INTERNATIONAL JOURNAL OF
NUMERICAL ANALYSIS AND MODELING
Volume 14, Number 2, Pages 255–282

HIGH-RESOLUTION IMAGE RECONSTRUCTION: AN env_{ℓ^1}/TV MODEL AND A FIXED-POINT PROXIMITY ALGORITHM

WENTING LONG, YAO LU, LIXIN SHEN, AND YUESHENG XU

Abstract. High-resolution image reconstruction obtains one high-resolution image from multiple low-resolution, shifted, degraded samples of a true scene. This is a typical ill-posed problem and optimization models such as the ℓ^2/TV model are previously studied for solving this problem. It is based on the assumption that during acquisition digital images are polluted by Gaussian noise. In this work, we propose a new optimization model arising from the statistical assumption for mixed Gaussian and impulse noises, which leads us to choose the Moreau envelop of the ℓ^1-norm as the fidelity term. The developed env_{ℓ^1}/TV model is effective to suppress mixed noises, combining the advantages of the ℓ^1/TV and the ℓ^2/TV models. Furthermore, a fixed-point proximity algorithm is developed for solving the proposed optimization model and convergence analysis is provided. An adaptive parameter choice strategy for the developed algorithm is also proposed for fast convergence. The experimental results confirm the superiority of the proposed model compared to the previous ℓ^2/TV model besides the robustness and effectiveness of the derived algorithm.

Key words. High-resolution image reconstruction, env_{ℓ^1}/TV model, proximity operator, fixed-point algorithm.

1. Introduction

High-resolution (HR) image reconstruction arises from many applications, such as remote sensing, surveillance, and medical imaging. HR image provides high pixel density and rich image details which are desired in many image-related areas. For example, HR tomography images can help doctors to diagnose nidus at an early stage, and HR remote sensing images contain rich spectrum and spatial information, such as texture and shapes. However, due to the limitation of solid-state sensors such as charge-coupled devices (CCD) and complementary metal oxide semiconductors (CMOS), the way of improving hardware resolution is expensive, and sometimes hardly meets the need of better image details and greater image clarity. HR image reconstruction is an approach that reconstructs one HR image from multiple shifted, degraded low-resolution (LR) images. It can break through the resolution limit of sensor manufacturing techniques and reduce the cost of high precision optics and imaging sensors based on existing imaging equipments.

The image degradation procedure involves blurring, downsampling, displacement-error and noise. The observation model can be written as [25]

$$c = Au + \eta, \tag{1}$$

where c is an observed LR image, η is the noise, u denotes the desired HR image, and A is the degradation system matrix. The observation model (1) can be specified with the estimated registration parameters and the given boundary condition. In this paper we focus on a specific HR image reconstruction problem based on a mathematical model for a prefabricated multi-sensors image acquisition system proposed in [9], which will be explained in section 2. This model is feasible to acquire subpixel image information by placing sensors in coupled shifted subpixel positions.

Received by the editors on January 6 ,2016 and accepted on October 14, 2016.
2000 *Mathematics Subject Classification.* 35R35, 49J40, 60G40.

The HR image reconstruction is an ill-posed problem due to the compactness of the degradation operators [45]. Regularization methods are an effective way to obtain a stable approximate solution. Regularization methods for HR reconstruction can be roughly divided into stochastic and deterministic approaches. Stochastic HR approach is based on statistical modeling of noise and image degradation process [51, 52, 53, 54]. Previous studies showed that the stochastic approach has major advantage in robustness. The deterministic approach exploits the prior assumption of images and reconstructs HR images via an optimization framework [1, 3, 42]. However, if there is no or lack of prior information, it is difficult to develop a suitable optimization model.

In considering the image degradation factors, the noise is sometimes modeled as Gaussian white noise. Previously, the ℓ^2/TV model was applied to reconstruct HR images with the ℓ^2-norm as fidelity term describing the model error [43]. Meanwhile, in real applications, the imaging acquisition systems may suffer from impulse noise which is usually caused by malfunctioning arrays in camera sensors, faulty memory locations in hardware, or transmission in a noise channel [11, 60]. In this case, the ℓ^2/TV model may perform poorly because the ℓ^2-norm has less ability to reduce the effect of outliers compared to the ℓ^1-norm [10, 12, 42, 47, 57]. Moreover, the estimated displacement-errors should be specifically considered during the reconstruction via a regularization method.

The first contribution of this paper is to propose a robust optimization model arising from a statistical assumption for mixed Gaussian and impulse noise, denoted as the env_{ℓ^1}/TV model, by replacing the fidelity term in the ℓ^2/TV model with the Moreau envelop of the ℓ^1-norm. It is based on the observation that the Moreau envelop of the ℓ^1-norm balances the advantages of the ℓ^1-norm and the ℓ^2-norm. Some researchers tried to combine the ℓ^1-norm and the ℓ^2-norm for the fidelity term [50, 58, 59]. However, the choice strategy of the tradeoff parameter between these two norms is not clear and implementing a choice strategy is time consuming [59]. In this study, we consider the joint distribution of Gaussian white noise and impulse noise. Using segmental maximum a posteriori (MAP) criteria, we derive the Moreau envelop of the ℓ^1-norm to define the fidelity term. The Moreau envelop of the ℓ^1-norm has some desired properties. First, the Moreau envelop of the ℓ^1-norm is differentiable and its derivative has a closed form. Second, the parameter appearing in the Moreau envelop has statistical meaning, and is related to the Gaussian noise level and incident probability of impulse noise, which can be estimated from images iteratively. In our experiment we propose a choice strategy of τ adaptively according to the number of iterations. The numerical result shows that the performance of the proposed model is stable for mixed Gaussian and impulse noise.

The second contribution of this paper is to propose a fixed-point proximity algorithm to solve the model that avoids finding the inverse of A^TA. In this paper, we focus on solving the models regularized by the TV-norm. It is difficult to minimize by conventional methods due to the non-smoothness of the TV regularization term. A number of numerical methods have been proposed to address this issue, such as the primal-dual method [13, 20, 24, 35, 61], the alternating direction method of multipliers [23], interior point algorithms [30] and fixed-point methods [19, 33, 36]. These algorithms have been broadly used in solving image deblurring problem and were reported with great performance. Most of the algorithms mentioned above require computing the inverse of A^TA. When facing the HR image reconstruction with the displacement-error problem, the fact that the size of matrix A is large and there is no fast algorithm (like FFT and DCT) to efficiently compute the inverse of A^TA would increase the computational cost. In this paper, we propose a new fixed-point algorithm that avoids computing the inverse of $A^T A$, which will substantially reduce the computational cost.

We conduct the convergence analysis of the proposed algorithm in a generalized setting first and then apply it to the HR reconstruction problem.

This paper is organized as follows: In section 2, we review the Bose-Boo mathematical formulation for HR resolution. Based on the analysis of the statistical model of mixed Gaussian and impulse noise, the env_{ℓ^1}/TV optimization model for HR reconstruction is derived via segmental MAP criteria in section 3. In section 4, we prove that the env_{ℓ^1}/TV optimization model is equivalent to solving a fixed-point operator equation, and an iterative splitting algorithm for solving the fixed-point equation is developed. Section 5 is devoted to the convergence analysis of the proposed algorithm. In sections 6 and 7, we present a discussion of the parameter estimates, algorithm modification, along with the experiments for HR image reconstruction under various noise levels and displacement-error levels, while discussion is given in the meantime. In section 8 we draw a conclusion for this study.

2. The Bose-Boo Mathematical Formulation

In this paper we study HR image reconstruction based on the Bose-Boo mathematical model, an observation model for the prefabricated multi-sensor image acquisition system proposed in [9].

The Bose-Boo model considers an image-formation system composed of a set of identical CCD sensor arrays to obtain multiple observed images. It splits the incoming light from the "taking lens" and then passes through the relay lenses before projecting onto the set of CCD sensor arrays. We locate the CCD sensor arrays in preset positions to obtain multiple down-sampled images, from which we reconstruct a desired high resolution image. The advantage of this model is that, by sampling a scene at shifted sub-pixel positions with multiple LR sensors, we can obtain sub-pixel image information to reconstruct a HR image without increasing the cost of devices.

We now describe in details the Bose-Boo model. We consider an image acquisition system with an $l_1 \times l_2$ sensor array in which each sensor has $N_1 \times N_2$ sensing units and the size of each sensing unit is $T_1 \times T_2$. For simplicity, without loss of generality, in the following discussion we work on the case when $l_1 = l_2 := L$ and L is an even number. Our aim is to reconstruct an image with resolution $O_1 \times O_2$ where $O_1 := L \times N_1$ and $O_2 := L \times N_2$. Let $\mathbf{u}$ be the original continuous scene and $\mathbf{C}$ be the digital data acquired by the LR sensors. For the $[p,q]$th sensor in the sensor array, the average intensity detected by the $[n_1,n_2]$-th unit is modeled by

$$\mathbf{C}_{p,q}[n_1,n_2] := \frac{1}{T_1T_2}\int_{T_1(n_1-\frac{1}{2})+d^x_{p,q}}^{T_1(n_1+\frac{1}{2})+d^x_{p,q}} \int_{T_2(n_2-\frac{1}{2})+d^y_{p,q}}^{T_2(n_2+\frac{1}{2})+d^y_{p,q}} \mathbf{u}(x,y)\,dy\,dx + \eta_{p,q}[n_1,n_2], \tag{2}$$

where $n_1 = 1,2,\dots,N_1, n_2 = 1,2,\dots,N_2, p,q = 1,2,\dots,L$, the $\eta_{p,q}[n_1,n_2]$ is noise and $d^x_{p,q}$ and $d^y_{p,q}$ are displacement distances of the $[p,q]$-th sensor in the x and y directions, respectively, which are defined as follows. The sensors are shifted from each other by $(\frac{T_1}{L}, \frac{T_2}{L})$. With respect to the $[0,0]$-th reference sensor, the $[p,q]$-th sensor has vertical and horizontal displacement-errors $\varepsilon^x_{p,q}$ and $\varepsilon^y_{p,q}$. We assume that

$$\left|\varepsilon^x_{p,q}\right| < \frac{1}{2} \quad \text{and} \quad \left|\varepsilon^y_{p,q}\right| < \frac{1}{2}$$

to avoid overlapping during the image capture process. Then the total displacement of the $[p,q]$-th sensor with respect to the $[0,0]$-th reference sensor in vertical and horizontal directions are, respectively,

$$d_{p,q}^x := \left(p + \varepsilon_{p,q}^x\right)\frac{T_1}{L} \quad \text{and} \quad d_{p,q}^y := \left(q + \varepsilon_{p,q}^y\right)\frac{T_2}{L}.$$

We combine all the LR images together to form an $O_1 \times O_2$ image $\mathbf{C}$ by assigning

$$\mathbf{C}\left[Ln_1 + p, Ln_2 + q\right] := \mathbf{C}_{p,q}\left[n_1, n_2\right].$$

This is the *observed* image. At the same time, The *ideal* HR image is modeled by

$$\mathbf{U}_{p,q}[n_1, n_2] := \frac{L^2}{T_1 T_2}\int_{\frac{T_1}{L}(n_1-\frac{1}{2})}^{\frac{T_1}{L}(n_1+\frac{1}{2})}\int_{\frac{T_2}{L}(n_2-\frac{1}{2})}^{\frac{T_2}{L}(n_2+\frac{1}{2})} \mathbf{u}(x,y)\, dy\, dx.$$

The purpose of the HR reconstruction is to restore the ideal HR image $\mathbf{U}$ from the observed image $\mathbf{C}$.

The blurring matrix corresponding to the $[p,q]$-th sensor modeled in (2) is given by

$$A_L\left(\varepsilon_{p,q}^x, \varepsilon_{p,q}^y\right) := A_L\left(\varepsilon_{p,q}^y\right)^T \otimes A_L\left(\varepsilon_{p,q}^x\right),$$

where $\otimes$ denotes the Kronecker tensor product defined by $\mathbf{B}\otimes\mathbf{D} := [b_{m,n}\mathbf{D}]$ with $\mathbf{B} := [b_{m,n}]$. $A_L\left(\varepsilon_{p,q}^x, \varepsilon_{p,q}^y\right)$ is explored to denote the blurring matrix of Bose-Boo model corresponding to the $[p,q]$-th sensor. When $L = 2$, the sensor's place with and without displacement error can be found in Figure 1 [16]. And the system matrices $A_L\left(\varepsilon_{p,q}^y\right)$ and $A_L\left(\varepsilon_{p,q}^x\right)$ vary under different boundary conditions. In this paper we consider the Neumann boundary condition. The sampling matrix is given as

$$D_{p,q} := J_q \otimes J_p$$

where $J_p := I_{N_1} \otimes (e_p e_p^T)$, $J_q := I_{N_2} \otimes (e_q e_q^T)$, and e_p and e_q are the p-th and q-th columns of the $L \times L$ identity matrix, respectively. In conclusion, the blurring matrix of Bose-Boo model considering all the sensors can be defined as

$$A_L\left(\varepsilon^x, \varepsilon^y\right) := \frac{1}{L \times L}\sum_{p,q=1}^{L} D_{p,q} A_L\left(\varepsilon_{p,q}^x, \varepsilon_{p,q}^y\right).$$

For the computational purpose, we reformulate the matrix as a column-wise vector. Let $m := O_1 O_2$, $u \in \mathbb{R}^m$ be the column-wise vector form of $\mathbf{U}$ and $c \in \mathbb{R}^m$ be the column-wise vector form of $\mathbf{C}$. Then the discrete model of the HR reconstruction is

$$c = A_L\left(\varepsilon^x, \varepsilon^y\right)u + \eta, \tag{3}$$

where η is the noise. We remark that the system matrix $A_L\left(\varepsilon^x, \varepsilon^y\right)$ is ill-conditioned and solving linear equation (3) is a typical ill-posed problem.

3. An Optimization Model for Mixed Noise

Traditional HR image reconstruction employs statistical models via some prior knowledge on noises. The ℓ^2/TV model is usually applied to treat the Gaussian white noise. However, impulse noise and outliers are common in practical applications, which are typically caused by malfunctioning arrays in camera sensors, faulty memory locations in hardware, or transmission in a noisy channel [11, 60]. Meanwhile, the Bose-Boo mathematical model contains the modeling error caused by small perturbation around the ideal subpixel locations of the sensor units. The perturbation built in the imaging system tends to be random, uncorrelated and is difficult to be calibrated by the system itself. Thus, the Bose-Boo model is sensitive to noise and large outliers [16]. This drives us to find a robust optimization model to overcome this shortcoming.

In this section, we propose a new optimization model for HR image reconstruction with the mixed Gaussian and impulse noises based on the assumption that these two types of noise are independent and each pixel has independent identical distribution. We explore the segmental MAP criterion [2, 29, 32] which maximizes the joint posterior probability density function (PDF).

We assume that the blurred image is contaminated by Gaussian noise first and then by impulse noise. Let M_1, M_2 and M_3 be random variables representing the original image, underlying blurred image with Gaussian noise and the observed image with both Gaussian noise and impulse noise, respectively. It can be naturally captured in the Hidden Markov Model (HMM) with two states. We recall that u is the column-wise vector form of the HR image under consideration and c is the column-wise vector form of the compact LR images. We denote by ξ the underlying medium state that the blurred image Au is contaminated by Gaussian noise. Let $f_{1,2|3}(u,\xi|c)$ be the joint conditional PDF of random variables M_1 and M_2 giving that $M_3 = c$. Then the segmental MAP criteria proposes to find a maximizer u^* such that

$$u^* = \arg\max_u \max_\xi \{f_{1,2|3}(u,\xi|c)\}. \tag{4}$$

By the Bayesian theorem, the model (4) is equivalent to

$$u^* = \arg\max_u \left\{ \max_\xi \{\ln f_{3,2|1}(c,\xi|u)\} + \ln f_1(u) \right\}, \tag{5}$$

where $f_{3,2|1}(c,\xi|u)$ is the joint conditional PDF of random variables M_3 and M_2 giving the random variable $M_1 = u$ and $f_1(u)$ is the marginal PDF of random variable $M_1 = u$. The PDF $f_1(u)$ embodies the prior information of the HR image and constrains the solution in a small space. By the assumption, Gaussian noise and impulse noise contamination processes are independent. Therefore, random variables $X_1 := M_1$, $X_2 := M_2 - AM_1$ and $X_3 := M_3 - M_2$ are independent. By the random variable transformation, we have that the Jacobian determinant

$$J = \begin{vmatrix} \frac{\partial X_1}{\partial M_1} & \frac{\partial X_1}{\partial M_2} & \frac{\partial X_1}{\partial M_3} \\ \frac{\partial X_2}{\partial M_1} & \frac{\partial X_2}{\partial M_2} & \frac{\partial X_2}{\partial M_3} \\ \frac{\partial X_3}{\partial M_1} & \frac{\partial X_3}{\partial M_2} & \frac{\partial X_3}{\partial M_3} \end{vmatrix},$$

with $|J| = 1$. Let $f_{1,2,3}(u,\xi,c)$ be the joint PDF of M_1, M_2 and M_3. Then $f_{X_1,X_2,X_3}(u, \xi - Au, c - \xi)$ is the joint PDF of X_1, X_2 and X_3. At the same time, $f_{X_1}(u)$, $f_{X_2}(\xi - Au)$ and $f_{X_3}(c - \xi)$ are the marginal PDF of X_1, X_2 and X_3, respectively. We have that

$$f_{3,2|1}(c,\xi|u) = \frac{f_{1,2,3}(u,\xi,c)}{f_1(u)} = \frac{f_{X_1,X_2,X_3}(u,\xi - Au, c-\xi)|J|}{f_{X_1}(u)} = f_{X_2}(\xi - Au) f_{X_3}(c - \xi). \tag{6}$$

We suppose that Gaussian noise distribution is of zero mean and variance σ^2 (σ is the standard deviation). Then there exists a constant k related to σ^2 such that

$$f_{X_2}(x) = k \exp\left[-\frac{1}{2\sigma^2} \|x\|_2^2 \right]. \tag{7}$$

For impulse noise modeled by the binomial distribution, there exists an incident probability $r \in [0,1]$ such that for $x \in \mathbb{R}^d$,

$$f_{X_3}(x) = \left(\frac{r}{2}\right)^{\|x\|_0} (1-r)^{d - \|x\|_0}. \tag{8}$$

Substituting (7) and (8) into (6) yields that

$$f_{3,2|1}(c,\xi|u) = k\exp\left[-\frac{1}{2\sigma^2}\|Au-\xi\|_2^2\right]\left(\frac{r}{2}\right)^{\|c-\xi\|_0}(1-r)^{d-\|c-\xi\|_0}.$$

A simple computation shows that $\max\limits_{\xi}\big\{\ln f_{3,2|1}(c,\xi|u)\big\}$ is equivalent to

$$\min_{\xi}\left\{\frac{1}{2\tau}\|Au-\xi\|_2^2+\|c-\xi\|_0\right\}, \tag{9}$$

where τ is a constant related to σ^2 and r. Because of the non-convexity of the ℓ^0-norm, we relax the ℓ^0-norm to ℓ^1-norm according to [12]. Therefore, problem (9) is twisted to the following optimization problem

$$\min_{\xi}\left\{\frac{1}{2\tau}\|Au-\xi\|_2^2+\|c-\xi\|_1\right\}. \tag{10}$$

We recall the definition of the *Moreau envelop* of function φ [4]. Let Γ_0 be defined as the class of all lower semicontinuous convex functions $f\colon \mathbb{R}^d \to (-\infty,+\infty]$ such that $\mathrm{dom}_f := \{x\in\mathbb{R}^d : f(x) < +\infty\} \neq \emptyset$. Suppose that $\varphi\in\Gamma_0(\mathbb{R}^d), \varphi : \mathbb{R}^d \to (-\infty,+\infty]$ and ν is a positive number. The Moreau envelop of φ with respect to parameter ν is defined by

$$\mathrm{env}_{\nu\varphi}(x) := \min_{y\in\mathbb{R}^d}\left\{\frac{1}{2\nu}\|x-y\|_2^2+\varphi(y)\right\}, \quad x\in\mathbb{R}^d.$$

The effect of the φ and the ℓ^2-norm is balanced and adjusted by parameter ν. From the definition of the Moreau envelop, the cost function in (10) is $-\mathrm{env}_{\tau\|\cdot\|_1}(\cdot)$. Meanwhile, we assume that images to be studied are piecewise continuous and contain sharp features. Hence, the marginal PDF of u can be described by the total variation norm, that is,

$$f_1(u) := k_2\exp[-\mu\|u\|_{\mathrm{TV}}],$$

where k_2 is a constant, μ is a regularization parameter and $\|u\|_{\mathrm{TV}}$ is the total variation norm [49]. Let $y := \xi - c$. By using the notion of $\mathrm{env}_{\nu\varphi}$, the optimization model (5) is rewritten as

$$u^* = \arg\min_{u\in\mathbb{R}^d}\left\{\mathrm{env}_{\tau\|\cdot\|_1}(Au-c)+\mu\|u\|_{\mathrm{TV}}\right\}. \tag{11}$$

We recall some properties of the Moreau envelop of the ℓ^1-norm. In one dimension, the Moreau envelop of the ℓ^1-norm reduces to

$$\mathrm{env}_{\tau|\cdot|}(x) = \begin{cases} |x|-\frac{\tau}{2}, & |x|>\tau, \\ \frac{1}{2\tau}x^2, & |x|\le\tau, \end{cases} \quad x\in\mathbb{R}.$$

We plot the figure of the ℓ^1-norm, the ℓ^2-norm and $\mathrm{env}_{\tau|\cdot|}(\cdot)$ with different choices of τ in Figure 1. From Figure 1, it is easy to see that $\mathrm{env}_{\tau|\cdot|}(\cdot)$ behaves like the ℓ^1-norm for large x and behaves like a quadratic function for small x. In addition, the Moreau envelop of function $\varphi\in\Gamma_0$ is continuously differentiable, and its gradient is formulated by

$$\nabla\mathrm{env}_{\tau\varphi}(x) = \frac{1}{\tau}\big(I-\mathrm{prox}_{\tau\varphi}(x)\big), \quad x\in\mathbb{R}^d.$$

We remark that to increase the robustness in dealing with large errors, some other maximum likelihood-type estimators (M-estimators) [7, 27] have been explored in multi-frame super-resolution image reconstruction [21, 28, 31, 37, 38]. However, these M-estimators are difficult to be formulated as convex optimization problems. Besides, the Huber loss function [26] was used in [46] as the fidelity term to obtain stable solutions and reject outliers [8]. Let $\tau > 0$. The Huber function $H_\tau : \mathbb{R} \to \mathbb{R}$ is defined as:

$$H_\tau(x) : = \begin{cases} \tau\left(|x| - \frac{\tau}{2}\right), & |x| > \tau, \\ \frac{1}{2}x^2, & |x| \le \tau. \end{cases}$$

Comparing it with the definition of $\text{env}_{\tau|\cdot|}$, we have that $H_\tau(x) = \tau \cdot \text{env}_{\tau|\cdot|}(x)$. If $\tau > 1$ which is the case in most image processing problems, the slope of $H_\tau(\cdot)$ is τ for large x. This implies that $H_\tau(\cdot)$ is of no use in reducing the outliers. In Figure 1, we plot the figures of the Huber loss function and $\text{env}_{\tau|\cdot|}$ with $\tau = 1.5$ for comparison.

We constrain u in $\Omega := \{x : x \in \mathbb{R}^d, 0 \le x_i \le 255, i \in \mathbb{N}^d\}$ which is related to the box constrain for visual presentation and 8-bit unsign integer storage in digital media [5, 17] and get a constrained convex optimization model

$$u^* = \arg\min_{u\in\Omega}\left\{\text{env}_{\tau\|\cdot\|_1}(Au - c) + \mu\|u\|_{\text{TV}}\right\}. \tag{12}$$

The new model is called by the env_{ℓ^1}/TV model. When assuming that the image is contaminated by impulse noise first and then by Gaussian noise, we obtain the same model because the Moreau envelop is a special infimal convolution [4]. To convert the constrained optimization problem (12) into a non-constrained optimization problem, we exploit the *indicator function*. The indicator function on the closed convex set Ω is defined as

$$\iota_\Omega(x) := \begin{cases} 0, & \text{if } x \in \Omega, \\ +\infty, & \text{otherwise.} \end{cases}$$

By introducing function $\iota_\Omega(\cdot)$, the optimization model (12) is equivalent to

$$u^* = \arg\min\left\{\text{env}_{\tau\|\cdot\|_1}(Au - c) + \mu\|u\|_{\text{TV}} + \iota_\Omega(u)\right\}. \tag{13}$$

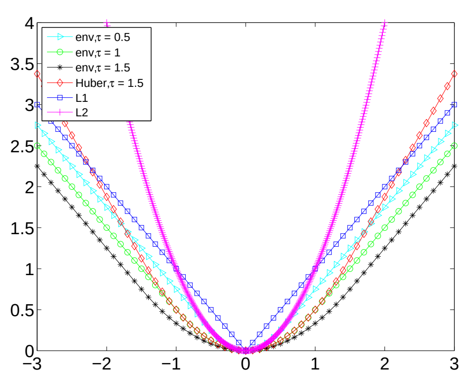

FIGURE 1. The plots of the ℓ^1-norm, the ℓ^2-nom, the Huber loss function and $\text{env}_{\tau\|\cdot\|_1}(\cdot)$ in one dimension.

4. Fixed-point Proximity Algorithm

In this section we propose a fixed-point proximity algorithm for a general convex optimization problem which takes problem (13) as a special case. The proposed algorithm is based on a system of fixed-point equations derived from the property of the solution of the above optimization problem via the proximity operator.

We now describe the optimization problem studied in this section. Suppose that $A \in \mathbb{R}^{d_1\times d}$ and $B \in \mathbb{R}^{d_2\times d}$. Suppose that $p_j \in \Gamma_0(\mathbb{R}^{d_j})$, for $j = 1, 2$, and $p_3 = \iota_w$ where w is a closed convex set in $\mathbb{R}^d$. For a vector $u \in \mathbb{R}^d$, the general optimization model that we study in this section is to find a minimizer $u^* \in \mathbb{R}^d$ that satisfies

$$u^* = \arg\min\{p_1(Au) + p_2(Bu) + p_3(u)\}, \tag{14}$$

where $p_1(A\cdot)$ is differentiable with a β-Lipschitz continuous gradient. When $p_2(B\cdot)$ is the TV-norm, $p_3 := \iota_w$ and $p_1 := \frac{1}{2}\|\cdot -c\|_2^2$, it reduces to the conventional ℓ^2/TV deblurring model. Likewise, when $p_2(B\cdot)$ is the TV-norm, $p_3 := \iota_w$ and $p_1 := \mathrm{env}_{\tau\|\cdot\|_1}(\cdot - c)$, it is the proposed env_{ℓ^1}/TV model.

Based on Fermat's rule, to solve the optimization model, we need to find the gradient of the objective function $\mathcal{P}(u) := p_1(Au) + p_2(Bu) + p_3(u)$ at point u^*. However, function p_2 is non-differentiable, the *sub-differential* operator is applied to overcome this difficulty. Let $f : \mathbb{R}^d \to (-\infty, +\infty]$ be proper. The sub-differential of f at given $x \in \mathbb{R}^d$ is a set defined by

$$\partial f(x) := \{y : y \in \mathbb{R}^d, f(z) \geq f(x) + \langle y, z - x\rangle, \text{ for all } z \in \mathbb{R}^d\}.$$

For a differentiable function f, the sub-differential of f at point x is a singleton set [4], that is, $\partial f(x) = \{\nabla f(x)\}$, where ∇ denotes the gradient operator.

We further introduce the *proximity* operator. For a symmetric positive definite matrix $H \in \mathbb{R}^{d\times d}$, we define the *weighted inner product* as

$$\langle x, y\rangle_H := \langle x, Hy\rangle \quad x, y \in \mathbb{R}^d,$$

and the induced *weighted norm* is accordingly defined by $\|x\|_H^2 := \langle x, x\rangle_H$. For a given point $x \in \mathbb{R}^d$, the *proximity operator* of f with respect to H is defined as

$$\mathrm{prox}_{f,H}(x) := \arg\min\left\{f(u) + \frac{1}{2}\|u - x\|_H^2, u \in \mathbb{R}^d\right\}.$$

Clearly, $\mathrm{prox}_{f,H}$ is a mapping on $\mathbb{R}^d$. In particular, when H is an identity matrix, we simplify $\mathrm{prox}_{f,H}$ as prox_f. The proximity operator plays a pivotal role in convex analysis [6, 44, 48]. Meanwhile, the sub-differential and the proximity operator can be mutually converted [35]. That is, for any $d \times d$ symmetric positive definite matrix H, $x \in \mathrm{dom}_f$ (the domain of f) and $y \in \mathbb{R}^d$,

$$Hy \in \partial f(x) \quad \text{if and only if} \quad x = \mathrm{prox}_{f,H}(x + y). \tag{15}$$

Meanwhile, we define a set $\mathbb{S}_+^d := \{S : S$ is a $d \times d$ diagonal matrix with positive diagonal entries$\}$.

With above preliminaries, the following theorem derives an important property of the solution of model (14).

Theorem 4.1. *Let $p_1 \in \Gamma_0(\mathbb{R}^{d_1})$, $p_2 \in \Gamma_0(\mathbb{R}^{d_2})$, $A \in \mathbb{R}^{d_1\times d}$, $B \in \mathbb{R}^{d_2\times d}$. If u is a solution of model (14) and function p_1 is differentiable, then for any $\lambda > 0$ and $S \in \mathbb{S}_+^d$, there exists $v \in \mathbb{R}^{d_2}$ such that the pair $(v, u) \in \mathbb{R}^{d_2} \times \mathbb{R}^d$ satisfies the following coupled fixed-point* *equations:*

$$v = (I - \text{prox}_{\lambda p_2})(v + Bu) \tag{16}$$

$$u = \text{prox}_{p_3, S^{-1}}\big(u - S(\nabla p_1(Au) + \frac{1}{\lambda}B^T v)\big). \tag{17}$$

Conversely, if $u \in \mathbb{R}^d$ *satisfies equations* (16)-(17) *for some* $v \in \mathbb{R}^d$*,* $\lambda > 0$ *and* $S \in \mathbb{S}^d_+$*, then it is a solution of model* (14)*.*

Proof. Let $u \in \mathbb{R}^d$ be a solution of model (1). According to Fermat's rule, Corollary 16.38 in [4] and the chain rule that $\partial(f \circ B) = B^T \circ \partial f \circ B$, the solution $u \in \mathbb{R}^d$ of model (14) needs to satisfy that

$$0 \in B^T \circ \partial p_2 \circ (Bu) + \partial(p_1 \circ A)(u) + \partial p_3(u). \tag{18}$$

Suppose that for any $\lambda > 0$, there exists a $v \in \mathbb{R}^{d_2}$ such that

$$v \in \lambda \partial p_2 \circ (Bu). \tag{19}$$

Then relation (18) can be rewritten as

$$0 \in \frac{1}{\lambda}B^T v + \partial(p_1 \circ A)(u) + \partial p_3(u). \tag{20}$$

Based on the assumption on the differentiablity of p_1, we have that $\partial(p_1 \circ A)(u) = \{\nabla p_1(Au)\}$. Multiplying both sides of relation (20) by $S \in \mathbb{S}^d_+$ yields that

$$-S\big(\nabla p_1(Au) + \frac{1}{\lambda}B^T v\big) \in S\partial p_3(u).$$

By employing property (15), we may rewrite relation (4.1) as

$$u = \text{prox}_{p_3, S^{-1}}\big(u - S(\nabla p_1(Au) + \frac{1}{\lambda}B^T v)\big).$$

Applying property (15) to (19), we obtain that

$$v = (I - \text{prox}_{\lambda p_2})(Bu + v).$$

We have proved that (v, u) satisfies the fixed-point equations (16)-(17).

Conversely, suppose that there exist $u \in \mathbb{R}^d$ and $v \in \mathbb{R}^{d_2}$, λ is a positive number and S is a symmetric positive definite square matrix. The necessary condition requires to prove that u is the solution of optimization problem (14) when equations (16)-(17) hold. To this end, we notice that equations (16), (19) are equivalent based on relation (15). Then by substituting relation (19) into (17) and employing property (15) again, we obtain that

$$0 \in B^T \circ \partial \circ p_2(Bu) + \nabla p_1(Au) + \partial p_3(u).$$

This ensures that u is a solution of model (14). □

Theorem 4.1 demonstrates that the solution of the minimization problem (14) can be obtained by solving two coupled fixed-point equations. The existence of a fixed-point (v, u) of the coupled equations (16)-(17) is a direct consequence of Theorem 4.1.

To develop a convergent iteration scheme, we first reformulate the system of fixed-point equations (16)-(17) into a compact form. Recall that for a proper function $f\colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, the conjugate function $f^*\colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is defined for $u \in \mathbb{R}^d$ by

$$f^*(u) := \sup_{x \in \mathbb{R}^d} \big\{\langle x, u\rangle - f(x)\big\}.$$

For a function $f \in \Gamma_0(\mathbb{R}^d)$ and a positive number λ, the proximity operators of f and its conjugate function f^* satisfy the following relationship (Theorem 14.3 (ii), [4]):

$$I = \mathrm{prox}_{\lambda f} + \lambda \mathrm{prox}_{\frac{1}{\lambda} f^*}\left(\frac{1}{\lambda} I\right).$$

Let $w := \frac{v}{\lambda}$. Then equation (16) can be rewritten as

$$w = \mathrm{prox}_{p_2^*, \lambda I}\big(\frac{1}{\lambda} Bu + w\big) \tag{21}$$

based on the definition of $\mathrm{prox}_{f,H}$. Combining equation (17) with (21), we get a system of fixed-point equations for vector (w, u):

$$\begin{cases} w = \mathrm{prox}_{p_2^*, \lambda I}(\frac{1}{\lambda} Bu + w), \\ u = \mathrm{prox}_{p_3, S^{-1}}\big(u - S(\nabla p_1(Au) + B^T w)\big). \end{cases} \tag{22}$$

Then we integrate two equations together by introducing a new function. Let $P := \lambda I, Q := S^{-1}, R := \mathrm{diag}(P, Q), V := \begin{bmatrix} w \\ u \end{bmatrix} \in \mathbb{R}^{d_2+d}$. We define convex function $\Phi(X) := p_2^*(x^1) + p_3(x^2) \colon \mathbb{R}^{d_2+d} \to \mathbb{R}$ for a vector $X := \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} \in \mathbb{R}^{d_2+d}$. By definition, a simple computation shows that the proximity operator of function Φ with respect to matrix R at vector V has the property that

$$\mathrm{prox}_{\Phi, R}(V) = \begin{bmatrix} \mathrm{prox}_{p_2^*, P}(w) \\ \mathrm{prox}_{p_3, Q}(u) \end{bmatrix}.$$

For the $d_2 \times d$ matrix B in model (14), we define matrix

$$E := \begin{bmatrix} I & \frac{1}{\lambda} B \\ -SB^T & I \end{bmatrix},$$

and for $d_1 \times d$ matrix A, we define

$$G(X) := \begin{bmatrix} 0 \\ \nabla p_1(Ax^2) \end{bmatrix}.$$

We remark that $G(X)$ is an operator on vector X and ∇p_1 may be nonlinear operator. For example, when $p_1 := \mathrm{env}_{\tau\|\cdot\|_1}(\cdot)$, $\nabla p_1 = \frac{(I - \mathrm{prox}_{\tau\|\cdot\|_1})(\cdot)}{\tau}$ is a nonlinear function. We also introduce $(d_2 + d) \times (d_2 + d)$ matrix

$$R := \begin{bmatrix} \lambda I & 0 \\ 0 & S^{-1} \end{bmatrix}.$$

With the help of the above definitions, equations (22) can be formed in a compact form

$$V = \mathrm{prox}_{\Phi, R}\left(EV - R^{-1} G(V)\right). \tag{23}$$

It is easy to see that finding a solution of model (14) is equivalent to computing a solution V of fixed-point equation (23). The existence and uniqueness of the solution of fixed-point equation (23) are guaranteed by Theorem 4.1. The next step is developing a convergent iteration scheme. Note that operator $\mathcal{Q} := E \cdot - R^{-1} G(\cdot)$ may be expanding. Motivated by our previous work [33], we split matrix E into two matrices. This yields that

$$V = \mathrm{prox}_{\Phi, R}\big((E - R^{-1}M)V + R^{-1}MV - R^{-1}G(V)\big),$$

where

$$M := \begin{bmatrix} \lambda I & B \\ B^T & S^{-1} \end{bmatrix}, \tag{24}$$

and E, R, G are given as above. Accordingly, an algorithm can be formulated as

$$V_{k+1} = \text{prox}_{\Phi,R}\big((E - R^{-1}M)V_{k+1} + R^{-1}MV_k - R^{-1}G(V_k)\big). \tag{25}$$

We unfold equation (25) to get a system of iteration equations

$$\begin{cases} v_{k+1} = (I - \text{prox}_{\lambda p_2})(Bu_k + v_k), \\ u_{k+1} = \text{prox}_{p_3,S^{-1}}\big(u_k - S(\nabla p_1(Au_k) + \frac{1}{\lambda}B^T(2v_{k+1} - v_k)\big). \end{cases}$$

This leads to a fixed-point algorithm based on the proximity operators and the pre-condition matrices described in Algorithm 1.

Algorithm 1 (Fixed-point algorithm based on the proximity and pre-condition operators.)
Initialization: $v_0 = 0, u_0 = 0, \lambda > 0$.
repeat
  Step 1: $v_{k+1} = (I - \text{prox}_{\lambda \text{p}_2})(Bu_k + v_k)$**,**
  Step 2: $u_{k+1} = \text{prox}_{p_3,S^{-1}}\Big(u_k - S\big(\nabla p_1(Au_k) + \frac{1}{\lambda}B^T(2v_{k+1} - v_k)\big)\Big)$**.**
until 'convergence'

This algorithm has advantages over most of the existing algorithms because it does not require computing the inverse of A^TA. As pointed out by a referee, a similar technique has been used in a different context [18, 56]. In addition, it explores pre-conditionor matrix S to further use the prior knowledge about A to accelerate the algorithm.

We remark that, for the non-constrained problem, $p_3 := \iota_{\mathbb{R}^d}$, $x \in \mathbb{R}^d$

$$\text{prox}_{\iota_{\mathbb{R}^d},S^{-1}}(x) = \arg\min_{u \in \mathbb{R}^d} \left\{ \frac{1}{2}\|u - x\|^2_{S^{-1}} \right\} = S^{-1}x$$

and the algorithm for solving this problem can be derived in the same way.

5. Convergence Analysis

In this section we analyze convergence of iteration (25).

We first introduce an operator related to the proximity operator. For each $X \in \mathbb{R}^{d_2+d}$, we let $Y \in \mathbb{R}^{d_2+d}$ satisfy that

$$Y = \text{prox}_{\Phi,R}\big((E - R^{-1}M)Y + R^{-1}MX - R^{-1}G(X)\big). \tag{26}$$

Noting that equation (26) defines a mapping from X to Y, we denote this mapping by NL_M. Thus, Y in equation (26) is well defined for a given X, with the solvability being guaranteed automatically. With the help of this new operator, we rewrite (25) as

$$V_{k+1} = \text{NL}_\text{M}(\text{V}_\text{k}).$$

For the convergence analysis of sequence $\mathcal{V} := \{V_k : k \in \mathbb{N}_0\}$, we shall establish a general theorem. We first present a lemma about the fixed-point of continuous operators.

Lemma 5.1. *Suppose that $F\colon \mathbb{R}^d \to \mathbb{R}^d$ is a continuous operator and H is an $d \times d$ symmetric positive definite matrix. Let $\mathfrak{v} := \{v_k\colon v_k \in \mathbb{R}^d, k \in \mathbb{N}\}$ be the sequence generated by iterative scheme $v_{k+1} = F(v_k)$ for any $v_0 \in \mathbb{R}^d$. If $\lim\limits_{k\to+\infty} \|v_{k+1} - v_k\|_H = 0$ and sequence $\mathfrak{v}$ has a cluster point $\tilde{v}$, then the cluster point is a fixed-point of operator F.*

Proof. Based on the assumption that $\mathfrak{v}$ has a cluster point $\tilde{v}$, there exist a subsequence $\{v_{k_n} : n \in \mathbb{N}\}$ that converges to the cluster point $\tilde{v} \in \mathbb{R}^d$. This together with the hypothesis that $\lim_{k\to+\infty} \|v_{k+1} - v_k\|_H = 0$ ensures

$$\lim_{n\to+\infty} F(v_{k_n}) = \lim_{n\to+\infty} v_{k_n+1} = \tilde{v}. \tag{27}$$

Meanwhile, since F is continuous, we observe that

$$\lim_{n\to+\infty} F(v_{k_n}) = F(\tilde{v}). \tag{28}$$

Comparing (27) and (28), we obtain that $F(\tilde{v}) = \tilde{v}$. This proves the desired result. □

We are now ready to prove the general theory about sequence convergence.

Theorem 5.2. *Suppose that $F\colon \mathbb{R}^d \to \mathbb{R}^d$ is a continuous operator and H is an $d \times d$ symmetric positive definite matrix. Let $\mathfrak{v} := \{v_k : v_k \in \mathbb{R}^d, k \in \mathbb{N}\}$ be the sequence generated by iterative scheme $v_{k+1} = F(v_k)$ for any $v_0 \in \mathbb{R}^d$ and C be the set of fixed-points of F. If the following conditions hold:*
(1) F is a continuous operator,
(2) for any $v \in C$, $\|v_{k+1} - v\|_H \le \|v_k - v\|_H$,
(3) $\|v_{k+1} - v_k\|_H \to 0$ as $k \to +\infty$,
then the sequence $\mathfrak{v}$ converges to a fixed-point of F.

Proof. With hypothesis (2), for any fixed-point v, we have that

$$\|v_k\|_H \le \|v_k - v\|_H + \|v\|_H \le \|v_0 - v\|_H + \|v\|_H.$$

Therefore, sequence $\mathfrak{v}$ is bounded. Hense, there exists a subsequence $\{v_{k_n}\}$ that converges to a cluster point $\tilde{v} \in \mathbb{R}^m$, that is,

$$\lim_{n\to+\infty} \|v_{k_n} - \tilde{v}\|_H = 0. \tag{29}$$

By Lemma 5.1, $\tilde{v}$ is a fixed-point of F.

On the other hand, from hypothesis (2) we also have that for any fixed-point v, $\|v_k - v\|_H$ is a bounded monotone decreasing sequence. Consequently,

$$\lim_{k\to+\infty} \|v_k - \tilde{v}\|_H = a \tag{30}$$

for some $a \ge 0$. Comparing equation (29) with (30), we conclude that $a = 0$, which proves the desired result. □

In the following, we prove convergence of sequence $\mathcal{V}$ by verifying the hypotheses of Theorem 5.2. We first prove the continuity of operator $\mathrm{NL_M}$, which is hypothesis (1) in Theorem 5.2.

We recall that the graph of function $p : \mathbb{R}^d \to \mathbb{R}^d$ is the set

$$\mathrm{gra}(p) := \big\{(x, y) \colon (x, y) \in \mathbb{R}^d \times \mathbb{R}^d, y = p(x)\big\}.$$

A mapping $J\colon \mathbb{R}^d \to \mathbb{R}^d$ is said to be firmly non-expansive, with respect to a symmetric positive definite matrix H. If for all $x, y \in \mathbb{R}^d$, it satisfies

$$\|Jx - Jy\|_H^2 \le \langle Jx - Jy, x - y\rangle_H.$$

Lemma 5.3. *Operator* $\mathrm{NL_M}$ *is continuous.*

Proof. Based on the closed graph theorem [39], we have that $\mathrm{NL_M}$ is continuous if and only if the graph of $\mathrm{NL_M}$ is a closed set. To this end, we will prove that for any sequence $(X_k, Y_k) \in \mathrm{gra}(\mathrm{NL_M})$ converging to $(X, Y) \in \mathbb{R}^{d_2+d} \times \mathbb{R}^{d_2+d}$, (X, Y) is in the graph of $\mathrm{NL_M}$.

Let

$$\widetilde{Y} := \mathrm{prox}_{\Phi,R}\big((E - R^{-1}M)Y + R^{-1}MX - R^{-1}G(X)\big). \tag{31}$$

Because $\mathrm{prox}_{\Phi,R}$ is firmly non-expansive with respect to R, we have that

$$\left\| Y_k - \widetilde{Y} \right\|_R^2 \le \left\langle Y_k - \widetilde{Y}, \big(E - R^{-1}M\big)(Y_k - Y) + R^{-1}M(X_k - X) - R^{-1}\big(G(X_k) - G(X)\big) \right\rangle_R. \tag{32}$$

Note that

$$G(X_k) - G(X) = \begin{bmatrix} 0 \\ \nabla p_1(Ax_k^2) \end{bmatrix} - \begin{bmatrix} 0 \\ \nabla p_1(Ax^2) \end{bmatrix} = \begin{bmatrix} 0 \\ \nabla p_1(Ax_k^2 - Ax^2) \end{bmatrix}.$$

Since ∇p_1 is continuous and A is linear, if $X_k \to X$ as $k \to \infty$, we have that $\lim_{k\to\infty} G(X_k) = G(X)$. Taking the limit on both sides of equation (32), we have that $\lim_{k\to+\infty} \|Y_k - \widetilde{Y}\|_R^2 = 0$. Then, combining this with the assumption that $\lim_{k\to+\infty}(X_k, Y_k) = (X, Y)$, we have that $Y = \widetilde{Y}$. Substituting this equation into (31), we observe that

$$Y = \mathrm{prox}_{\Phi,R}\big((E - R^{-1}M)Y + R^{-1}MX - R^{-1}G(X)\big).$$

This completes the proof. $\square$

Next, we verify that sequence $\mathcal{V}$ satisfies hypotheses (2) and (3) of Theorem 5.2. To this end, for the matrices M defined by (24) and $\widetilde{M} := M - L$, where $L := \begin{bmatrix} 0 & 0 \\ 0 & 2\beta I \end{bmatrix}$, we prove that they are symmetric positive definite.

Lemma 5.4. *If $\lambda > 0$, $S \in \mathbb{S}_+^d$, $\|S\|_2 < \frac{\lambda}{\|B\|_2^2 + 2\lambda\beta}$, then matrices M and $\widetilde{M}$ are symmetric positive definite matrices.*

Proof. It is easy to see that $\widetilde{M}$ is a symmetric matrix. Now we prove that $\widetilde{M}$ is positive definite. Based on the assumption that $\|S\|_2 < \frac{\lambda}{\|B\|_2^2+2\lambda\beta} < \frac{1}{2\beta}$, we have that $S^{-1} - 2\beta I \in \mathbb{S}_+^d$. We define

$$\Phi := \begin{bmatrix} \sqrt{\lambda} I & 0 \\ 0 & (S^{-1} - 2\beta I)^{\frac{1}{2}} \end{bmatrix} \quad \text{and} \quad \Psi := \begin{bmatrix} I & \frac{1}{\sqrt{\lambda}} B(S^{-1} - 2\beta I)^{-\frac{1}{2}} \\ \frac{1}{\sqrt{\lambda}}(S^{-1} - 2\beta I)^{-\frac{1}{2}} B^T & I \end{bmatrix}.$$

Then, we factor matrix $\widetilde{M}$ as $\widetilde{M} = \Phi\Psi\Phi^T$. Therefore, we need only to prove that matrix Ψ is positive definite. It suffices to prove that

$$\left\| \frac{1}{\sqrt{\lambda}}(S^{-1} - 2\beta I)^{-\frac{1}{2}} B^T \right\|_2 < 1. \tag{33}$$

We denote by $\mathrm{eig}(S)$ the set of the eigenvalues of matrix S, that is, $\mathrm{eig}(S) = \{\lambda_1, \ldots, \lambda_d\}$. Suppose that $\lambda_1 \le \cdots \le \lambda_d$. Then we have $\lambda_d \le \frac{\lambda}{\|B\|_2^2+2\lambda\beta}$, which means $(\frac{1}{\lambda_d} - 2\beta)^{-1} < \frac{\lambda}{\|B\|_2^2}$. It implies that

$$(\frac{1}{\lambda_1} - 2\beta)^{-1} \le \cdots \le (\frac{1}{\lambda_{d_1}} - 2\beta)^{-1} < \frac{\lambda}{\|B\|_2^2}.$$

Since

$$\mathrm{eig}((S^{-1}-2\beta I)^{-1}) = \big\{(\frac{1}{\lambda_1}-2\beta)^{-1}, \ldots, (\frac{1}{\lambda_d}-2\beta)^{-1}\big\},$$

we have that $\|(S^{-1}-2\beta I)^{-1}\|_2 < \frac{\lambda}{\|B\|_2^2}$. Then we obtain that

$$\big\|(S^{-1}-2\beta I)^{-\frac{1}{2}}(B^T)\big\|_2^2 \le \big\|(S^{-1}-2\beta I)^{-1}\big\|_2 \big\|B^T\big\|_2^2 < \lambda,$$

which yields (33). Based on equation (33), we have that $\widetilde{M}$ is a symmetric positive definite matrix. Likewise, we can prove that M is symmetric positive definite. □

We recall that a real-valued function p is Lipschitz continuous with Lipschitz constant K provided that

$$|p(x)-p(y)| \le K\|x-y\|, \text{ for all } x, y \in \mathbb{R}^d.$$

Lemma 5.5. *Suppose that function $h_1 := p_1 \circ A$ has Lipschitz continuous gradient with Lipschitz constant β, sequence $\mathcal{V}$ is generated by iterative scheme* (25) *and V is a fixed-point of* $\mathrm{NL_M}$. *If $\lambda > 0$, $S \in \mathbb{S}_+^d$, $\|S\|_2 < \frac{\lambda}{\|B\|_2^2+2\lambda\beta}$, then $\lim\limits_{k\to+\infty} \|V_{k+1}-V_k\|_M = 0$ and $\|V_{k+1}-V\|_M \le \|V_k - V\|_M$ for any initial vector $V_0 \in \mathbb{R}^{d_2+d}$.*

Proof. Notice that operator $\mathrm{prox}_{\Phi,R}$ that appears in the definition of operator $\mathrm{NL_M}$ is firmly non-expansive. By choosing $(V,V), (V_k, V_{k+1}) \in \mathrm{gra}(\mathrm{NL_M})$, it follows that

(34)
$$\Big\|V_{k+1}-V\Big\|_R^2 \le \Big\langle V_{k+1}-V, \big((E-R^{-1}M)(V_{k+1}-V)+R^{-1}M(V_k-V)-R^{-1}(G(V_k)-G(V))\Big\rangle_R.$$

Because

$$\begin{bmatrix} w^T & u^T \end{bmatrix} \begin{bmatrix} \lambda I & 0 \\ 0 & S^{-1} \end{bmatrix} \begin{bmatrix} w \\ u \end{bmatrix} = \begin{bmatrix} w^T & u^T \end{bmatrix} \begin{bmatrix} \lambda I & B \\ -B^T & S^{-1} \end{bmatrix} \begin{bmatrix} w \\ u \end{bmatrix},$$

for any $V \in \mathbb{R}^{d_2+d}$, we have that $\|V\|_R^2 = \langle V, EV\rangle_R$. Thus, we have that

$$\|V_{k+1}-V\|_R^2 = \big\langle V_{k+1}-V, E(V_{k+1}-V)\big\rangle_R. \tag{35}$$

Substituting (35) into (34), we obtain that

$$\big\langle V_{k+1}-V, M(V_{k+1}-V)\big\rangle \le \big\langle V_{k+1}-V, M(V_k-V)\big\rangle + \big\langle V_{k+1}-V, G(V)-G(V_k)\big\rangle. \tag{36}$$

We first consider reshaping the term $\mathcal{P}_k := \big\langle V_{k+1}-V, G(V)-G(V_k)\big\rangle$ by exploring the property of convex functions. Because

$$\mathcal{P}_k = \big\langle u_{k+1}-u, \nabla h_1(u) - \nabla h_1(u_k)\big\rangle,$$

we find that

$$\begin{aligned} \mathcal{P}_k = &\big\langle u_k - u_{k+1}, \nabla h_1(u_k) - \nabla h_1(u_{k+1})\big\rangle + \big\langle u_k - u_{k+1}, \nabla h_1(u_{k+1})\big\rangle \\ &+ \big\langle u_{k+1}-u, \nabla h_1(u)\big\rangle + \big\langle u - u_k, \nabla h_1(u_k)\big\rangle. \end{aligned} \tag{37}$$

Noting that h_1 is a Lipschitz continuous gradient with Lipschitz constant β, we have that

$$\big\langle u_k - u_{k+1}, \nabla h_1(u_k) - \nabla h_1(u_{k+1})\big\rangle \le \beta\|u_k - u_{k+1}\|_2^2. \tag{38}$$

Then using the convexity of h_1, we obtain that

$$\big\langle u_k - u_{k+1}, \nabla h_1(u_{k+1})\big\rangle \le h_1(u_k) - h_1(u_{k+1}), \tag{39}$$

$$\big\langle u_{k+1} - u, \nabla h_1(u)\big\rangle \le h_1(u_{k+1}) - h_1(u), \tag{40}$$

$$\big\langle u - u_k, \nabla h_1(u_k)\big\rangle \le h_1(u) - h_1(u_k). \tag{41}$$

Substituting inequalities (38), (39), (40), and (41) back into (37), we obtain that

$$2\mathcal{P}_k \leq 2\beta\langle u_k - u_{k+1}, u_k - u_{k+1}\rangle = \big\langle V_k - V_{k+1}, L(V_k - V_{k+1})\big\rangle \tag{42}$$

where $L = \begin{bmatrix} 0 & 0 \\ 0 & 2\beta I \end{bmatrix}$. From (42) and (36), we have that

$$2\big\langle V_{k+1} - V, M(V_{k+1} - V)\big\rangle \leq 2\big\langle V_{k+1} - V, M(V_k - V)\big\rangle + \big\langle V_k - V_{k+1}, L(V_k - V_{k+1})\big\rangle.$$

Base on the proof that M is a symmetric positive definite matrix, we have that

$$\begin{aligned} & \big\langle V_{k+1} - V, M(V_{k+1} - V)\big\rangle \\ \leq\ & -\big\langle V_{k+1} - V, M(V_{k+1} - V)\big\rangle + 2\big\langle V_{k+1} - V, M(V_k - V)\big\rangle + \big\langle V_k - V_{k+1}, L(V_k - V_{k+1})\big\rangle \\ =\ & -\big\langle V_k - V_{k+1}, M(V_k - V_{k+1})\big\rangle + \big\langle V_k - V, M(V_k - V)\big\rangle + \big\langle V_k - V_{k+1}, L(V_k - V_{k+1})\big\rangle. \end{aligned}$$

Meanwhile, $\widetilde{M}$ is also a symmetric positive definite matrix. Thus, we observe that

$$\|V_{k+1} - V\|_M^2 \leq \|V_k - V\|_M^2 - \|V_{k+1} - V_k\|_{\widetilde{M}}^2.$$

Summing the above inequality over k from 0 to n, we obtain that

$$\|V_{n+1} - V\|_M^2 + \sum_{k=0}^{n} \big\|V_{k+1} - V_k\big\|_{\widetilde{M}}^2 \leq \|V_0 - V\|_M^2.$$

From above inequality, we conclude that $\|V_{k+1} - V\|_M \leq \|V_k - V\|_M$, for $k = 0, 1, \ldots,$ and $\lim\limits_{k\to+\infty} \|V_{k+1} - V_k\|_{\widetilde{M}} = 0$. □

In conclusion, we have the following theorem:

Theorem 5.6. *If function $h_1 = p_1 \circ A$ has a Lipschitz continuous gradient with Lipschitz constant β, $\lambda > 0$, $S \in \mathbb{S}_+^d$, $\|S\|_2 < \frac{\lambda}{\|B\|_2^2 + 2\lambda\beta}$, then for any initial $V_0 \in \mathbb{R}^{d_2+d}$, sequence $\mathcal{V}$ generated from iterative scheme* (25) *converges to the fixed-point of* $\mathrm{NL_M}$.

Proof. This theorem is a direct consequence of Theorem 5.2, Lemmas 5.3, 5.4 and 5.5. □

6. High Resolution Image Reconstruction

In this section we apply the developed fixed-point algorithm based on the proximity and pre-condition operators (Algorithm 1) to the high-resolution image reconstruction model described in (13).

We first present the formulation of the TV-norm that we consider in this paper. Let D_M denote the $M \times M$ difference matrix, whose specific form depends on the boundary condition used. Let I_M denote the $M \times M$ identity matrix. The anisotropic total variation $\|u\|_{\mathrm{TV}}$ is defined by

$$\|u\|_{\mathrm{TV}} = \|Bu\|_1,$$

where

$$B :\, = \begin{bmatrix} I_{M_2} \otimes D_{M_1} \\ D_{M_2} \otimes I_{M_1} \end{bmatrix}. \tag{43}$$

Previous studies show that the Neumann boundary condition (assuming that the scene immediately outside is a reflection of the original scene at the boundary) gives better reconstruction results in the HR image reconstruction than the zero boundary condition and the periodic boundary condition [14, 15, 41]. Under the Neumann boundary condition hypothesis, D_N in (43) is defined by

$$D_N := \begin{bmatrix} 0 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}.$$

Next, we analyze the Lipschitz constant β of the gradient of $\mathrm{env}_{\tau\|\cdot\|_1}(A_L(\varepsilon^x,\varepsilon^y)\cdot -c)$. For simplicity, when $\varepsilon=0$, we write $A_L(\varepsilon^x,\varepsilon^y)$ as A_L.

Proposition 6.1. *If $h_1 := \mathrm{env}_{\tau\|\cdot\|_1}\big(A_L(\varepsilon^x,\varepsilon^y)\cdot -c\big)$, then ∇h_1 is Lipschitz continuous with Lipschitz constant $\beta := \frac{(\|A_L\|_2+\frac{4}{L}\varepsilon^*)^2}{\tau}$.*

Proof. For any $x,y\in\mathbb{R}^d$, we have that

$$\begin{aligned}&\big\|\nabla h_1(x)-\nabla h_1(y)\big\|_2\\ &=\frac{1}{\tau}\big\|A_L\big(\varepsilon^x,\varepsilon^y)^T\big[(I-\mathrm{prox}_{\tau\|\cdot\|_1})(A_L(\varepsilon^x,\varepsilon^y)x-c)-(I-\mathrm{prox}_{\tau\|\cdot\|_1})(A_L(\varepsilon^x,\varepsilon^y)y-c)\big]\big\|_2.\end{aligned} \tag{44}$$

Recall that $I-\mathrm{prox}_{\tau\|\cdot\|_1}$ is a firmly non-expansive operator [22]. Thus, for all $x,y\in\mathbb{R}^d$, we have that

$$\big\|(I-\mathrm{prox}_{\tau\|\cdot\|_1})(A_L(\varepsilon^x,\varepsilon^y)x-c)-(I-\mathrm{prox}_{\tau\|\cdot\|_1})(A_L(\varepsilon^x,\varepsilon^y)y-c)\big\|_2\le\big\|A_L(\varepsilon^x,\varepsilon^y)(x-y)\big\|_2. \tag{45}$$

From previous study $\big($Lemma 1, [40] $\big)$, for any given integer L,

$$\|A_L(\varepsilon^x,\varepsilon^y)-A_L\|_2\le\frac{4}{L}\varepsilon^*.$$

From this estimation, we obtain that

$$\big\|A_L(\varepsilon^x,\varepsilon^y)\big\|_2\le\big\|A_L\big\|_2+\frac{4}{L}\varepsilon^*. \tag{46}$$

By combining (44), (45) and (46), we obtain that

$$\big\|\nabla h_1(x)-\nabla h_1(y)\big\|_2\le\beta\big\|x-y\big\|_2,$$

where

$$\beta := \frac{(\|A_L\|_2+\frac{4}{L}\varepsilon^*)^2}{\tau}. \tag{47}$$

This completes the proof. ☐

We consider exploring Algorithm 1 to solve the optimization problem (13) with all the data being generated from the Bose-Boo model. The Neumann boundary condition is under consideration.

Theorem 6.2. *Let $\{u_k, u_k\in\mathbb{R}^m, k\in\mathbb{N}\}$ be the sequence generated from Algorithm 1 in the context described above. If $\lambda>0$, $S\in\mathbb{S}_+^m$, $\|S\|_2<\frac{\lambda\tau}{8\tau+2\lambda(1+\frac{4}{L})^2}$, then for any initial value $u_0\in\mathbb{R}^m$, sequence $\{u_k\}$ converges to the ideal HR image u.*

Proof. We recall that for the Bose-Boo model, $\|A_L\|_2 < 1$ (Lemma 2, [34]). Meanwhile, for the Neumann boundary condition, we have that $\|B\|_2^2 < 8$, [36]. We combine Theorem 5.6 and Proposition 6.1 to obtain this theorem based on the assumption that $\varepsilon^* = \max_{0\le p,q\le L-1}\left\{|\varepsilon^x_{p,q}|, |\varepsilon^y_{p,q}|\right\} < 1$. □

In the high resolution image reconstruction application, we choose the precondition matrix $S = sI$ for simplicity. Therefore, based on Theorem 5.6, we have $s < \frac{\lambda}{\|B\|_2^2+2\lambda\beta}$. We appropriately amplify the range and obtain that $0 < s < \frac{1}{2\beta}$. By the estimate of (47), we choose $\delta\colon = \frac{1}{2(\|A_L\|_2+\frac{4}{L}\varepsilon^*)^2}$. Then s and τ are required to satisfy

$$0 < \frac{s}{\tau} < \delta. \tag{48}$$

Based on the above discussion, we propose an accelerated algorithm by proposing an adaptive parameter choice strategy for Algorithm 1. In Algorithm 2, we set three parameters called s_t, *Adaptive_Frequency* and *Layer_Number*. We propose to select relatively large s and τ at the first several steps to quickly decline in the descent direction. So, the initial values of s and τ are chosen to be 128. Then s and τ are reduced by half. The reduction frequency is controlled by *Adaptive_Frequency* and it is given by experiences. After searching for several steps, τ is fixed to search in small scale. The fixed τ is for the purpose of the convergence condition (48). Meanwhile, a lower boundary of τ is defined by s_t, which is related to the Gaussian noise variance, to make sure that the ℓ^1-norm is used to deal with the salt-pepper noise and outliers and the ℓ^2-norm is used to deal with most Gaussian noise. When there is no Gaussian noise, s_t is set to be zero. When $i >$ *Layer_Number*, s and τ are fixed to be a constant.

Algorithm 2: (Fixed-point algorithm based on the proximity operator accelerated by updating parameters)
Initialization: $v_0 = 0, u_0 = 0, \lambda > 0$, *Adaptive_Frequency*, *Layer_number*, $i = 0$, $\tau = 128$, σ is the standard deviation of the Gaussian noise, $s\_t = 3\sigma$.
repeat
  $z = k \ mod \ Adaptive\_Frequency$;
  if $z == 0$ **and** $i < Layer\_Number$
  $i = i + 1$;
  else
  end
    $s = 128/2^i$;
    if $\tau > s\_t$
      $\tau = 128/2^i$;
    else
      $\tau = s\_t$;
    end
  $v_{k+1} = (I - \mathrm{prox}_{\lambda \mathrm{p}_2})(Bu_k + v_k)$
  $u_{k+1} = \mathrm{prox}_{p_3,s^{-1}}\left(u_k - s(\nabla p_1^\tau(Au_k) + \frac{1}{\lambda}B^T\left(2v_{k+1} - v_k\right))\right)$
until ‘convergence’

(a) (b)

FIGURE 2. The orignail figures: (a) Lena and (b) Cameraman.

TABLE 1. PSNR(dB) results of 2×2 times reconstructed images with different displacement-errors and different noise levels.

		Lena		Cam	
	Noise $\sigma^2 + r$	ℓ^2/TV	env$_{\ell^1}$/TV	ℓ^2/TV	env$_{\ell^1}$/TV
$disp = 0$	$0 + 0.1$	26.6768	**35.8171**	25.5093	**33.0426**
	$0.001 + 0$	**33.3435**	**33.0933**	**31.6235**	**31.0419**
	$0.001 + 0.05$	26.8779	**32.1452**	25.6236	**30.5133**
$disp = 0.5$	$0 + 0.1$	26.2636	**35.2317**	**25.3206**	**32.4254**
	$0.001 + 0$	**32.5381**	**33.0786**	**31.5269**	**30.9662**
	$0.001 + 0.05$	26.5678	**32.0478**	25.6972	**30.4003**

7. Numerical Result

In this section, we compare numerical performance of the ℓ^2/TV model and the env$_{\ell^1}$/TV model. The optimal parameters are selected by using the grid search method at the high-performance computer cluster equiped with MATLAB R2013a and the experiments are repeated for testing robustness. After the best combination of parameters are selected, the experiments with optimal parameters are implemented on the OS X system of version 10.9.5 with 2.4 GHz dual-core Intel Core i5. The compiler is MATLAB R2015a.

We discuss computation of $A_L(\varepsilon^x_{p,q})$ only and that of $A_L(\varepsilon^y_{p,q})$ may be likewise discussed. For example, when $L = 2$, we have

$$A_2 = \frac{1}{4}\begin{bmatrix} \frac{3}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{3}{2} \end{bmatrix}, A_2(\varepsilon^x_{p,q}) = \frac{1}{4}\begin{bmatrix} \frac{3}{2} - \varepsilon^x_{p,q} & \frac{1}{2} + \varepsilon^x_{p,q} & 0 & 0 \\ \frac{1}{2} - \varepsilon^x_{p,q} & 1 & \frac{1}{2} + \varepsilon^x_{p,q} & 0 \\ 0 & \frac{1}{2} - \varepsilon^x_{p,q} & 1 & \frac{1}{2} + \varepsilon^x_{p,q} \\ 0 & 0 & \frac{1}{2} - \varepsilon^x_{p,q} & \frac{3}{2} + \varepsilon^x_{p,q} \end{bmatrix}$$

and

$$A_2(\varepsilon^y_{p,q}) = \frac{1}{4}\begin{bmatrix} \frac{3}{2} - \varepsilon^y_{p,q} & \frac{1}{2} - \varepsilon^y_{p,q} & 0 & 0 \\ \frac{1}{2} + \varepsilon^y_{p,q} & 1 & \frac{1}{2} - \varepsilon^y_{p,q} & 0 \\ 0 & \frac{1}{2} + \varepsilon^y_{p,q} & 1 & \frac{1}{2} - \varepsilon^y_{p,q} \\ 0 & 0 & \frac{1}{2} + \varepsilon^y_{p,q} & \frac{3}{2} + \varepsilon^y_{p,q} \end{bmatrix}$$

under Neumann boundary condition. In addition, no matter which boundary condition is imposed on the model, the interior row of $A_L(\varepsilon^x_{p,q})$ is given by

$$\frac{1}{L}\left[0,\ldots,0,\frac{1}{2}+\varepsilon^x_{p,q},\underbrace{1,\ldots,1}_{L-1},\frac{1}{2}-\varepsilon^x_{p,q},0\ldots,0\right]$$

which can be considered as a low-pass filter acting on the image U. This low-pass filter is a tensor product of the univariate low-pass filter

$$m_{0,L,\varepsilon} := \frac{1}{L}\left[\frac{1}{2}+\varepsilon,\underbrace{1,\ldots,1}_{L-1},\frac{1}{2}-\varepsilon\right],$$

where the parameter ε may vary for each sensor. Let

$$m_{0,L,0} := \frac{1}{L}\left[\frac{1}{2},\underbrace{1,\ldots,1}_{L-1},\frac{1}{2}\right] \quad \text{and} \quad m_{1,L,0} := \frac{1}{L}\left[1,\underbrace{0,\ldots,0}_{L-1},-1\right].$$

Then we split the low-pass filter as

$$m_{0,L,\varepsilon} = m_{0,L,0} + \varepsilon\ m_{1,L,0}.$$

Based on this splitting, we get $m_{0,L,0}$ which is not related to ε. It is computed first and stored for later use. This method can reduce the computational cost tremendously. We realize the filters by exploring the MATLAB command $imfilter(A_L, m_{0,L,\varepsilon},'replicate','same','conv')$.

TABLE 2. PSNR(dB) results of 4×4 times reconstructed images with different displacement-errors and different noise levels.

		Lena		Cam	
	Noise σ^2+r	ℓ^2/TV	env_{ℓ^1}/TV	ℓ^2/TV	env_{ℓ^1}/TV
$disp=0$	$0+0.1$	26.4766	**32.5910**	25.1103	**29.3599**
	$0.001+0$	**30.9348**	**30.7920**	**28.7026**	**28.1657**
	$0.001+0.05$	26.7846	**30.4191**	26.0110	**28.0595**
$disp=0.5$	$0+0.1$	26.4028	**32.5549**	25.1892	**29.383**
	$0.001+0$	**30.9956**	30.7814	**28.7246**	**28.1951**
	$0.001+0.05$	26.5804	**30.429**	25.024	**28.0847**

TABLE 3. SSIM results of 2×2 times reconstructed images with different displacement-errors and different noise levels.

		Lena		Cam	
	Noise σ^2+r	ℓ^2/TV	env_{ℓ^1}/TV	ℓ^2/TV	env_{ℓ^1}/TV
$disp=0$	$0+0.1$	0.4082	**0.8791**	0.2331	**0.7622**
	$0.001+0$	**0.6903**	**0.6698**	**0.4531**	**0.4484**
	$0.001+0.05$	0.4561	**0.6577**	0.2849	**0.4375**
$disp=0.5$	$0+0.1$	0.4060	**0.8718**	**0.2300**	**0.7501**
	$0.001+0$	**0.6423**	**0.6658**	**0.4537**	**0.4492**
	$0.001+0.05$	0.4314	**0.6464**	0.2809	**0.4363**

We use the MATLAB command $imnoise(Image,' gaussian', 0, \sigma^2)$ to add Gaussian noise, in which the mean value is given by 0 and the variance is σ^2 when the image is scaled between $[0, 1]$. We remark that salt-pepper noise is a kind of special impulse noise and all numerical experiments are implemented using salt-pepper noise. For a salt-pepper noise generator, we explore the command $imnoise(Image,' salt \ \& \ pepper', r)$, where r is the noise density, which indicates the percentage of corrupted pixels. In our simulation, the displacement-error matrices ε^x and ε^y for the $L \times L$ sensor array are simulated by the following three MATLAB commands: *rand('seed',100)*, $\varepsilon^x = z * (rand(L) - 0.5)$ and $\varepsilon^y = z * (rand(L) - 0.5)$, where z is chosen in $(0, 1)$.

TABLE 4. SSIM results of 4×4 times reconstructed images with different displacement-errors and different noise levels.

		Lena		Cam	
	Noise $\sigma^2 + r$	ℓ^2/TV	env$_{\ell^1}$/TV	ℓ^2/TV	env$_{\ell^1}$/TV
$disp = 0$	$0 + 0.1$	0.3865	**0.7710**	0.1962	**0.5940**
	$0.001 + 0$	**0.6149**	**0.6040**	**0.3760**	**0.3582**
	$0.001 + 0.05$	0.4029	**0.5925**	0.2444	**0.3537**
$disp = 0.5$	$0 + 0.1$	0.3974	**0.7693**	0.2011	**0.5935**
	$0.001 + 0$	**0.6073**	0.6017	**0.3847**	**0.3659**
	$0.001 + 0.05$	0.3903	**0.5877**	0.2237	**0.3535**

In the numerical result to be reported, we compare the performance of the ℓ^2/TV model and the env$_{\ell^1}$/TV model. For this reason, we use the most suitable algorithm to solve each of the models. Specifically, we use Algorithm 1 to solve the ℓ^2/TV model and Algorithm 2 to solve the env$_{\ell^1}$/TV model. Iterations will terminate when the following condition is satisfied

$$\frac{\|x^n - x^{n+1}\|_2}{\|x^{n+1}\|_2} \leq \mathrm{TOL}.$$

For the fixed stopping criteria $\mathrm{TOL} = 1.0 \times 10^{-4}$, the env$_{\ell^1}$/TV model stop at $60 \sim 80$ steps in most cases while the ℓ^2/TV model stop at $170 \sim 190$ steps.

We show the reconstructed images in Figures 3 and 4 for the situation that the displacement-error is of $\frac{1}{2}$ pixel size and the noise is the mixed Gaussian noise $\sigma^2 = 0.001$ and salt-pepper noise with $r = 5\%$. We remark here that the *Adaptive_Frequency* and *Layer_Number* are expirically selected to obtain highest peak signal to noise ratio (PSNR) result. It can be clearly seen from Figures 5 and 6 that the env$_{\ell^1}$/TV model is more effective than the ℓ^2/TV model. We also explore both the highest peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) [55] to measure the quality of the reconstructed images. Tables 1 and 2 give the PSNR value when the resolution is enhanced by 2×2 times and 4×4 times under different displacement-error levels, respectively, while Tables 3 and 4 show the SSIM value correspondently. The results are stable with the change of displacement-errors. We conclude that the env$_{\ell^1}$/TV model is better than the ℓ^2/TV model in the existence of impulse noise, while for Gaussian noise the results of the two models are comparable. Meanwhile, the env$_{\ell^1}$/TV model performs well and is stable for the HR image reconstruction with displacement-errors.

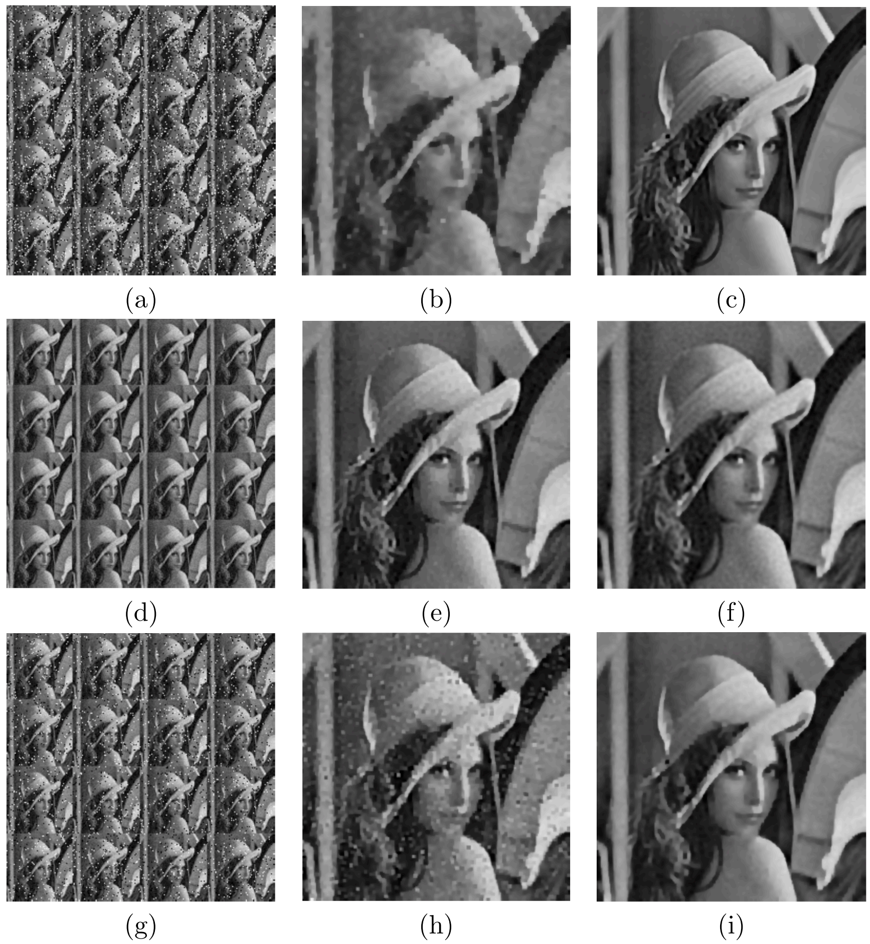

FIGURE 3. HR image reconstruction results for the corrupted 'Lena' with $\frac{1}{2}$ pixel size displacement-error. Noise level: 1. 10% salt-pepper noise, 2. $\sigma^2 = 0.001$ Gaussian noise, 3. 0.05% salt-pepper noise and $\sigma^2 = 0.001$ Gaussian noise. (a) (d) (g) are spliced LR images by 16 small LR images with noise levels 1-3 respectively, (b) (e) (h) are reconstructed results of the ℓ^2/TV deblurring model and (c) (f) (i) are reconstructed results of the env_{ℓ^1}/TV deblurring model.

8. Conclusion

We propose an env_{ℓ^1}/TV model for HR image reconstruction with displacement-errors to get higher image reconstruction quality. A fixed-point proximity scheme is developed for solving the model and convergence of the proposed algorithm is proved. A new iterative fixed-point proximity algorithm is developed to accelerate the convergence with an adaptive parameter choice strategy. Numerical experiment results show that the proposed model is stable and effective for the mixed Gaussian noise, impulse noise and displacement error. The env_{ℓ^1}/TV model outpeforms the ℓ^2/TV model for the mixed Gaussian and impulse noise.

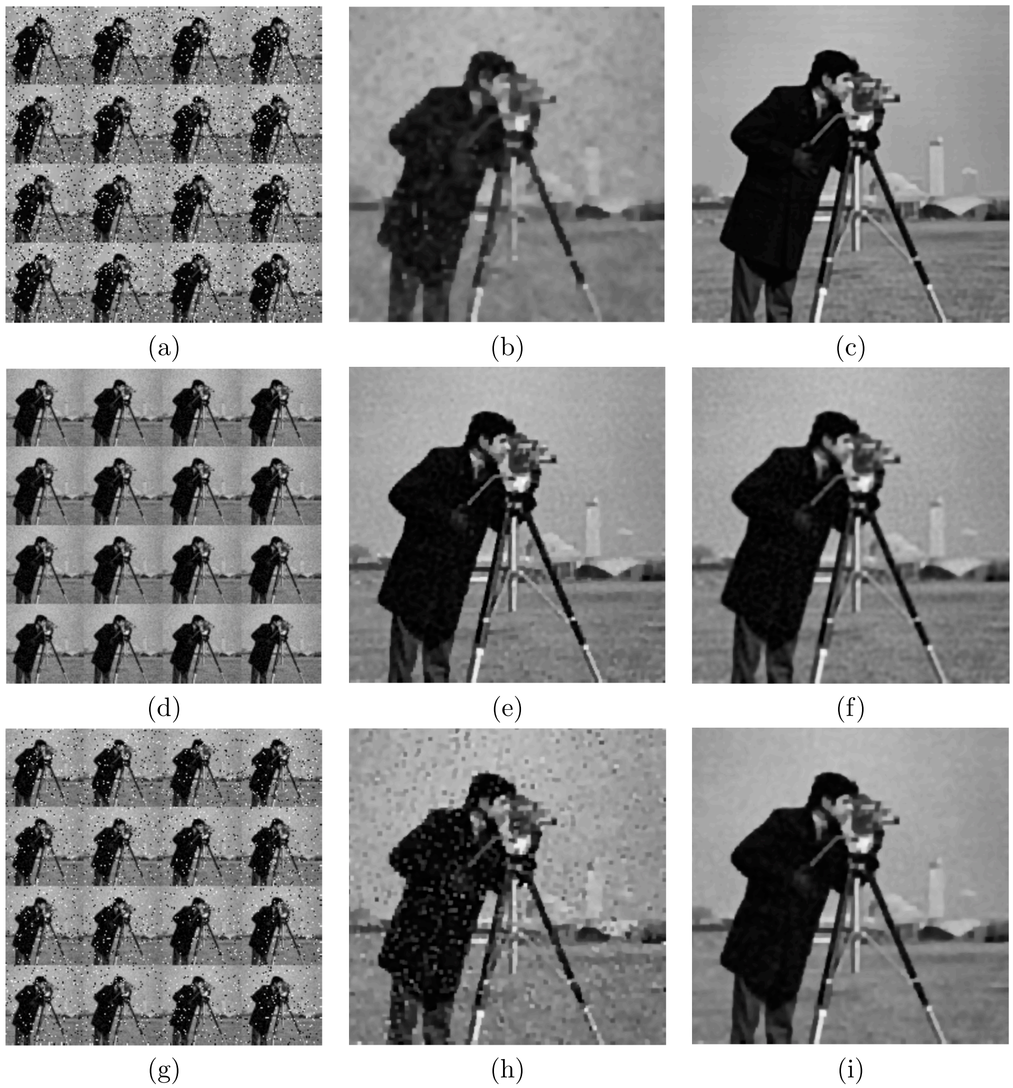

FIGURE 4. HR image reconstruction results for the corrupted 'Cameraman' with $\frac{1}{2}$ pixel size displacement-error. Noise level: 1. 10% salt-pepper noise, 2. $\sigma^2 = 0.001$ Gaussian noise, 3. 0.05% salt-pepper noise and $\sigma^2 = 0.001$ Gaussian noise. (a) (d) (g) are spliced LR images by 16 small LR images with noise levels 1-3 respectively, (b) (e) (h) are reconstructed results of the ℓ^2/TV deblurring model and (c) (f) (i) are reconstructed results of the env_{ℓ^1}/TV deblurring model.

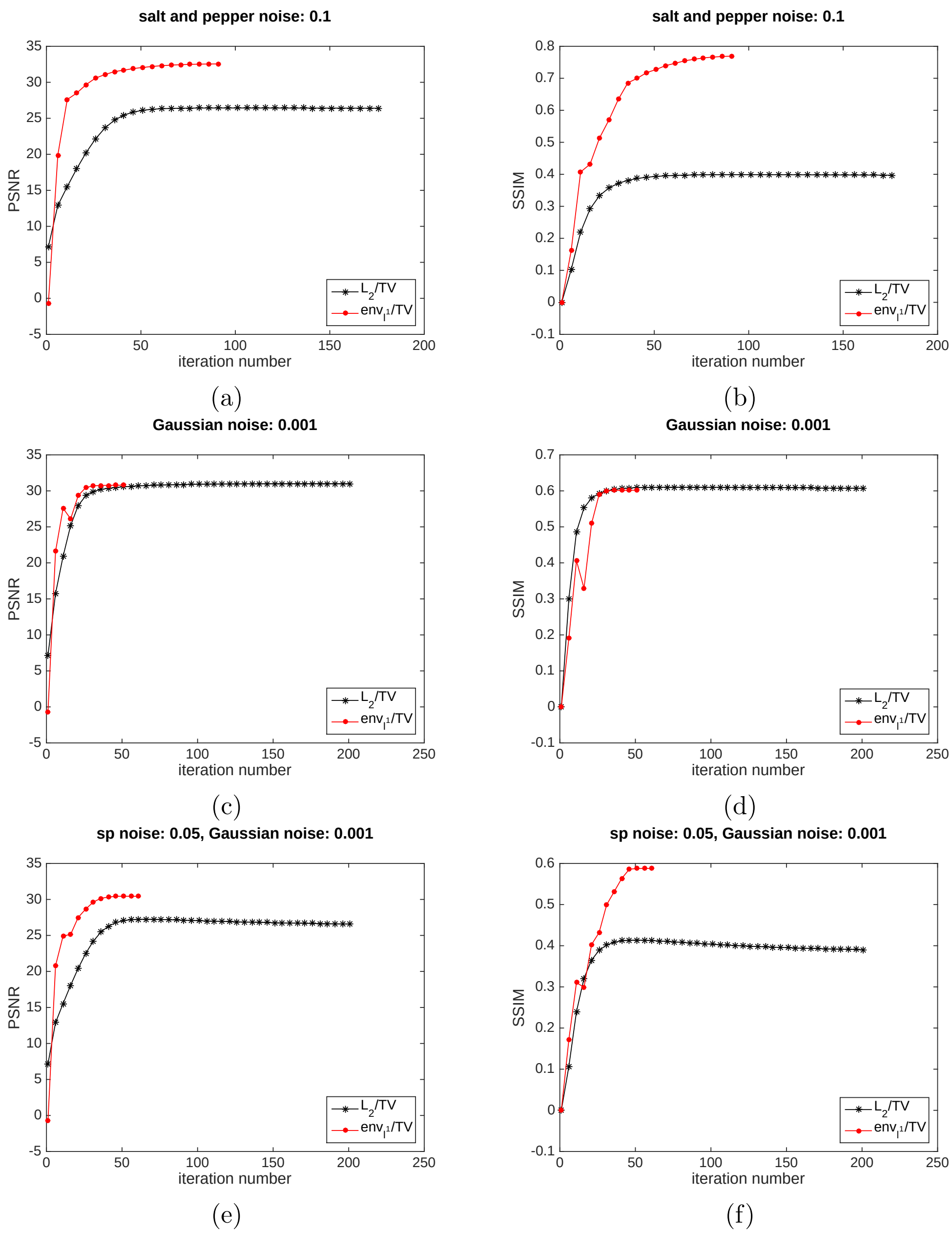

FIGURE 5. PSNR and SSIM values for resolution improved 16 times for the corrupted 'Lena' with $\frac{1}{2}$ pixel size displacement-error. Noise level: 1. 10% salt-pepper noise, 2. $\sigma^2 = 0.001$ Gaussian noise, 3. 0.05% salt-pepper noise and $\sigma^2 = 0.001$ Gaussian noise. (a),(c) and (e) are figures of PSNR values while (b),(d) and (f) are figures of SSIM values.

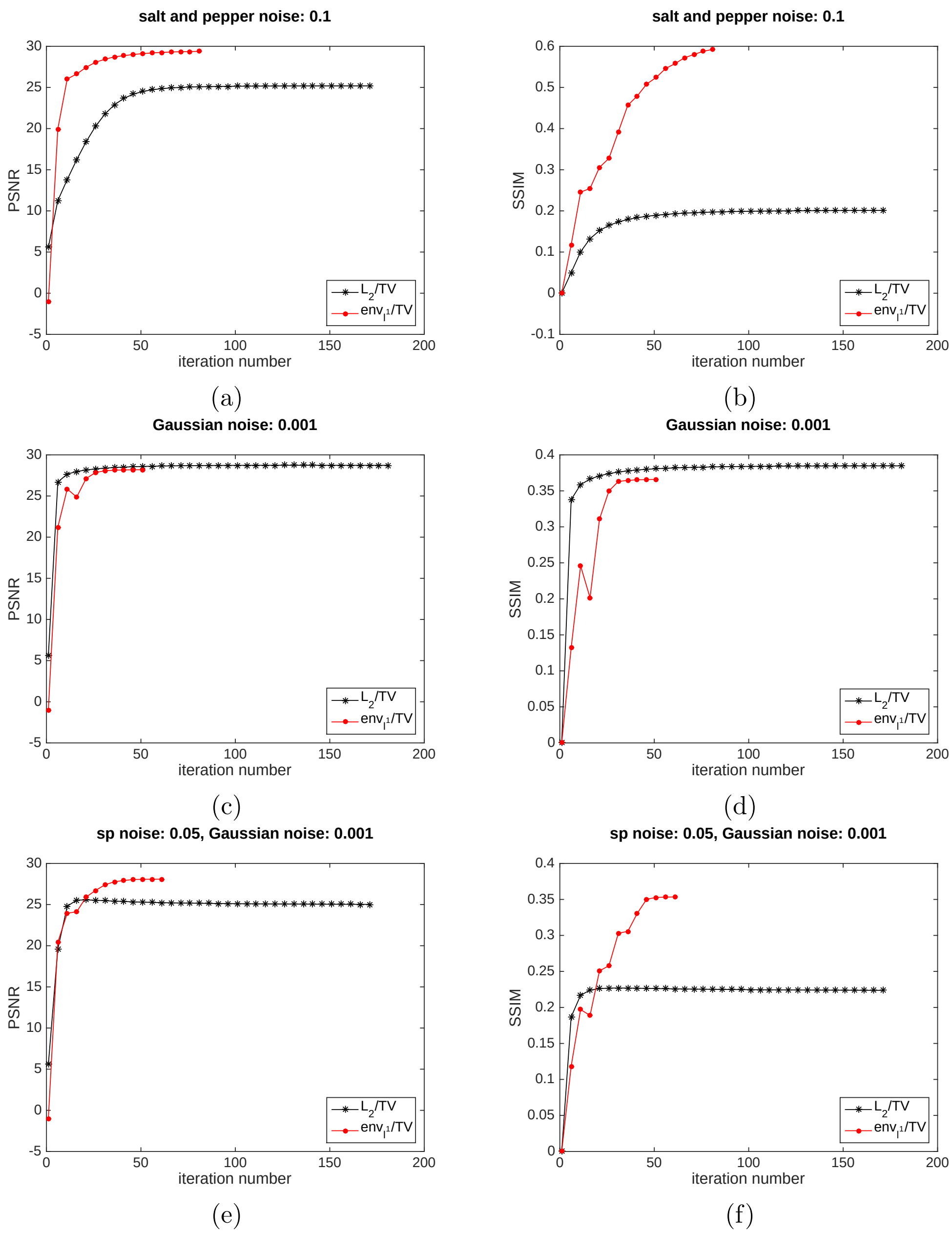

FIGURE 6. PSNR and SSIM values for resolution improved 16 times for the corrupted 'Cameraman' with $\frac{1}{2}$ pixel size displacement-error. Noise level: 1. 10% salt-pepper noise, 2. $\sigma^2 = 0.001$ Gaussian noise, 3. 0.05% salt-pepper noise and $\sigma^2 = 0.001$ Gaussian noise. (a),(c) and (e) are figures of PSNR values while (b),(d) and (f) are figures of SSIM values.

9. Acknowledgement

The research was supported in part by the Ministry of Science and Technology of China under the Special Project on High-performance Computing under the National Key R&D Program (No. 2016YFB0200602),the US National Science Foundation under grant DMS-1522332, by the US Department of Energy under grant 20166700041050011, by Guangdong Provincial Government of China through the Computational Science Innovative Research Team program, by the Natural Science Foundation of China under grants 11471013, 11401601 and 91530117, by Innovation Key Fund of Guangdong Province under grants 2016B030307003, 2015B010110003, 2015B020233008 and by Guangdong Key Fund 20151019 1740058.

Appendix A. The Closed Form of Proximity Operators

The closed form of the proximity operator of the ℓ^1-norm can be found in [36].

Next, we present the exact form of the sub-differential and the proximity operator of $\iota_\Omega(\cdot)$, where Ω is defined in (12). By computation, for given $x = (x_i, i \in \mathbb{N}^d)^T \in \Omega$, the sub-differential of ι_Ω at x is

$$\partial\iota_\Omega(x) = P(x)\mathbb{R}^d_+,$$

where $P(x) := \mathrm{diag}(\zeta(x_i), i \in \mathbb{N}^d)$ and $\zeta : [0, 255] \to \{-1, 0, 1\}$,

$$\zeta(x_i) := \begin{cases} -1 & x_i = 0 \\ 0 & x_i \in (0, 255) \\ 1 & x_i = 255. \end{cases}$$

We define a set $\mathbb{S}^d_+ := \{S : S$ is a $d \times d$ diagonal matrix with positive diagonal entries$\}$. We also let the set $\mathbb{K}$ for $H \in \mathbb{R}^{d\times d}$ be defined as

$$H\mathbb{K} := \{Hk : k \in \mathbb{K}\}.$$

Then, for any $x \in \Omega, S \in \mathbb{S}^d_+$, we have the property

$$S\partial\iota_\Omega(x) = SP(x)\mathbb{R}^d_+ = P(x)S\mathbb{R}^d_+ = P(x)\mathbb{R}^d_+ = \partial\iota_\Omega(x)$$

based on the fact that $S\mathbb{R}^d_+ = \mathbb{R}^d_+$. For a given convex set w, the *projection operator* is denoted by Pr_{w}. Thus, we have the conclusion $S\partial\iota_\Omega = \partial\iota_\Omega$, which implies

$$\mathrm{prox}_{\iota_\Omega, S^{-1}} = \mathrm{prox}_{\iota_\Omega} = \mathrm{Pr}_{\iota_\Omega}.$$

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(4):913–917, 1997.
[2] R. Andersson, C. Bruder, A. Piotrowski, U. Menzel, H. Nord, J. Sandgren, T. Hvidsten, T. de Ståhl, J. Dumanski, and J. Komorowski. A segmental maximum a posteriori approach to genome-wide copy number profiling. Bioinformatics, 24(6):751–758, 2008.
[3] H. C. Andrews and B. R. Hunt. Digital image restoration. Prentice-Hall Signal Processing Series, 1, 1977.
[4] H. H. Bauschke and P. L. Combettes. Convex analysis and monotone operator theory in Hilbert spaces. Springer, New York, 2011.
[5] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. IEEE Transactions on Image Processing, 18(11):2419–2434, 2009.
[6] D. P. Bertsekas. Convex Optimization Algorithms. Athena Scientific, Massachusetts, 2015.
[7] M. J. Black. Robust incremental optical flow. PhD thesis, Yale University, 1992.
[8] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger. Robust anisotropic diffusion. IEEE Transactions on Image Processing, 7(3):421–432, 1998.

[9] N. K. Bose and K. J. Boo. High-resolution image reconstruction with multisensors. International Journal of Imaging Systems and Technology, 9(4):294–304, 1998.
[10] A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. SIAM review, 51(1):34–81, 2009.
[11] J. Cai, R. Chan, and M. Nikolova. Fast two-phase image deblurring under impulse noise. Journal of Mathematical Imaging and Vision, 36(1):46–53, 2010.
[12] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory, 52(2):489–509, 2006.
[13] 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(1):120–145, 2011.
[14] R. Chan, T. Chan, M. Ng, W. Tang, and C. Wong. Preconditioned iterative methods for high-resolution image reconstruction with multisensors. SPIE proceedings series, 3461:348–357, 1998.
[15] R. Chan, T. Chan, L. Shen, and Z. Shen. Wavelet algorithms for high-resolution image reconstruction. SIAM Journal on Scientific Computing, 24(4):1408–1432, 2003.
[16] R. Chan, S. D. Riemenschneider, L. Shen, and Z. Shen. Tight frame: an efficient way for high-resolution image reconstruction. Applied and Computational Harmonic Analysis, 17(1):91–115, 2004.
[17] R. Chan, M. Tao, and X. Yuan. Constrained total variation deblurring models and fast algorithms based on alternating direction method of multipliers. SIAM Journal on imaging Sciences, 6(1):680–697, 2013.
[18] R Chan, H Yang, and T Zeng. A two-stage image segmentation method for blurry images with poisson or multiplicative gamma noise. Siam Journal on Imaging Sciences, 7(1):98–127, 2014.
[19] F. Chen, L. Shen, Y. Xu, and X. Zeng. The moreau envelope approach for the l1/tv image denoising model. Inverse Problems and Imaging, 8(1):53–77, 2014.
[20] P. Chen, J. Huang, and X. Zhang. A primal–dual fixed point algorithm for convex separable minimization with applications to image restoration. Inverse Problems, 29(2):025011, 2013.
[21] Q. Chen, C. A. Micchelli, S. Peng, and Y. Xu. Multivariate filter banks having matrix factorizations. SIAM Journal on Matrix Analysis and Applications, 25(2):517–531, 2003.
[22] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. Multiscale Modeling & Simulation, 4(4):1168–1200, 2005.
[23] J. Eckstein and W. Yao. Augmented lagrangian and alternating direction methods for convex optimization: A tutorial and some illustrative computational results. RUTCOR Research Reports, 32, 2012.
[24] E. Esser, X. Zhang, and T. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. SIAM Journal on Imaging Sciences, 3(4):1015–1046, 2010.
[25] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. IEEE Transactions on Image processing, 13(10):1327–1344, 2004.
[26] P. J. Huber. Robust estimation of a location parameter. The Annals of Mathematical Statistics, 35(1):73–101, 1964.
[27] P. J. Huber. Robust statistics. Springer, Berlin, 2011.
[28] M. Irani and S. Peleg. Improving resolution by image registration. CVGIP: Graphical models and image processing, 53(3):231–239, 1991.
[29] B. H. Juang and L. R. Rabiner. The segmental k-means algorithm for estimating parameters of hidden markov models. IEEE Transactions on Acoustics, Speech and Signal Processing, 38(9):1639–1641, 1990.
[30] S. J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. An interior-point method for large-scale l 1-regularized least squares. IEEE Journal of Selected Topics in Signal Processing, 1(4):606–617, 2007.
[31] S. P. Kim, N. K. Bose, and H. M. Valenzuela. Recursive reconstruction of high resolution image from noisy undersampled multiframes. IEEE Transactions on Acoustics, Speech and Signal Processing, 38(6):1013–1027, 1990.
[32] C. H. Lee and J. L. Gauvain. Speaker adaptation based on map estimation of hmm parameters. In 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing., volume 2, pages 558–561. IEEE, 1993.
[33] Q. Li, L. Shen, Y. Xu, and N. Zhang. Multi-step fixed-point proximity algorithms for solving a class of optimization problems arising from image processing. Advances in Computational Mathematics, 41(2):387–422, 2015.

[34] Y. Lu, L. Shen, and Y. Xu. Multi-parameter regularization methods for high-resolution image reconstruction with displacement errors. IEEE Transactions on Circuits and Systems I: Regular Papers, 54(8):1788–1799, 2007.
[35] C. A. Micchelli, L. Shen, and Y. Xu. Proximity algorithms for image models: denoising. Inverse Problems, 27(4):045009, 2011.
[36] C. A. Micchelli, L. Shen, Y. Xu, and X. Zeng. Proximity algorithms for the l1/tv image denoising model. Advances in Computational Mathematics, 38(2):401–426, 2013.
[37] C. A. Micchelli and Y. Xu. Using the matrix refinement equation for the construction of wavelets on invariant sets. Applied and Computational Harmonic Analysis, 1(4):391–401, 1994.
[38] C. A. Micchelli and Y. Xu. Reconstruction and decomposition algorithms for biorthogonal multiwavelets. Multidimensional Systems and Signal Processing, 8(1-2):31–69, 1997.
[39] J. Munkres. Topology. Prenctice Hall, Inc, 2000.
[40] M. Ng and N. K. Bose. Analysis of displacement errors in high-resolution image reconstruction with multisensors. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 49(6):806–813, 2002.
[41] M. Ng and A. Yip. A fast map algorithm for high-resolution image reconstruction with multisensors. Multidimensional Systems and Signal Processing, 12(2):143–164, 2001.
[42] M. Nikolova. A variational approach to remove outliers and impulse noise. Journal of Mathematical Imaging and Vision, 20(1-2):99–120, 2004.
[43] A. Panagiotopoulou and V. Anastassopoulos. Super-resolution image reconstruction techniques: Trade-offs between the data-fidelity and regularization terms. Information Fusion, 13(3):185–195, 2012.
[44] N. Parikh and S. Boyd. Proximal algorithms. Foundations and Trends in optimization, 1(3):123–231, 2013.
[45] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: a technical overview. Signal Processing Magazine, 20(3):21–36, 2003.
[46] V. Patanavijit and S. Jitapunkul. A robust iterative multiframe super-resolution reconstruction using a huber bayesian approach with huber-tikhonov regularization. In International Symposium on Intelligent Signal Processing and Communications, pages 13–16. IEEE, 2006.
[47] V. Patanavijit and S. Jitapunkul. A lorentzian stochastic estimation for a robust iterative multiframe super-resolution reconstruction with lorentzian-tikhonov regularization. EURASIP Journal on Advances in Signal Processing, 2007(2):21–21, 2007.
[48] N. Pustelnik, C. Chaux, and J. C. Pesquet. Parallel proximal algorithm for image restoration using hybrid regularization. IEEE Transactions on Image Processing, 20(9):2450–2462, 2011.
[49] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena, 60(1):259–268, 1992.
[50] H. Song, L. Zhang, P. Wang, K. Zhang, and X. Li. An adaptive l1–l2 hybrid error model to super-resolution. In IEEE International Conference on 17th Image Processing (ICIP), pages 2821–2824. IEEE, 2010.
[51] J. Tian and K. Ma. Stochastic super-resolution image reconstruction. Journal of Visual Communication and Image Representation, 21(3):232–244, 2010.
[52] S. Villena, M. Vega, S. D. Babacan, R. Molina, and A. K. Katsaggelos. Bayesian combination of sparse and non-sparse priors in image super resolution. Digital Signal Processing, 23(2):530–541, 2013.
[53] M. Vrigkas, C. Nikou, and L. P. Kondi. Accurate image registration for map image super-resolution. Signal Processing: Image Communication, 28(5):494–508, 2013.
[54] C. Wang, J. Chen, and R. Liu. Development and evaluation of mri based bayesian image reconstruction methods for pet. Computerized Medical Imaging and Graphics, 28(4):177–184, 2004.
[55] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4):600–612, 2004.
[56] Y Wen, R Chan, and T Zeng. Primal-dual algorithms for total variation based image restoration under poisson noise dedicated to professor lin qun on the occasion of his 80th birthday. Science China Mathematics, 59(1):141–160, 2016.
[57] G. Winkler. Random fields and markov chain monte carlo cmethods. Springer, Berlin, 2003.
[58] Y. Xiao, T. Zeng, J. Yu, and M. Ng. Restoration of images corrupted by mixed gaussian-impulse noise via l1–l0 minimization. Pattern Recognition, 44(8):1708–1720, 2011.
[59] L. Yue, H. Shen, Q. Yuan, and L. Zhang. A locally adaptive l1- l2 norm for multi-frame super-resolution of images with mixed noise and outliers. Signal Processing, 105:156–174, 2014.

[60] M. E. Zervakis and A. N. Venetsanopoulos. Linear and nonlinear image restoration under the presence of mixed noise. IEEE Transactions on Circuits and Systems, 38(3):258–272, 1991.

[61] M. Zhu and T. Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. UCLA CAM Report, pages 08–34, 2008.

School of Data and Computer Science and Guangdong Province Key Lab of Computational Science, Sun Yat-sen University, Guangzhou 510006, P. R. China

E-mail: `longwenting@gmail.com`

School of Data and Computer Science and Guangdong Province Key Lab of Computational Science, Sun Yat-sen University, Guangzhou 510006, P. R. China

E-mail: `luyao23@mail.sysu.edu.cn`

Department of Mathematics, Syracuse University, Syracuse, NY 13244, USA

E-mail: `lshen03@syr.edu`

School of Data and Computer Science and Guangdong Province Key Lab of Computational Science, Sun Yat-sen University, Guangzhou 510006, P. R. China. This author is also an emeritus professor of Mathematics Department, Syracuse University, Syracuse, NY 13244, USA. All correspondence should be sent to this author.

E-mail: `yxu06@syr.edu`