymbol{g}\|_1\} + \alpha_3\|\boldsymbol{u}\|_{TV}\}.\end{aligned} \tag{1.10}$$

According to (1.10), we can interpret the total variation minimization scheme to denoise the deblurred image $\boldsymbol{f}$. Here α_1 measures the trade off between a deblurred image $\boldsymbol{f}$ and a denoised image $\boldsymbol{u}$. On the other hand, α_2 measures the amount of noise removal on the observed image $\boldsymbol{g}$, and α_3 measures the amount of regularization to a denoising image $\boldsymbol{u}$. The main advantage of the proposed method is that a TV norm is used in the image restoration process. Therefore the new method has the ability to preserve edges very well in the restored image.

The proposed model in (1.9) is different from the model in (1.7). In (1.9), we consider the norm of the difference between the deblurred image and the denoised image in the model. However, in (1.7), Wang *et al.* consider the norm of the difference between the gradient vector of the restored image and the new auxiliary variables in the TV calculation. In Section 4, we will compare the results of the proposed model and the model in (1.7).

An alternating minimization algorithm is employed to solve the proposed total variation minimization problem. Our experimental results show that the quality of restored images by the proposed method is competitive with those restored by the other tested methods such as the FTVd method and the LAD method. We also show that convergence of the alternating minimization algorithm and demonstrate that the algorithm is very efficient.

The outline of this paper is as follows. In Section 2, an iterative algorithm is developed. In Section 3, We show the convergence of the iterative algorithm. In Section 4, numerical examples are given to demonstrate the effectiveness of the proposed model. Conclusions are made in Section 5.

2. The Iterative Algorithm. In the paper, we propose to use an alternating minimization algorithm to solve (1.9). Starting from an initial guess $\boldsymbol{w}^{(0)}, \boldsymbol{u}^{(0)}$, this method computes a sequence of iterates

$$\boldsymbol{f}^{(1)}, \boldsymbol{w}^{(1)}, \boldsymbol{u}^{(1)}, \boldsymbol{f}^{(2)}, \boldsymbol{w}^{(2)}, \boldsymbol{u}^{(2)}, \ldots, \boldsymbol{f}^{(i)}, \boldsymbol{w}^{(i)}, \boldsymbol{u}^{(i)}, \ldots$$

such that

$$\begin{cases} \mathcal{S}_h(\boldsymbol{w}^{(i-1)}, \boldsymbol{u}^{(i-1)}) := \boldsymbol{f}^{(i)} = \operatorname{argmin}_{\boldsymbol{f}} \ \|\boldsymbol{H}\boldsymbol{f} - \boldsymbol{w}^{(i-1)}\|_2^2 + \alpha_1 \|\boldsymbol{f} - \boldsymbol{u}^{(i-1)}\|_2^2 \\ \mathcal{S}_{l_1}(\boldsymbol{f}^{(i)}) := \boldsymbol{w}^{(i)} = \operatorname{argmin}_{\boldsymbol{w}} \ \|\boldsymbol{H}\boldsymbol{f}^{(i)} - \boldsymbol{w}\|_2^2 + \alpha_2 \|\boldsymbol{w} - \boldsymbol{g}\|_1 \\ \mathcal{S}_{tv}(\boldsymbol{f}^{(i)}) := \boldsymbol{u}^{(i)} = \operatorname{argmin}_{\boldsymbol{u}} \ \alpha_1 \|\boldsymbol{f}^{(i)} - \boldsymbol{u}\|_2^2 + \alpha_3 \|\boldsymbol{u}\|_{TV} \end{cases}$$

for $i = 1, 2, \cdots$. Therefore, we can express the following relationship:

$$\begin{aligned} \boldsymbol{w}^{(i)} &= \mathcal{S}_{l_1}(\boldsymbol{f}^{(i)}) = \mathcal{S}_{l_1}(\mathcal{S}_h(\boldsymbol{w}^{(i-1)}, \boldsymbol{u}^{(i-1)})), \quad i = 1, 2, \cdots, \\ \boldsymbol{u}^{(i)} &= \mathcal{S}_{tv}(\boldsymbol{f}^{(i)}) = \mathcal{S}_{tv}(\mathcal{S}_h(\boldsymbol{w}^{(i-1)}, \boldsymbol{u}^{(i-1)})), \quad i = 1, 2, \cdots. \end{aligned}$$

For simplicity, we denote

$$\boldsymbol{w}^{(i)} = \mathcal{T}_1(\boldsymbol{w}^{(i-1)}, \boldsymbol{u}^{(i-1)}) \quad \text{and} \quad \boldsymbol{u}^{(i)} = \mathcal{T}_2(\boldsymbol{w}^{(i-1)}, \boldsymbol{u}^{(i-1)}), \tag{2.1}$$

where

$$\mathcal{T}_1(\cdot) = \mathcal{S}_{l_1}(\mathcal{S}_h(\cdot, \boldsymbol{u})) \quad \text{and} \quad \mathcal{T}_2(\cdot) = \mathcal{S}_{tv}(\mathcal{S}_h(\boldsymbol{w}, \cdot)).$$

In the next section, we will analyze the convergence of $\boldsymbol{w}^{(i)}$ and $\boldsymbol{u}^{(i)}$ under $\mathcal{T}_1$ and $\mathcal{T}_2$. Let us first study the computational cost of the alternating minimization algorithm.

Step 1. The first step of the method is to perform the deblurring. The minimizer of the optimization problem:

$$\min_{\boldsymbol{f}} \|\boldsymbol{H}\boldsymbol{f} - \boldsymbol{w}^{(i-1)}\|_2^2 + \alpha_1 \|\boldsymbol{f} - \boldsymbol{u}^{(i-1)}\|_2^2$$

is equivalent to solving a linear system:

$$(\boldsymbol{H}^T\boldsymbol{H} + \alpha_1 \boldsymbol{I}) f = \boldsymbol{H}^T \boldsymbol{w}^{(i-1)} + \alpha_1 \boldsymbol{u}^{(i-1)}. \tag{2.2}$$

Because of the regularization term $\alpha_1 I$, the coefficient matrix $(\boldsymbol{H}^t\boldsymbol{H} + \alpha_1 I)$ is always invertible even $\boldsymbol{H}^t\boldsymbol{H}$ is singular, and the matrix $(\boldsymbol{H}^t\boldsymbol{H} + \alpha_1 I)$ is symmetry positive definite. The conjugate gradient method can be used to solve (2.2) at each iteration. Convergence can be improved using preconditioning techniques. Transform-based preconditioning techniques have been proved to be very successful [23].

Step 2. The minimizer of the second optimization problem is equivalent to solve N^2 minimizers of the function $\psi(s) = |t - s|^2 + \rho|s|$, where $\rho > 0$. In [9], the exact minimizer of $\psi(s)$ is given by the following

$$s = \phi(t) = \begin{cases} t - \frac{\rho}{2}, & t > \frac{\rho}{2}, \\ 0, & |t| \le \frac{\rho}{2}, \\ t + \frac{\rho}{2}, & t < -\frac{\rho}{2}. \end{cases} \tag{2.3}$$

Therefore, the minimizer of the optimization problem $\min_{\boldsymbol{w}} \|\boldsymbol{H}\boldsymbol{f}^{(i)}-\boldsymbol{w}\|_2^2+\alpha_2\|\boldsymbol{w}-\boldsymbol{g}\|_1$ is given by

$$\boldsymbol{w}_{j,k} = \begin{cases} (\boldsymbol{H}\boldsymbol{f}^{(i)})_{j,k} - \frac{\alpha_2}{2}, & (\boldsymbol{H}\boldsymbol{f}^{(i)} - \boldsymbol{g})_{j,k} > \frac{\alpha_2}{2}, \\ \boldsymbol{g}_{j,k}, & |(\boldsymbol{H}\boldsymbol{f}^{(i)} - \boldsymbol{g})_{j,k}| \le \frac{\alpha_2}{2}, \\ (\boldsymbol{H}\boldsymbol{f}^{(i)})_{j,k} + \frac{\alpha_2}{2}, & (\boldsymbol{H}\boldsymbol{f}^{(i)} - \boldsymbol{g})_{j,k} < -\frac{\alpha_2}{2}, \end{cases}$$

for $1 \le j,k \le N$, i.e., the function $\mathcal{S}_{l_1}(\boldsymbol{f}^{(i)}) = \phi(\boldsymbol{H}\boldsymbol{f}^{(i)} - \boldsymbol{g}) + \boldsymbol{g}$ according to the above formula.

Step 3. The third optimization problem

$$\alpha_1\|\boldsymbol{f}^{(i)} - \boldsymbol{u}\|_2^2 + \alpha_3\|\boldsymbol{u}\|_{TV} \tag{2.4}$$

can be solved by many TV denoising methods like Chambolle's projection algorithm [7], semismooth Newton's method [25], multilevel optimization method [12] and graph-based optimization method [8]. In the Chambolle scheme,we solve the following constrained minimization problem:

$$\min_{\boldsymbol{p}} \|\boldsymbol{f}^{(i)} - \frac{\alpha_3}{2\alpha_1}\mathrm{div}\boldsymbol{p}\|_2^2$$

subject to $|\boldsymbol{p}_{j,k}| \le 1$ for all $j,k = 1,2,\cdots,N$. Here

$$\boldsymbol{p}_{j,k} = \begin{bmatrix} \boldsymbol{p}_{j,k}^x \\ \boldsymbol{p}_{j,k}^y \end{bmatrix}$$

is the dual variable at the (j,k)th pixel location, $\boldsymbol{p}$ is the concatenation of all $\boldsymbol{p}_{j,k}$, and the discrete divergence of $\boldsymbol{p}$ is defined such that

$$(\mathrm{div}\boldsymbol{p})_{j,k} \equiv \boldsymbol{p}_{j,k}^x - \boldsymbol{p}_{j-1,k}^x + \boldsymbol{p}_{j,k}^y - \boldsymbol{p}_{j,k-1}^y$$

with $\boldsymbol{p}_{0,k}^x = \boldsymbol{p}_{j,0}^y = 0$. The vector $\mathrm{div}\boldsymbol{p}$ is the concatenation of all $(\mathrm{div}\boldsymbol{p})_{j,k}$. For simplicity, we denote $\lambda = \alpha_3/2\alpha_1$. When the minimizer $\boldsymbol{p}^*$ of the constrained optimization problem in (2.4)is determined, the denoised image $\boldsymbol{u}^{(i)}$ can be generated as follows:

$$\boldsymbol{u}^{(i)} = \boldsymbol{f}^{(i)} - \lambda\mathrm{div}\boldsymbol{p}^*.$$

In [7], the iterative scheme for computing the optimal solution $\boldsymbol{p}$ is given as follows:

$$\boldsymbol{p}_{j,k}^{l+1,x} = \frac{\boldsymbol{p}_{j,k}^{(l,x)} + \tau\nabla(\mathrm{div}\boldsymbol{p}^{(l)} - \boldsymbol{f}^{(i)}/\lambda)_{j,k}^x}{1 + \tau|\nabla(\mathrm{div}\boldsymbol{p}^{(l)} - \boldsymbol{f}^{(i)}/\lambda)_{j,k}|}, \quad \forall j,k = 1,2,\cdots,N$$

and

$$\boldsymbol{p}_{j,k}^{l+1,y} = \frac{\boldsymbol{p}_{j,k}^{(l,y)} + \tau\nabla(\mathrm{div}\boldsymbol{p}^{(l)} - \boldsymbol{f}^{(i)}/\lambda)_{j,k}^y}{1 + \tau|\nabla(\mathrm{div}\boldsymbol{p}^{(l)} - \boldsymbol{f}^{(i)}/\lambda)_{j,k}|}, \quad \forall j,k = 1,2,\cdots,N$$

where $\boldsymbol{p}_{j,k}^{(l,z)}(z \in \{x,y\})$ is the lth iterate of the iterative method for the minimizer, $\nabla(\cdot)_{j,k}^z(z \in \{x,y\})$ is defined as previous, and τ is a parameter introduced in the projection gradient method, see [7] for details.

3. The Convergence of Analysis. In this section, we study the convergence of the alternating minimization algorithm. We make use of the results by Browder and Petryshyn [6] and Opial [29] to show that the algorithm converges to a minimizer.

THEOREM 3.1. *If a nonexpansive mapping $\mathcal{T} : \mathcal{X} \to \mathcal{X}$ of a Hilbert space $\mathcal{X}$ into itself is asymptotically regular and has at least one fixed point then, for any $x \in \mathcal{X}$, a weak limit of a weakly convergence subsequence of the sequence of successive approximations $\{\mathcal{T}^n x\}$ is a fixed point.*

We note when $\mathcal{X}$ is $\mathbb{R}^{N^2}$, the sequence of successive approximations can converge strongly to a fixed point.

In order to use the theorem, we need to show that the objective function $\mathcal{J}$ in (1.9) is coercive, and $\mathcal{T}_1$ and $\mathcal{T}_2$ in (2.1) are non-expansive and asymptotically regular.

DEFINITION 3.2. *An operator $\mathcal{P}$ is called non-expansive if for any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{R}^{N^2}$, we have*

$$\|\mathcal{P}(\boldsymbol{x}_1) - \mathcal{P}(\boldsymbol{x}_2)\|_2 \leq \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2$$

If there exists some non-expansive operator $\mathcal{A}$ and $\alpha \in (0,1)$ such that $\mathcal{P} = (1-\alpha)\mathcal{I} + \alpha\mathcal{A}$, then $\mathcal{P}$ is called α-averaged non-expansive.

LEMMA 3.3. *[15] Let φ be convex and semi-continuous and $\alpha > 0$. Suppose $\hat{\boldsymbol{x}}$ is defined as follows:*

$$\hat{\boldsymbol{x}} = argmin_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2 + \alpha\varphi(\boldsymbol{x})$$

Define $\mathcal{S}$ such that $\hat{\boldsymbol{x}} = \mathcal{S}(\boldsymbol{y})$ for each $\boldsymbol{y}$. Then $\mathcal{S}$ is $\frac{1}{2}$-averaged non-expansive.

With Lemma 3.3, we know that $\mathcal{S}_{tv}$ is non-expansive. Next we show that the operators $\mathcal{T}_1$ and $\mathcal{T}_2$ defined in (2.1) are non-expansive. We first prove the operator $\mathcal{S}_{l_1}(\boldsymbol{u})$ is non-expansive, or we just show that $s = \phi(t)$ in (2.3) is non-expansive.

LEMMA 3.4. *The operator $s = \phi(t)$ defined in (2.3) is non-expansive.*

Proof. Let us consider the following four cases:

case 1. If $t_1, t_2 > \frac{\rho}{2}$ or $t_1, t_2 < -\frac{\rho}{2}$, then

$$|s_1 - s_2| = |\phi(t_1) - \phi(t_2)| = |t_1 - t_2|$$

case 2. If $|t_1| \leq \frac{\rho}{2}, |t_2| \leq \frac{\rho}{2}$, then

$$0 = |s_1 - s_2| = |\phi(t_1) - \phi(t_2)| \leq |t_1 - t_2|$$

case 3. If $t_1 > \frac{\rho}{2}$ and $|t_2| \leq \frac{\rho}{2}$, then

$$|s_1 - s_2| = |\phi(t_1) - \phi(t_2)| = |t_1 - \frac{\rho}{2} - 0| = |t_1 - \frac{\rho}{2}| \leq |t_1 - |t_2|| \leq |t_1 - t_2|$$

case 4. If $t_1 > \frac{\rho}{2}$ and $t_2 < -\frac{\rho}{2}$, then by noting that $t_1 - t_2 > \rho$, we have

$$|s_1 - s_2| = |\phi(t_1) - \phi(t_2)| = |t_1 - t_2 - \rho| \leq |t_1 - t_2|.$$

The result follows. □

LEMMA 3.5. *The operators $\mathcal{T}_1$ and $\mathcal{T}_2$ in (2.1) are non-expansive.*

Proof. We note that

$$\begin{aligned} &\|\mathcal{T}_1(\boldsymbol{w}_1) - \mathcal{T}_1(\boldsymbol{w}_2)\|_2 \\ &= \|\mathcal{S}_{l_1}(\mathcal{S}_h(\boldsymbol{w}_1, \boldsymbol{u})) - \mathcal{S}_{l_1}(\mathcal{S}_h(\boldsymbol{w}_2, \boldsymbol{u}))\|_2 \end{aligned}$$

$$
\begin{aligned}
&= \|\phi(\boldsymbol{H}\mathcal{S}_h(\boldsymbol{w}_1,\boldsymbol{u}) - \boldsymbol{g}) + \boldsymbol{g} - \phi(\boldsymbol{H}\mathcal{S}_h(\boldsymbol{w}_2,\boldsymbol{u}) - \boldsymbol{g}) - \boldsymbol{g}\| \\
&\leq \|\boldsymbol{H}\mathcal{S}_h(\boldsymbol{w}_1,\boldsymbol{u}) - \boldsymbol{H}\mathcal{S}_h(\boldsymbol{w}_1,\boldsymbol{u})\|_2 \\
&= \|\boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{H} + \alpha_1\boldsymbol{I})^{-1}(\boldsymbol{H}^T\boldsymbol{w}_1 + \alpha_1\boldsymbol{u}) - \boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{H} + \alpha_1\boldsymbol{I})^{-1}(\boldsymbol{H}^T\boldsymbol{w}_2 + \alpha_1\boldsymbol{u})\|_2 \\
&= \|\boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{H} + \alpha_1\boldsymbol{I})^{-1}\boldsymbol{H}^T(\boldsymbol{w}_1 - \boldsymbol{w}_2)\|_2 \\
&\leq \|\boldsymbol{w}_1 - \boldsymbol{w}_2\|_2,
\end{aligned}
$$

and

$$
\begin{aligned}
&\|\mathcal{T}_2(\boldsymbol{u}_1) - \mathcal{T}_2(\boldsymbol{u}_2)\|_2 \\
&= \|\mathcal{S}_{tv}(\mathcal{S}_h(\boldsymbol{w},\boldsymbol{u}_1)) - \mathcal{S}_{tv}(\mathcal{S}_h(\boldsymbol{w},\boldsymbol{u}_2))\|_2 \\
&\leq \|\mathcal{S}_h(\boldsymbol{w},\boldsymbol{u}_1) - \mathcal{S}_h(\boldsymbol{w},\boldsymbol{u}_2)\|_2 \\
&= \|(\boldsymbol{H}^T\boldsymbol{H} + \alpha_1\boldsymbol{I})^{-1}(\boldsymbol{H}^T\boldsymbol{w} + \alpha_1\boldsymbol{u}_1) - (\boldsymbol{H}^T\boldsymbol{H} + \alpha_1\boldsymbol{I})^{-1}(\boldsymbol{H}^T\boldsymbol{w} + \alpha_1\boldsymbol{u}_2)\|_2 \\
&= \|\alpha_1(\boldsymbol{H}^T\boldsymbol{H} + \alpha_1\boldsymbol{I})^{-1}(\boldsymbol{u}_1 - \boldsymbol{u}_2)\|_2 \\
&\leq \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_2
\end{aligned}
$$

The results follow. □

LEMMA 3.6. *Let $\boldsymbol{w}^{(i)}$ and $\boldsymbol{u}^{(i)}$ generated by (2.1), then both $\sum_{i=1}^{\infty}\|\boldsymbol{u}^{(i)} - \boldsymbol{u}^{(i-1)}\|_2^2$ and $\sum_{i=1}^{\infty}\|\boldsymbol{w}^{(i)} - \boldsymbol{w}^{(i-1)}\|_2^2$ converge.*

Proof. In (1.9), we consider the Taylor series expansion of $\mathcal{J}(\boldsymbol{w}^{(i)},\boldsymbol{f},\boldsymbol{u}^{(i)})$ in the second variable and set $\boldsymbol{f} = \boldsymbol{f}^{(i)}$ in the calculation, we have

$$
\begin{aligned}
\mathcal{J}(\boldsymbol{w}^{(i)},\boldsymbol{f}^{(i)},\boldsymbol{u}^{(i)}) = \mathcal{J}(\boldsymbol{w}^{(i)},\boldsymbol{f}^{(i+1)},\boldsymbol{u}^{(i)}) &+ (\boldsymbol{f}^{(i)} - \boldsymbol{f}^{(i+1)})^T\frac{\partial\mathcal{J}}{\partial\boldsymbol{f}}(\boldsymbol{w}^{(i)},\boldsymbol{f}^{(i+1)},\boldsymbol{u}^{(i)}) \\
&+\frac{1}{2}(\boldsymbol{f}^{(i)} - \boldsymbol{f}^{(i+1)})^T\frac{\partial^2\mathcal{J}}{\partial\boldsymbol{f}^2}(\boldsymbol{w}^{(i)},\boldsymbol{f}^{(i+1)},\boldsymbol{u}^{(i)})(\boldsymbol{f}^{(i)} - \boldsymbol{f}^{(i+1)})
\end{aligned}
$$

Since $\boldsymbol{f}^{(i+1)}$ is the minimizer of $\mathcal{J}(\boldsymbol{w}^{(i)},\boldsymbol{f},\boldsymbol{u}^{(i)})$, we have

$$
\frac{\partial\mathcal{J}}{\partial\boldsymbol{f}}(\boldsymbol{w}^{(i)},\boldsymbol{f}^{(i+1)},\boldsymbol{u}^{(i)}) = 0.
$$

Next we can compute

$$
\frac{\partial^2\mathcal{J}}{\partial\boldsymbol{f}^2}(\boldsymbol{w}^{(i)},\boldsymbol{f}^{(i+1)},\boldsymbol{u}^{(i)}) = 2(\boldsymbol{H}^T\boldsymbol{H} + \alpha_1\boldsymbol{I}) \geq 2\alpha_1\boldsymbol{I}.
$$

Therefore, we obtain

$$
\mathcal{J}(\boldsymbol{w}^{(i)},\boldsymbol{f}^{(i)},\boldsymbol{u}^{(i)}) - \mathcal{J}(\boldsymbol{w}^{(i)},\boldsymbol{f}^{(i+1)},\boldsymbol{u}^{(i)}) \geq \alpha_1\|\boldsymbol{f}^{(i)} - \boldsymbol{f}^{(i+1)}\|_2^2.
$$

As $\mathcal{J}(\boldsymbol{w}^{(i+1)},\boldsymbol{f}^{(i+1)},\boldsymbol{u}^{(i+1)}) \leq \mathcal{J}(\boldsymbol{w}^{(i)},\boldsymbol{f}^{(i+1)},\boldsymbol{u}^{(i)})$, we obtain

$$
\begin{aligned}
\mathcal{J}(\boldsymbol{w}^{(i)},\boldsymbol{f}^{(i)},\boldsymbol{u}^{(i)}) - \mathcal{J}(\boldsymbol{w}^{(i+1)},\boldsymbol{f}^{(i+1)},\boldsymbol{u}^{(i+1)}) &\geq \mathcal{J}(\boldsymbol{w}^{(i)},\boldsymbol{f}^{(i)},\boldsymbol{u}^{(i)}) - \mathcal{J}(\boldsymbol{w}^{(i)},\boldsymbol{f}^{(i+1)},\boldsymbol{u}^{(i)}) \\
&\geq \alpha_1\|\boldsymbol{f}^{(i)} - \boldsymbol{f}^{(i+1)}\|_2^2.
\end{aligned}
$$

Since the operator $\mathcal{S}_{tv}$ is non-expansive, we have

$$
\|\boldsymbol{f}^{(i)} - \boldsymbol{f}^{(i+1)}\|_2^2 \geq \|\mathcal{S}_{tv}(\boldsymbol{f}^{(i)}) - \mathcal{S}_{tv}(\boldsymbol{f}^{(i+1)})\|_2^2 = \|\boldsymbol{u}^{(i)} - \boldsymbol{u}^{(i+1)}\|_2^2.
$$

Hence we get

$$
\mathcal{J}(\boldsymbol{w}^{(i)},\boldsymbol{f}^{(i)},\boldsymbol{u}^{(i)}) - \mathcal{J}(\boldsymbol{w}^{(i+1)},\boldsymbol{f}^{(i+1)},\boldsymbol{u}^{(i+1)}) \geq \alpha_1\|\boldsymbol{u}^{(i)} - \boldsymbol{u}^{(i+1)}\|_2^2
$$

It follows that $\sum_{i=1}^{\infty} \|\boldsymbol{u}^{(i)} - \boldsymbol{u}^{(i-1)}\|_2^2$ is bounded, and $\sum_{i=1}^{\infty} \|\boldsymbol{u}^{(i)} - \boldsymbol{u}^{(i-1)}\|_2^2$ converges. By using a similar argument, we can obtain that $\sum_{i=1}^{\infty} \|\boldsymbol{w}^{(i)} - \boldsymbol{w}^{(i-1)}\|_2^2$ is bounded, and therefore $\sum_{i=1}^{\infty} \|\boldsymbol{w}^{(i)} - \boldsymbol{w}^{(i-1)}\|_2^2$ converges. □

Immediately, we have the following lemma, which states that the operators $\mathcal{T}_1$ and $\mathcal{T}_2$ are asymptotically regular.

LEMMA 3.7. *For any initial guess $\boldsymbol{w}^{(0)}, \boldsymbol{u}^{(0)} \in \mathbb{R}^{N^2}$, suppose $\{\boldsymbol{w}^{(i)}\}$ and $\{\boldsymbol{u}^{(i)}\}$ are generated by (2.1), then both $\mathcal{T}_1$ and $\mathcal{T}_2$ are asymptotically regular, i.e.,*

$$\lim_{i\to\infty} \|\boldsymbol{w}^{(i+1)} - \boldsymbol{w}^{(i)}\|_2 = \lim_{i\to\infty} \|\mathcal{T}_1^{i+1}(\boldsymbol{w}^{(0)}) - \mathcal{T}_1^{i}(\boldsymbol{w}^{(0)})\|_2 = 0$$
$$\lim_{i\to\infty} \|\boldsymbol{u}^{(i+1)} - \boldsymbol{u}^{(i)}\|_2 = \lim_{i\to\infty} \|\mathcal{T}_2^{i+1}(\boldsymbol{u}^{(0)}) - \mathcal{T}_2^{i}(\boldsymbol{u}^{(0)})\|_2 = 0$$

To show the coerciveness of $\mathcal{J}$, we introduce the following definition.

DEFINITION 3.8. *A function $\phi : \mathbb{R}^{N^2} \to \mathbb{R}$ is proper over a set $X \subset \mathbb{R}^{N^2}$ if $\phi(\boldsymbol{x}) < \infty$ for at least one $\boldsymbol{x} \in X$ and $\phi(\boldsymbol{x}) > -\infty$ for all $\boldsymbol{x} \in X$. A function $\phi : \mathbb{R}^{N^2} \to \mathbb{R}$ is coercive over a set $X \subset \mathbb{R}^{N^2}$ if for every sequence $\{\boldsymbol{x}_k\} \subset X$ such that $\|\boldsymbol{x}_k\|_2 \to \infty$, we have*

$$\lim_{k\to\infty} \phi(\boldsymbol{x}_k) = \infty.$$

When $X = \mathbb{R}^{N^2}$, we say that ϕ is coercive on $\mathbb{R}^{N^2}$.

LEMMA 3.9. *[4] Let $\varphi : \mathbb{R}^{N^2} \to \mathbb{R}$ be a closed, proper and coercive function. Then the set of minima of φ over $\mathbb{R}^{N^2}$ is nonempty and compact.*

The following lemma states the objective function $\mathcal{J}(\boldsymbol{w}, \boldsymbol{f}, \boldsymbol{u})$ is coercive under certain condition.

LEMMA 3.10. *Let $\boldsymbol{L}_h$ and $\boldsymbol{L}_v$ be the one-side difference matrix on the horizontal direction and the vertical direction respectively, and*

$$\boldsymbol{L} = \begin{pmatrix} \boldsymbol{L}_h \\ \boldsymbol{L}_v \end{pmatrix}$$

The function $\mathcal{J}(\boldsymbol{w}, \boldsymbol{f}, \boldsymbol{u})$ is coercive if $Null(\boldsymbol{H}) \cap Null(\boldsymbol{L}) = \emptyset$, where $Null(\cdot)$ denotes the null space of the corresponding matrix.

Proof. When $\| \begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{u} \\ \boldsymbol{w} \end{pmatrix} \|_2 \to \infty$, we must have either $\| \begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{u} \end{pmatrix} \|_2 \to \infty$ or $\|\boldsymbol{w}\|_2 \to \infty$. Let us discuss these two cases. In the first case, when $\|\boldsymbol{w}\|_2 \to \infty$, it is clear that $\|\boldsymbol{w} - \boldsymbol{g}\|_1 \to \infty$, and therefore $\mathcal{J}(\boldsymbol{w}, \boldsymbol{f}, \boldsymbol{u}) \to \infty$, the result follows. In the second case, $\|\boldsymbol{w}\|_2$ is bounded, otherwise the result follows. We would like to show that $\mathcal{J}(\boldsymbol{w}, \boldsymbol{f}, \boldsymbol{u}) \to \infty$ when $\| \begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{u} \end{pmatrix} \|_2 \to \infty$.

The lower bound of the discrete total variation is given by

$$\begin{aligned} \|\boldsymbol{u}\|_{TV} &= \sum_{1\le j,k\le N} |(\nabla \boldsymbol{u})_{j,k}|_2 \\ &= \sum_{1\le j,k\le N} \sqrt{|(\nabla \boldsymbol{u})^x_{j,k}|^2 + |(\nabla \boldsymbol{u})^y_{j,k}|^2} \\ &\ge \frac{1}{\sqrt{2}} \sum_{1\le j,k\le N} |(\nabla \boldsymbol{u})^x_{j,k}| + |(\nabla \boldsymbol{u})^y_{j,k}| = \frac{1}{\sqrt{2}} \|\boldsymbol{L}\boldsymbol{u}\|_1. \end{aligned}$$

By using the above inequality, we have

$$\begin{aligned}(3.1)\qquad \mathcal{J}(\boldsymbol{w},\boldsymbol{f},\boldsymbol{u}) &\geq \|\boldsymbol{H}\boldsymbol{f}-\boldsymbol{w}\|_2^2+\alpha_1\|\boldsymbol{f}-\boldsymbol{u}\|_2^2+\frac{\alpha_3}{\sqrt{2}}\|\boldsymbol{L}\boldsymbol{u}\|_1\\ &\geq \|\boldsymbol{H}\boldsymbol{f}-\boldsymbol{w}\|_2^2+\alpha_1\|\boldsymbol{f}-\boldsymbol{u}\|_2^2+\frac{\alpha_3}{\sqrt{2}}\|\boldsymbol{L}\boldsymbol{u}\|_2\\ &= \left\|\begin{pmatrix}\boldsymbol{H} & 0\\ \sqrt{\alpha_1}I & -\sqrt{\alpha_1}I\\ 0 & \frac{\alpha_3}{\sqrt{2}}L\end{pmatrix}\begin{pmatrix}\boldsymbol{f}\\ \boldsymbol{u}\end{pmatrix}-\begin{pmatrix}\boldsymbol{w}\\ 0\\ 0\end{pmatrix}\right\|_2^2.\end{aligned}$$

Let

$$\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\\ \boldsymbol{z}\end{pmatrix}=\begin{pmatrix}\boldsymbol{H} & 0\\ \sqrt{\alpha_1}I & -\sqrt{\alpha_1}I\\ 0 & \frac{\alpha_3}{\sqrt{2}}L\end{pmatrix}\begin{pmatrix}\boldsymbol{f}\\ \boldsymbol{u}\end{pmatrix}$$

We note that

$$\begin{pmatrix}\boldsymbol{H} & 0\\ \sqrt{\alpha_1}I & -\sqrt{\alpha_1}I\\ 0 & \frac{\alpha_3}{\sqrt{2}}L\end{pmatrix}=\begin{pmatrix}I & 0 & 0\\ 0 & I & 0\\ 0 & -\frac{\alpha_3}{\sqrt{2\alpha_1}}L & I\end{pmatrix}\begin{pmatrix}\boldsymbol{H} & 0\\ \sqrt{\alpha_1}I & -\sqrt{\alpha_1}I\\ \frac{\alpha_3}{\sqrt{2}}L & 0\end{pmatrix}$$

and the above matrix is full rank as $Null(\boldsymbol{H})\cap Null(\boldsymbol{L})=\emptyset$. Hence when $\|\begin{pmatrix}\boldsymbol{f}\\ \boldsymbol{u}\end{pmatrix}\|_2\rightarrow\infty$, we have either $\|\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}\|_2\rightarrow\infty$ or $\|z\|_2\rightarrow\infty$. By using (3.1), we get $\mathcal{J}(\boldsymbol{w},\boldsymbol{f},\boldsymbol{u})\rightarrow\infty$, The result follows. □

REMARK 1. *If $\boldsymbol{x}\in Null(\boldsymbol{L})$, then $\boldsymbol{x}_{i,j}=c$ for any i and j, where c is a nonzero constant. Since $\boldsymbol{H}$ is a blurring matrix (all the nonzero entries should be positive), it is clear that $\boldsymbol{H}\boldsymbol{x}$ is a nonzero vector. It follows that the assumption $Null(\boldsymbol{H})\cap Null(\boldsymbol{L})=\emptyset$ holds in general.*

Now we show that the set of fixed points of $\mathcal{T}_1$ and $\mathcal{T}_2$ are non-empty.

LEMMA 3.11. *Suppose $Null(\boldsymbol{H})\cap Null(\boldsymbol{L})=\emptyset$. Then the set of fixed points of $\mathcal{T}_1$ and $\mathcal{T}_2$ are non-empty.*

Proof. Since the objective function $\mathcal{J}$ is coercive, the set of minimizers of $\mathcal{J}$ is non-empty. Assume $(\boldsymbol{w}',\boldsymbol{f}',\boldsymbol{u}')$ is a minimizer of $\mathcal{J}$, i.e.,

$$\begin{pmatrix}\frac{\partial\mathcal{J}}{\partial\boldsymbol{w}}(\boldsymbol{w}',\boldsymbol{f}',\boldsymbol{u}')\\ \frac{\partial\mathcal{J}}{\partial\boldsymbol{f}}(\boldsymbol{w}',\boldsymbol{f}',\boldsymbol{u}')\\ \frac{\partial\mathcal{J}}{\partial\boldsymbol{u}}(\boldsymbol{w}',\boldsymbol{f}',\boldsymbol{u}')\end{pmatrix}=\begin{pmatrix}0\\ 0\\ 0\end{pmatrix}$$

It implies that

$$\begin{cases}\boldsymbol{f}' = \mathcal{S}_h(\boldsymbol{w}',\boldsymbol{u}') = \text{argmin}\mathcal{J}(\boldsymbol{w}',\cdot,\boldsymbol{u}')\\ \boldsymbol{w}' = \mathcal{S}_{l_1}(\boldsymbol{f}') = \text{argmin}\mathcal{J}(\cdot,\boldsymbol{f}',\boldsymbol{u}')\\ \boldsymbol{u}' = \mathcal{S}_{tv}(\boldsymbol{f}') = \text{argmin}\mathcal{J}(\boldsymbol{w}',\boldsymbol{f}',\cdot)\end{cases}$$

Thus we obtain

$$\begin{aligned}\boldsymbol{w}' &= \mathcal{S}_{l_1}(\mathcal{S}_h(\boldsymbol{w}',\boldsymbol{u}'))=\mathcal{T}_1(\boldsymbol{w}')\\ \boldsymbol{u}' &= \mathcal{S}_{tv}(\mathcal{S}_h(\boldsymbol{w}',\boldsymbol{u}'))=\mathcal{T}_2(\boldsymbol{u}')\end{aligned}$$

i.e., $\boldsymbol{w}'$ and $\boldsymbol{u}'$ are the fixed points of $\mathcal{T}_1$ and $\mathcal{T}_2$ respectively. The result follows. □

We remark that when $\boldsymbol{H}$ is a matrix of full column rank, then the objective function $\mathcal{J}$ in (1.9) is strictly convex. Therefore a fixed point of $\mathcal{T}$ is also a global minimizer of $\mathcal{J}$. According to Theorem 3.1, the sequence $\{\boldsymbol{u}^{(i)}\}$ converges to a fixed point of $\mathcal{J}$, i.e., a minimizer of $\mathcal{J}$.

THEOREM 3.12. *Suppose Null* $(\boldsymbol{H}) \bigcap$ *Null* $(\boldsymbol{L}) = \emptyset$. *For any initial guess* $\boldsymbol{w}^{(0)}, \boldsymbol{u}^{(0)} \in \mathbb{R}^{N^2}$, *If* $\boldsymbol{w}^{(i)}$ *and* $\boldsymbol{u}^{(i)}$ *generated by (2.1), then both* $\boldsymbol{w}^{(i)}$ *and* $\boldsymbol{u}^{(i)}$ *converge to a stationary point of* $\mathcal{J}$.

We remark that when $\boldsymbol{H}$ is a matrix of full column rank, both $\boldsymbol{u}^{(i)}$ and $\boldsymbol{w}^{(i)}$ converge to a minimizer of $\mathcal{J}$.

4. Experimental Results. In this section, we illustrate the performance of our proposed algorithm for image restoration problems. Relative error, Blurred Signal to Noise Ratio (BSNR) and the highest peak Signal-to-Noise Ratio (PSNR) are used to measure the quality of the restored image. They are defined as follows:

$$\text{ReErr} = \frac{\|\tilde{\boldsymbol{f}} - \boldsymbol{f}\|_2}{\|\boldsymbol{f}\|_2}, \quad \text{BSNR} = 20\log_{10}\left(\frac{\|\boldsymbol{g}\|_2}{\|\boldsymbol{n}\|_2}\right),$$

$$\text{SNR} = 10\log_{10}\frac{\|\boldsymbol{f} - \boldsymbol{f}_{\text{mean}}\|_2}{\|\boldsymbol{f} - \tilde{\boldsymbol{f}}\|_2}, \quad \text{PSNR} = -20\log_{10}\left(\frac{\|\tilde{\boldsymbol{f}} - \boldsymbol{f}\|_2}{N}\right)$$

where $\boldsymbol{f}, \boldsymbol{g}, \boldsymbol{n}$ and $\tilde{\boldsymbol{f}}$ are the original image, the observed image, the noise vector added in the test, and the restored image respectively. We compare the proposed method (ℓ_1-TV) with the LAD method [18] where the objective function is given in (1.8), and the FTVd algorithm [33] where the objective function is given in (1.7).

4.1. The Proposed Method and The LAD Method. Three original images ("Cameraman", "Satellite" and "Wheel") are used to test the proposed method, and they are shown in Figure 4.1. The blurring function is chosen to be a two-dimensional Gaussian,

$$h(i,j) = e^{-2(i/3)^2 - 2(j/3)^2},$$

truncated such that the function has support of 7×7, which is used in [18]. In the tests, three kinds of noise on the blurred image are used, namely, Gaussian noise, uniform noise and Laplace noise. The blurred and noisy "Cameraman", "Satellite" and "Wheel" images are shown in Figures 4.2–4.4(a), Figures 4.5–4.7(a) and Figures 4.8–4.10(a) respectively. The stopping criterion of the two methods is that the relative difference between the successive iterate of the restored image should satisfy the following inequality:

$$\frac{\|\tilde{\boldsymbol{f}}^{(i+1)} - \tilde{\boldsymbol{f}}^{(i)}\|_2}{\|\tilde{\boldsymbol{f}}^{(i+1)}\|_2} < 10^{-4}, \tag{4.1}$$

where $\tilde{\boldsymbol{f}}^{(i)}$ is the computed image at the ith iteration of the proposed method.

The restored images by using the LAD method are shown in Figures 4.2-4.10(b). The restored images by using the ℓ_1-TV method are shown in Figures 4.2-4.10(c). According to the figures, the quality of the restored images by using the ℓ_1-TV method is better than those by using the LAD method. In Tables 4.1-4.3, we summarized the

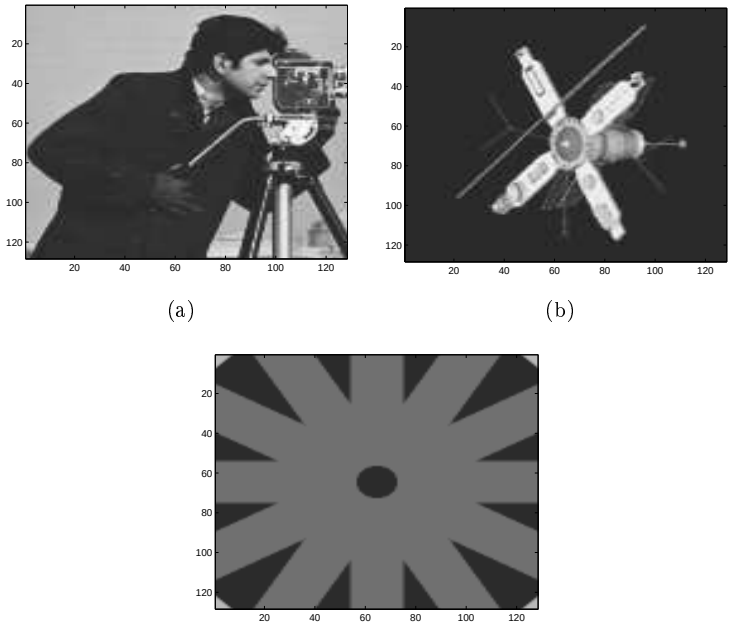


FIG. 4.1. *(a) The original "Cameraman" image; (b) the original "Satellite" image; (c) the original "Wheel" image.*

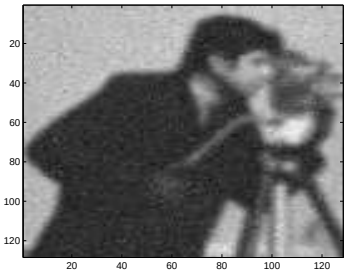
restoration results of the three tested images. In the tables, we see that the PSNR of the restored images by the ℓ_1-TV method is larger than those by the LAD method. Also the computational time required by the ℓ_1-TV method is significantly less that that required by the LAD method. It shows that the proposed method is very effective and efficient.

In Table 4.5, we show three regularization parameters used in the ℓ_1-TV method for restoration of four tested images. We determine the best value of α_1, α_2 and α_3 such that the relative error of the restored image $\tilde{\boldsymbol{f}}$ with respect to the original image $\boldsymbol{f}$ is the smallest, i.e.,

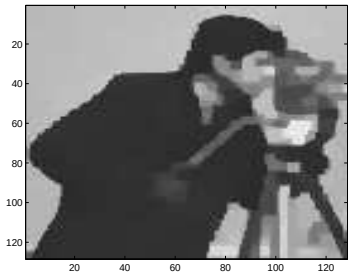
$$\frac{\|\tilde{\boldsymbol{f}} - \boldsymbol{f}\|_2}{\|\boldsymbol{f}\|_2}$$

is the smallest among all tested values of α_1, α_2 and α_3. We remark that the noise added to the blurred image is independent and identically distributed (i.i.d.), i.e., each pixel location has the same probability distribution as the others and all are mutually independent. Also the same blurring matrix is applied to the original image, therefore we use the same set of parameters for different images when the same kind of noise is considered. We note that such implementation may not be optimal with respect to the relative error of the restored image and the original image. However, the computational cost of searching regularization parameters can be reduced.

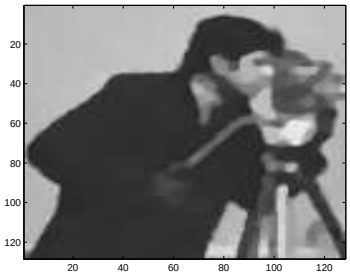
4.2. The Proposed Method and the FTVd method. In this subsection, we compare the performance of the proposed method with that of the FTVd method. We



(a)

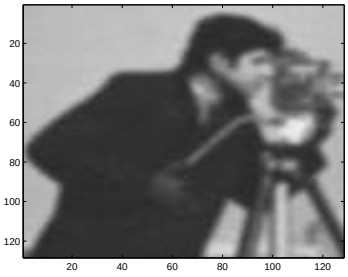


(b)



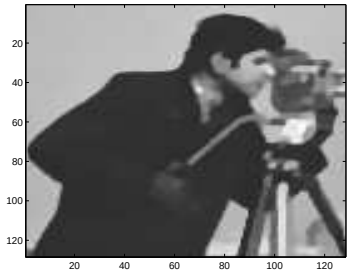
(c)

FIG. 4.2. *(a) The blurred noisy image with 50% of the pixels contaminated by Gaussian noise. (b) The restored image using the LAD method. (c) The restored image using the L1-TV method.*



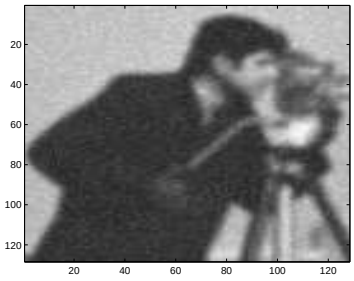
(a)

(b)

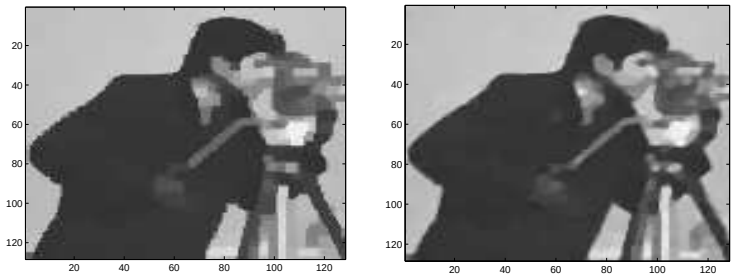


(c)

FIG. 4.3. *(a) The blurred noisy image with 50% of the pixels contaminated by Laplace noise. (b) The restored image using the LAD method. (c) The restored image using the L1-TV method.*

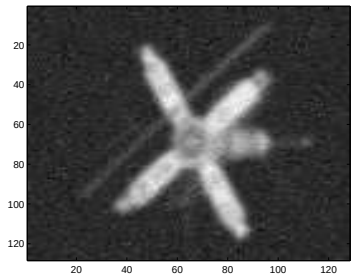


(a)

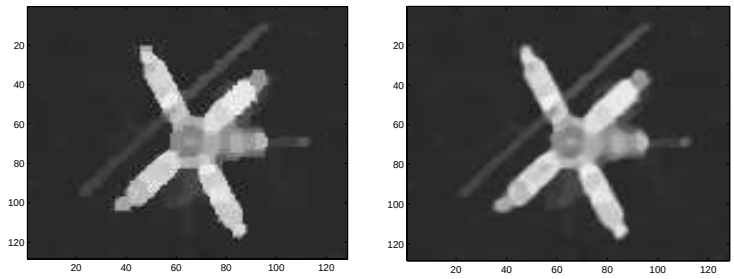


(b) (c)

FIG. 4.4. *(a) The blurred noisy image with* 50% *of the pixels contaminated by uniform noise. (b) The restored image using the LAD method. (c) The restored image using the L1-TV method.*

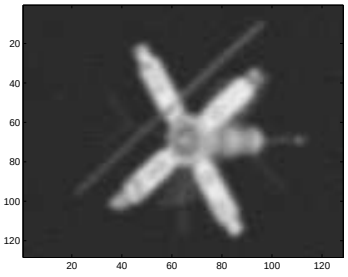


(a)

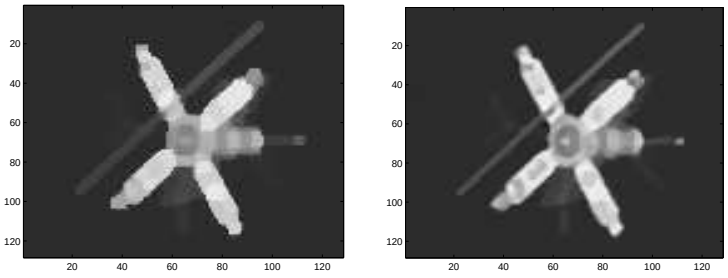


(b) (c)

FIG. 4.5. *(a) The blurred noisy image with* 50% *of the pixels contaminated by Gaussian noise. (b) The restored image using the LAD method. (c) The restored image using the L1-TV method.*

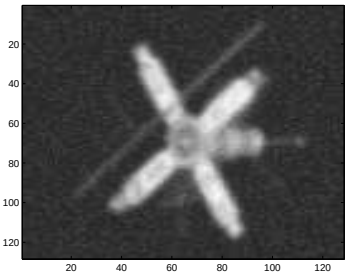


(a)

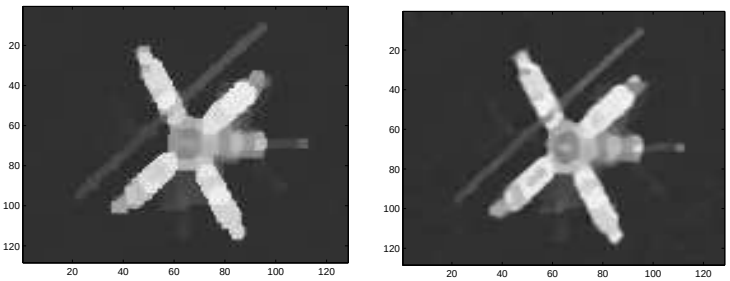


(b) (c)

FIG. 4.6. *(a) The blurred noisy image with* 50% *of the pixels contaminated by Laplace noise. (b) The restored image using the LAD method. (c) The restored image using the L1-TV method.*



(a)



(b) (c)

FIG. 4.7. *(a) The blurred noisy image with* 50% *of the pixels contaminated by uniform noise. (b) The restored image using the LAD method. (c) The restored image using the L1-TV method.*

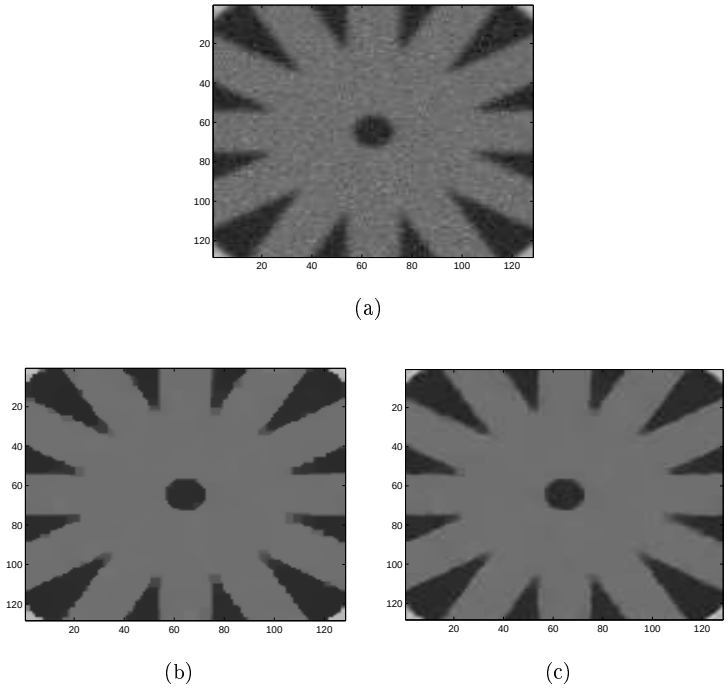


FIG. 4.8. *(a) The blurred noisy image with 50% of the pixels contaminated by Gaussian noise. (b) The restored image using the LAD method. (c) The restored image using the L1-TV method.*

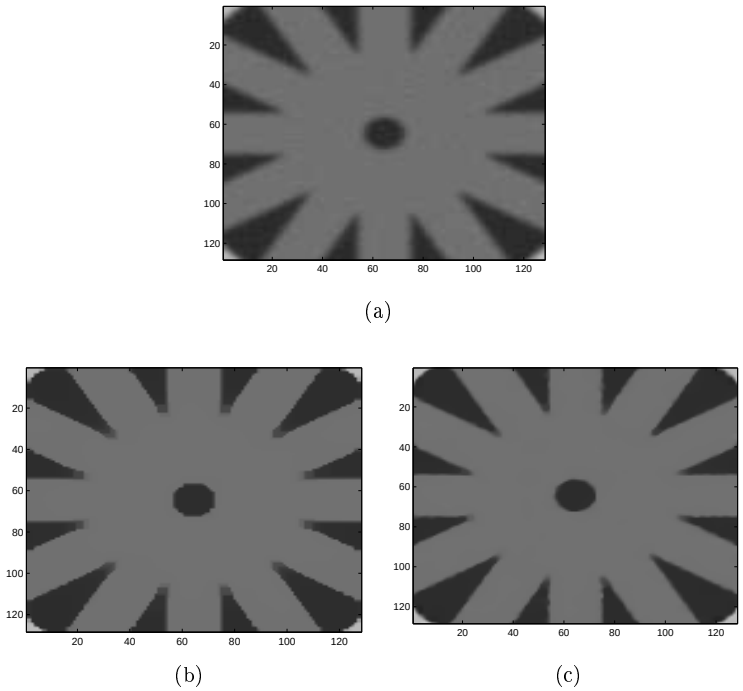


FIG. 4.9. *(a) The blurred noisy image with 50% of the pixels contaminated by Laplace noise. (b) The restored image using the LAD method. (c) The restored image using the L1-TV method.*

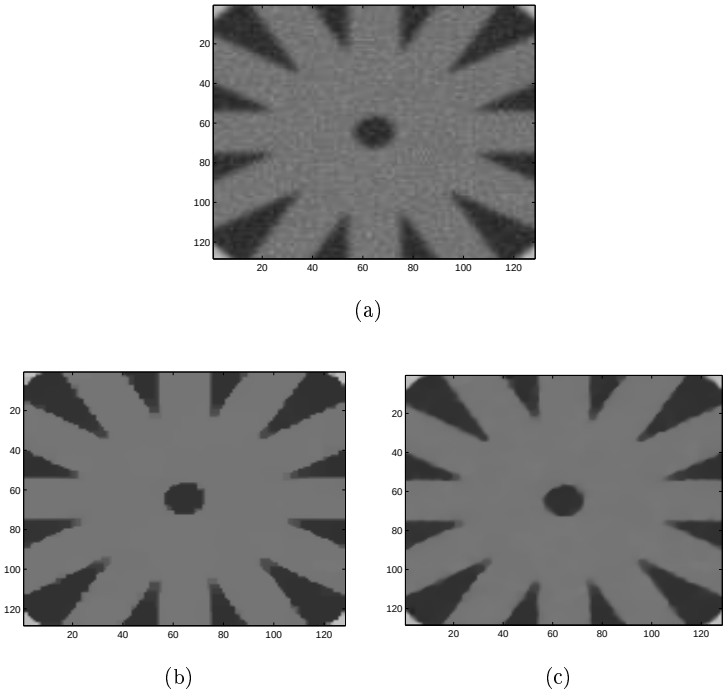


FIG. 4.10. *(a) The blurred noisy image with* 50% *of the pixels contaminated by uniform noise. (b) The restored image using the LAD method. (c) The restored image using the L1-TV method.*

use the "Cameraman" image as the original image. The salt-and-pepper noise is added to the Gaussian blurred images, see [33]. The implementation of the FTVd method can be found in HTTP://WWW.CAAM.RICE.EDU/ OPTIMIZATION/L1/FTVD/. In [33], a continuation scheme on γ and β in (1.7) is used. We follow the same continuation scheme on α_1 and α_2 in (1.9). For the FTVd method, we download the program and generate the image restoration results for comparison. The stopping criterion of the FTVd is set in the program, see [33] for detail. The stopping criterion of the proposed method is the same as that of the previous subsection: the relative difference between the successive iterate of the restored image should satisfy the following inequality in (4.1). The image restoration results are summarized in Table 4.5, and the restored images are shown in Figures 4.11 and 4.12. We see from the figures that both methods can restore images with about the same visual quality. In Table 4.5, we summarize the restoration results of the two methods. We find that the SNRs and PSNRs of the restored images by using two methods are about the same, and the computational time required by the proposed method is slightly faster than that required by the FTVd method.

5. Concluding Remarks. In this paper, we have studied a fast ℓ_1 data fitting plus total variation minimization method for image restoration. We have employed an alternating minimization algorithm to solve the proposed total variation minimization problem. Our experimental results show that the quality of restored images by the proposed method is competitive with those restored by the LAD method and the FTVd method. The most important contribution is that the proposed algorithm is

TABLE 4.1
Summary results for "Cameraman" image.

Noise	Gaussian BSNR=22.8853		Laplace BSNR=33.8851		Uniform BSNR=22.8004	
	LAD	ℓ_1-TV	LAD	ℓ_1-TV	LAD	ℓ_1-TV
PSNR (dB)	23.4816	24.0763	23.7781	25.6610	22.9887	24.0077
Relative error	0.1329	0.1241	0.1284	0.1034	0.1405	0.1251
Time (seconds)	49.2	14.4	91.2	5.00	52.0	6.55

TABLE 4.2
Summary results for "Satellite" image.

Noise	Gaussian BSNR=18.7376		Laplace BSNR=29.8226		Uniform BSNR=18.9050	
	LAD	ℓ_1-TV	LAD	ℓ_1-TV	LAD	ℓ_1-TV
PSNR (dB)	26.0406	27.3872	26.1986	29.3262	25.2148	26.7752
Relative error	0.1565	0.1340	0.1536	0.1077	0.1721	0.1438
Time (seconds)	67.0	5.9	49.8	3.81	68.4	4.14

TABLE 4.3
Summary results for "Wheel" image.

Noise	Gaussian BSNR=20.9160		Laplace BSNR=31.9807		Uniform BSNR=21.0149	
	LAD	ℓ_1-TV	LAD	ℓ_1-TV	LAD	ℓ_1-TV
PSNR (dB)	31.8726	33.3470	33.5339	35.0963	29.4192	30.1312
Relative error	0.0641	0.0541	0.0530	0.0443	0.0851	0.0784
Time (seconds)	39.4	9.72	52.5	4.34	71.7	10.3

TABLE 4.4
The parameters used for restored images by the ℓ_1-TV method.

Noise	Image	α_1	α_2	$\lambda = \alpha_3/2\alpha_1$
Gaussian	Cameraman	0.5	0.026	0.018
	Satellite	0.5	0.026	0.018
	Wheel	0.5	0.026	0.018
Laplace	Cameraman	0.0545	0.055	0.018
	Satellite	0.0545	0.055	0.018
	Wheel	0.0545	0.055	0.018
Uniform	Cameraman	0.2	0.05	0.020
	Satellite	0.2	0.05	0.020
	Wheel	0.2	0.05	0.020

very efficient.

The future research plan is to investigate how to select the regularization parameters in (1.9). We can make use of the continuation scheme to make α_1 and α_2 in (1.9) larger and larger such that $\boldsymbol{f} \approx \boldsymbol{u}$ and $\boldsymbol{w} \approx \boldsymbol{g}$. However, the regularization parameter α_3 is required to choose in order to restore a high quality of an image. In [22], we have developed a fast total variation image restoration method with automatic selection of regularization parameters to restore blurred and noisy images when $\ell 2$-norm is used

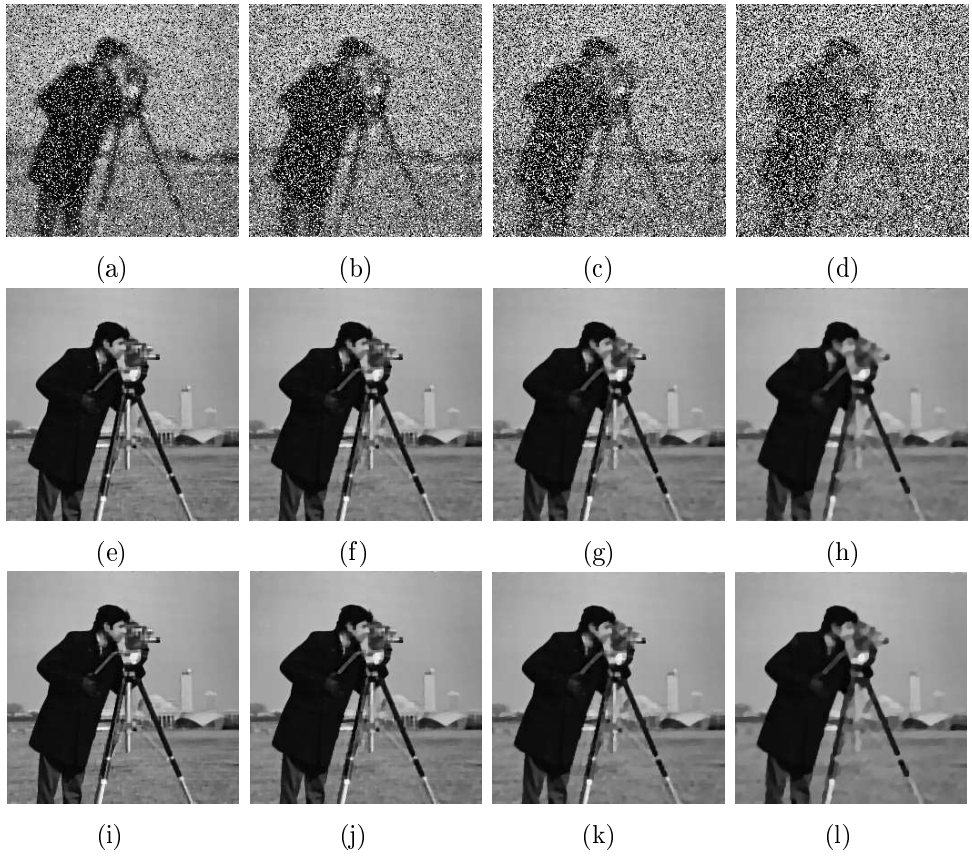


FIG. 4.11. *(a)-(d) Images with Gaussian blur of size* 7×7 *with standard derviation = 5 and salt-and-pepper noise from 30% to 60%; (e)-(h) restored images by the FTVd method; (i)-(l) restored images by the proposed method.*

in the data-fitting term. The method exploits the generalized cross-validation (GCV) technique to determine inexpensively how much regularization used in each restoration step. By updating these regularization parameters in the iterative procedure, the restored image can be obtained by choosing the one having the minimum GCV value. The experimental results in [22] show that the quality of restored images by the proposed method with no prior knowledge of the original image is quite well.

In our problem, $\ell 1$-norm is used in the data-fitting term. In the proposed model (1.9), we make use of a new variable to measure the data-fitting term in $\ell 2$-norm, and therefore a deblurring problem in the form of Tikhonov regularization is required to solve in Step 1 of the proposed algorithm. We can adapt the GCV technique into this step to determine a suitable regularization parameter. The image restoration results can be selected with the smallest GCV value in the whole iterative process of the proposed algorithm. However, the additive noise does not satisfy the Gaussian assumption. In the next step, we will test this approach to evaluate the performance of such GCV-based $\ell 1$-TV algorithm for image restoration problems with non-Gaussian additive noise.

REFERENCES

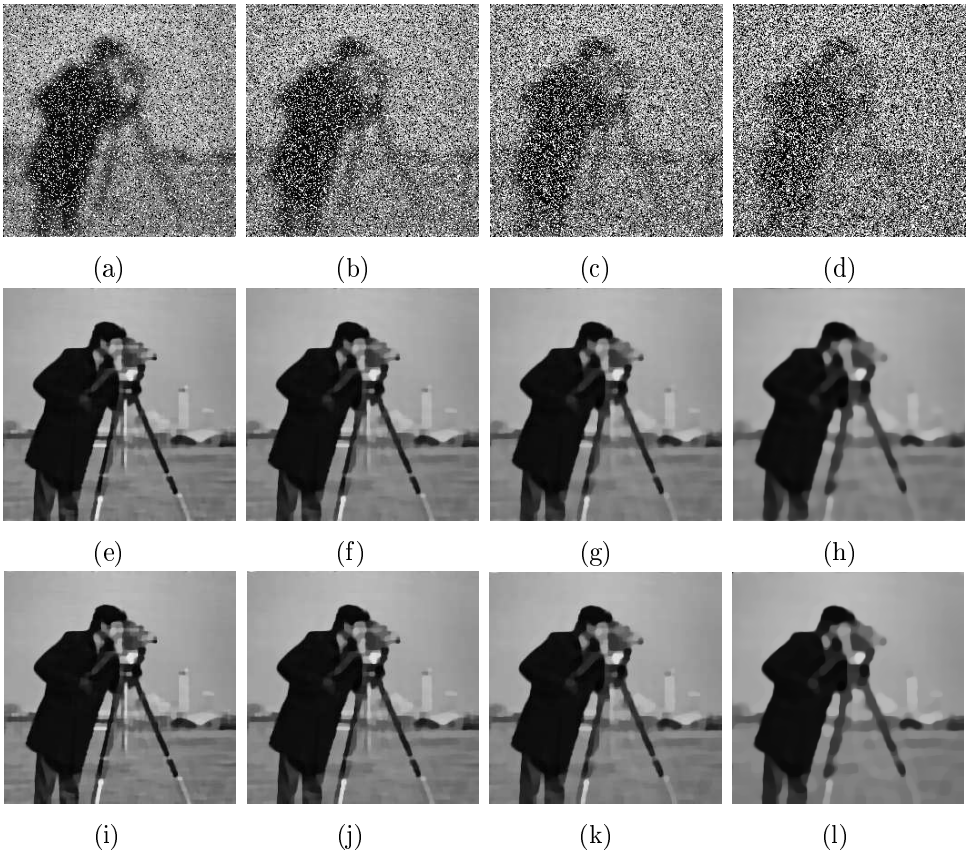


FIG. 4.12. *(a)-(d) Images with Gaussian blur of size* 15×15 *with standard derviation = 5 and salt-and-pepper noise from 30% to 60%; (e)-(h) restored images by the FTVd method; (i)-(l) restored images by the proposed method.*

[1] N. Abdelmalek and N. Otsu, Restoration of images with missing high-frequency components by minimizing the $\ell 1$ norm of the solution vector, *Appl. Opt.*, V24 (1985), pp. 1415-1420.

[2] S. Alliney and S. Ruzinsky, An algorithm for the minimization of mixed $\ell 1$ and $\ell 2$ norms with application to Bayesian estimation, *IEEE Trans. Signal Process.*, V42 (1994), pp. 618-627.

[3] J. E. Aujol, G. Gilboa, T. Chan and S. Osher, Structure-Texture Image Decomposition–Modeling, Algorithms, and Parameter Selection, *International Journal of Computer Vision.*, V67 (2006), pp. 111-136.

[4] D. Bertsekas, A. Nedic and E. Ozdaglar, Convex analysis and optimization, *Athena Scientific*, 2003.

[5] P. Bloomfield and W. Steiger, Least absolute deviations:Theory, Applications and Algorithms, *Birkhäuser, Boston.*, 1983.

[6] F. E. Browder and W. V. Petryshyn, The solution by iteration of nonlinear functional equations in Banach spaces, *Bull. Amer. Math. Soc.*, V72 (1966), pp. 571-575.

[7] A. Chambolle, An algorithm for total variation minimization and applications, *J. Math. Imaging Vision.*, V20 (2004), pp. 89-97.

[8] A. Chambolle, Total variation minimization and a class of binary MRF models, Energy Minimization Methods in Computer Vision and Pattern Recognition, *Lecture Notes in Computer Science, Springer Berlin.*, V3757 (2005), pp. 136-152.

[9] A. Chambolle, R. A. De Vore, N. Lee, and B. J. Lucier, Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage, *IEEE Trans. Image Process.*, V7 (1998), pp. 319-335.

[10] A. Chambolle and P. L. Lions, Image recovery via total variation minimization an related

TABLE 4.5
The summary of the restoration results of the proposed method and the FTVd method.

Gaussian blur	Noise level	Method	SNR (dB)	PSNR (dB)	Relative error	Time (seconds)
7x7	30%	FTVd	14.5	26.7	0.0876	18.4
		Proposed	14.7	27.0	0.0852	14.8
7x7	40%	FTVd	13.4	25.7	0.0989	22.9
		Proposed	13.5	25.8	0.0978	16.2
7x7	50%	FTVd	12.6	24.8	0.1090	22.2
		Proposed	12.7	24.9	0.1083	17.5
7x7	60%	FTVd	11.2	23.5	0.1277	23.0
		Proposed	11.3	23.5	0.1268	21.1
15x15	30%	FTVd	12.1	24.3	0.1154	19.7
		Proposed	12.1	24.4	0.1151	15.5
15x15	40%	FTVd	11.5	23.7	0.1238	24.0
		Proposed	11.5	23.7	0.1242	16.8
15x15	50%	FTVd	10.8	23.0	0.1343	21.0
		Proposed	10.8	23.0	0.1343	16.6
15x15	60%	FTVd	9.6	21.8	0.1542	19.1
		Proposed	9.6	21.9	0.1536	22.1

problems, *Numer. Math.*, V76 (1997), pp. 167-188.

[11] R. Chan, T. Chan, L. Shen and Z. Shen, Wavelet deblurring algorithms for spatially varying blur from high-resolution image reconstruction, *Linear Alg. Appl.*, V366 (2003), pp. 139-155.

[12] T. Chan and K. Chen, An optimization based total variation image denoising, *SIAM J. Multi. Model. Simul.*, V5 (2006), pp. 615-645.

[13] T. Chan and S. Esedoglu, Aspects of total variation regularized $\ell 1$ function approximation, *Technical report, university of california at los angeles.*, 2004.

[14] T. Chan and J. Shen, *Image Processing and Analysis - Variational, PDE, Wavelet, and Stochastic Methods*, SIAM Publisher, Philadelphia, 2005.

[15] P. L. Combettes and V. R. Wajs, Signal recovery by proximal forward-backward splitting, *SIAM J. Multi. Model. Simul.*, V4 (2005), pp. 1168-1200.

[16] I. Daubechies and G. Teschke, Variational image restoration by means of wavelets: simultaneous decomposition, deblurring and denoising, *Appl. Comput. Harmon. Anal.*, V19 (2005), pp. 1-16.

[17] A. Guitton and D. J. Verschuur, Adaptive subtraction of multiples using the $\ell 1$-norm, *Geophys. Prospecting.*, V52 (2004), pp. 1-27.

[18] H. Y. Fu, M. Ng, M. Nikolova, and J. L. Barlow Efficient minimization methods of mixed $\ell 2 - \ell 1$ and $\ell 1 - \ell 1$ norms for image restoration, *SIAM J. Sci. Comput.*, V4 (2005), pp. 1168-1200.

[19] D. Krishnan, P. Lin and A. Yip, A primal-dual active-set method for non-negativity constrained total variation deblurring problems, *IEEE Transactions on Image Processing*, V16 (2007), pp. 2766-2777.

[20] T. Kärkkäinen , K. Kunisch, and K. Majava Denoising of smooth images using L^1-fitting, *Computing.* V74 (2005), pp. 353-376.

[21] S. Kuo and R. Mammone, Image restoration by convex projections using adaptive contraints and the $\ell 1$ norm, *IEEE Trans. Signal Process.* V40 (1992), pp. 1159-1168.

[22] H. Liao, F. Li and M. Ng, Generalized cross-validation for total variation image restoration, *ICM Research Report 08-11* (http://www.math.hkbu.edu.hk/ICM/pdf/08-11.pdf), Institute for Computational Mathematics, Hong Kong Baptist University.

[23] M. Ng, Iterative methods for Toeplitz systems, *Oxford University Press*, 2004.

[24] M. Ng, R. Chan and W. Tang, A fast algorithm for deblurring models with Neumann boundary conditions, *SIAM J. Sci. Comput.*, V21 (1999), pp. 851-866.

[25] M. Ng, L. Oi, Y. Yang and Y. Huang, On semismooth Newton's methods for totel variation

minimization, *J. Math. Imaging Vision.*, V27 (2007), pp. 265-276.
[26] T. Miyashita, Super-resolved image restoration of holographic image by $\ell 1$-norm minization with clutter rejection, *Acoustical Imaging.*, V19 (1992), pp. 77-82.
[27] M. Nikolova, Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers, *SIAM J. Numer. Anal.*, V40 (2002), pp. 965-994.
[28] M. Nikolova, A variational approach to remove outliers and impulse noise, *J. Math. Imaging Vision*, V20 (2004), pp. 99-120.
[29] Z. Opial, Weak convergence of the sequence of successive approximations for nonexpansive mapping, *Bullet. Amer. Math. Soc.*, V73 (1967), pp. 591-597.
[30] L. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica D*, V60 (1992), pp. 259-268.
[31] A. Tikhonov and V. Arsenin, Solution of ill-poised problems, *Winston, Washington, DC*, 1977.
[32] Y. Wang, J. Yang, W. Yin and Y. Zhang, A new alternating minimization algorithm for total variation image reconstruction, *SIAM J. Imaging Sciences*, V1 (2008), pp. 248-272.
[33] J. Yang, W. Yin, Y. Zhang and Y. Wang, A fast algorithm for edge-preserving variational multichannel image restoration, Tech. Report 08-09, CAAM, Rice University.
[34] J. Yang, Y. Zhang and W. Yin, An efficient TVL1 algorithm for deblurring multichannel images corrupted by impulsive noise, http://www.caam.rice.edu/ optimization/L1/ftvd/.
[35] W. Yin, D. Goldfarb and S. Osher, The totel variation regularized L_1 model for multiscale decomposition, *SIAM J. Multi. Model. Simul.*, V6 (2006), pp. 190-211.