Fast Nonconvex Nonsmooth Minimization Methods for Image Restoration and Reconstruction

Mila Nikolova, *Senior Member, IEEE,* Michael K. Ng, and Chi-Pan Tam

Abstract—Nonconvex nonsmooth regularization has advantages over convex regularization for restoring images with neat edges. However, its practical interest used to be limited by the difficulty of the computational stage which requires a nonconvex nonsmooth minimization. In this paper, we deal with nonconvex nonsmooth minimization methods for image restoration and reconstruction. Our theoretical results show that the solution of the nonconvex nonsmooth minimization problem is composed of constant regions surrounded by closed contours and neat edges. The main goal of this paper is to develop fast minimization algorithms to solve the nonconvex nonsmooth minimization problem. Our experimental results show that the effectiveness and efficiency of the proposed algorithms.

Index Terms—Continuation methods, fast Fourier transform, image reconstruction, image restoration, nonconvex nonsmooth global minimization, nonconvex nonsmooth regularization, total variation.

I. INTRODUCTION

DIGITAL image restoration and reconstruction plays an important part in various applied areas such as medical and astronomical imaging, film restoration, image and video coding and many others [20], [16]. We focus on the most common data production model where the observed data $\mathbf{g} \in \mathbb{R}^q$ are related to the underlying $n \times m$ image, rearranged into a vector $\mathbf{f} \in \mathbb{R}^p$ $p = mn$ according to

$$\mathbf{g} = H\mathbf{f} + \mathbf{n} \tag{1}$$

where $\mathbf{n} \in \mathbb{R}^q$ accounts for the perturbations and H is a $q \times p$ matrix representing for instance optical blurring, distortion wavelets in seismic imaging and nondestructive evaluation, a Radon transform in X-ray tomography, a Fourier transform in diffraction tomography. In most of the applications, the information provided by the forward model (1) alone is not sufficient to find an acceptable solution $\hat{\mathbf{f}}$. Prior information on the underlying image is needed to restore a convenient $\hat{\mathbf{f}}$—which is close

M. Nikolova is with Centre de Mathématiques et de Leurs Applications (CMLA) CNRS (UMR 8536)—ENS Cachan, 94235 Cachan Cedex, France (e-mail: nikolova@cmla.ens-cachan.fr).

M. K. Ng and C.-P. Tam are with the 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; cptam@math.hkbu.edu.hk).

to data production model (1) and satisfies some prior requirements. A flexible means to define such a solution is regularization, e.g., [1], [5], [10], and [15], where $\hat{\mathbf{f}}$ is a minimizer of a cost function (commonly called *an energy*) $J(\mathbf{f})$ of the form

$$J(\mathbf{f}) = \Theta(H\mathbf{f} - \mathbf{g}) + \beta \Phi(\mathbf{f}). \tag{2}$$

In this expression, Θ forces closeness to data according to (1), Φ embodies the priors and $\beta > 0$ is a parameter that controls the tradeoff between these two terms. The most usual choice for Θ is

$$\Theta(\mathbf{v}) = \|\mathbf{v}\|_2^2. \tag{3}$$

Our computational method is designed for Θ as given in (3), or for any smooth convex function Θ. Since [29], data terms $\Theta(\mathbf{v}) = \|\mathbf{v}\|_1$ were shown to be useful if some data entries have to be preserved, which is appreciable for instance if $\mathbf{n}$ is impulse noise [30], [3] or in image decomposition [2], or in hybrid restoration methods [11]. Our method is easy to extend to the latter case using some ideas of [18].

A. Choice of Energy: Nonconvex Nonsmooth Regularization

In many image processing applications, the regularization term Φ reads

$$\Phi(\mathbf{f}) = \sum_{i \in I} \varphi(\|\mathrm{D}_i \mathbf{f}\|_2) \tag{4}$$

where I denotes the set of all pixels of the image $I = \{1, \ldots, p\} \mathrm{D}_i : \mathbb{R}^p \to \mathbb{R}^s$, for $s \geq 1$ is a linear operator yielding a vector containing the differences between pixel i and its neighbors, and $\varphi : \mathbb{R} \mapsto \mathbb{R}_+$ is called *a potential function* (PF). Each D_i can be seen as an $s \times p$ matrix. If $\{\mathrm{D}_i : i \in I\}$ corresponds to the forward discretization of the gradient operator, we have $\mathrm{D}_i \in \mathbb{R}^{2 \times p}$ and in particular, for every $i \in I$, shown in the equation at the bottom of the next page. For any $s \geq 1$ and for any $i \in I$, we define the $s \times p$ matrix D_i as given in the following:

$$\mathrm{D}_i = \begin{bmatrix} \mathrm{D}_i^1 \\ \cdots \\ \mathrm{D}_i^s \end{bmatrix}$$

Remark 1: When $\varphi(t) = |t|$ and $\{\mathrm{D}_i\}$ corresponds to the discrete analog of the gradient operator, (4) is the well-known Total Variation (TV) regularization function [37]. In a continuous setting, regularization involving $\|\nabla^k \mathbf{f}\|_2, k = 1, 2, \cdots$, is rotation invariant. In order to lighten the numerical intricacies
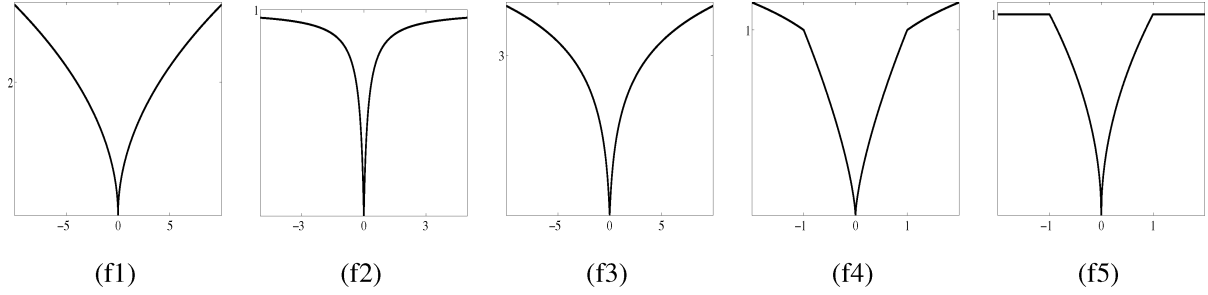
Fig. 1.   Potential functions from Table I (remind that φ is symmetric on $\mathbb{R}$).

relevant to the discrete variant $\|\mathrm{D}_i\mathbf{f}\|_2$, a common approach is to replace it by functions of the form

$$\Phi(\mathbf{f}) = \sum_{i \in I} \varphi\left(\left|\mathbf{d}_i^T\mathbf{f}\right|\right) \tag{5}$$

where the superscript $\cdot^T$ stands for transpose, and $\mathbf{d}_i \in \mathbb{R}^p$, for $i \in I$, yield e.g., the first-order differences between each pixel and its four or eight adjacent neighbors. Note that this kind of regularizer are customary in Markov random field modeling, e.g., the classical survey paper [5]. Even though rotation invariance is not well defined in the discrete setting, it is usually observed that the ℓ_2 norm of the discrete gradient, i.e., $\|\mathrm{D}_i\mathbf{f}\|_2$, yields image restorations of better quality than $|\mathbf{d}_i^T\mathbf{f}|$.

Various potential functions (PFs) φ have been used in the literature, a review can be found for instance in [6]. An important requirement is that φ allows the recovery of large differences $\|\mathrm{D}_i\mathbf{f}\|_2$ at the locations of edges and smooth the other differences. It is well known that this requirement cannot be met by $\varphi(t) = t^2$ which was originally used in [39]. Since the pioneering work of Geman & Geman [15], different nonconvex functions φ have been considered either in a statistical or in a variational framework, e.g., [5], [13], [14], [23], and [25]. In order to avoid the numerical intricacies arising with nonconvex regularization, since [17], [22], [38] in 1990, an important effort was done to derive *convex* edge-preserving PFs, see [1] for an excellent account. Nevertheless, *nonconvex nonsmooth regularization* offers much richer possibilities to restore high quality images with neat edges: for regularizer functions of the form (5) a theoretical explanation was provided in [31] while numerical examples can be found in numerous articles, e.g., [13], [14], [27], [33], and [36].

This paper provides two main contributions. The theoretical one is to prove that the minimizers of energies of the form (3)–(4), where $\varphi(\|\cdot\|_2)$ is nonconvex and nonsmooth at zero

and $\{\mathrm{D}_i\}$ is the discrete gradient, are composed of constant regions surrounded by closed contours and neat edges (a more general result is stated in Section II). This result is not a trivial extension of [31] which considers (5) and the case $\|\mathrm{D}_i\mathbf{f}\|_2$ for $\mathrm{D}_i \in \mathbb{R}^{s \times p}, s \geq 2$ cannot be handled in a similar way. The practical contribution is quite challenging: we derive fast algorithms to approximate faithfully the global minimizer of these nonconvex and nonsmooth energies (Section III). Our experimental results (Section IV) show clearly the effectiveness and efficiency of the proposed numerical schemes. Concluding remarks and perspectives are sketched in Section V.

II. ESSENTIAL PROPERTIES OF MINIMIZERS OF J

In this section, we study the properties of minimizers of J defined according to (2), (3) and (4) under customary, weak assumptions. We will assume the following:

H1: $\ker(H) \cap \ker(\mathrm{D}) = \{0\}$, where where $\mathrm{D} = [\mathrm{D}_1^T, \ldots, \mathrm{D}_p^T]^T$;

H2: φ is continuous and symmetric on $\mathbb{R}$, increasing on $\mathbb{R}_+$, with $\varphi(0) = 0$ and $\varphi'(0^+) > 0$;

H3: φ is $\mathcal{C}^2$ on $\mathbb{R}_+^* \setminus M$ where $M \not\ni 0$ is a finite set of points (possibly empty) such that if $t \in M$ then $\varphi'(t^-) > \varphi'(t^+)$ and $-\infty < \ddot{\varphi}(t^-) \overset{\text{def}}{=} \lim_{\tau \nearrow t} \varphi''(t) < \ddot{\varphi}(t^+) \overset{\text{def}}{=} \lim_{\tau \searrow t} \varphi''(t) \leq 0$;

H4: $\ddot{\varphi}(0^+) \overset{\text{def}}{=} \lim_{t \searrow 0} \varphi''(t) < 0, \varphi''$ is increasing with $\varphi''(t) \leq 0 \quad \forall t \in \mathbb{R}_+^* \setminus M$ and $\lim_{t \to \infty} \varphi''(t) = 0$.

Here we use the common notations

$$\mathbb{R}_+ \overset{\text{def}}{=} \{t \in \mathbb{R} : t \geq 0\} \quad \text{and} \quad \mathbb{R}_+^* \overset{\text{def}}{=} \{t \in \mathbb{R} : t > 0\}.$$

Examples of PFs φ satisfying these assumptions are given in Table I and plotted in Fig. 1. The restriction of φ'' on $\mathbb{R}_+^* \setminus M$ for all these functions in plotted in Fig. 2. Note that by H2, all terms of Φ such that $\|\mathrm{D}_i\mathbf{f}\|_2 = 0$ are nondifferentiable.

$$\begin{cases} i \notin \{n, 2n \ldots, mn\} & \Rightarrow \mathrm{D}_i^1[i] = -1, \mathrm{D}_i^1[i+1] = +1, \quad \mathrm{D}_i^1[j] = 0, \quad \forall j \notin \{i, i+1\} \\ \text{otherwise} & \Rightarrow \mathrm{D}_i^1[j] = 0, \quad \forall j = 1, \ldots, p \end{cases}$$

$$\begin{cases} i \notin \{(m-1)n+1, (m-1)n+2, \ldots, mn\} & \Rightarrow \mathrm{D}_i^2[i] = -1, \quad \mathrm{D}_i^2[i+n] = 1, \\ & \quad \mathrm{D}_i^2[j] = 0, \quad \forall j \notin \{i, i+n\} \\ \text{otherwise} & \Rightarrow \mathrm{D}_i^2[j] = 0 \quad \forall j = 1, \ldots, p \end{cases}$$
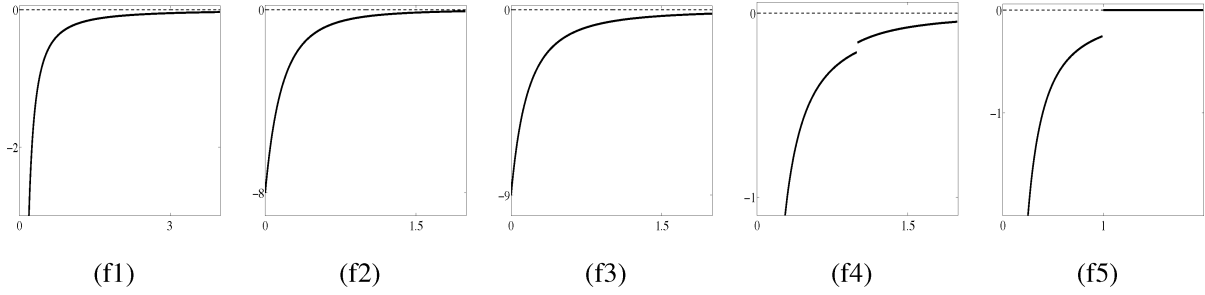
Fig. 2. Restriction of φ'' on $\mathbb{R}_+^* \setminus M \to \mathbb{R}$ for all the potential functions in Table I.

TABLE I
RESTRICTION ON $\mathbb{R}_+^*$ OF SEVERAL PFS YIELDING NONSMOOTH AND NONCONVEX REGULARIZATION AND SATISFYING H2, H3 AND H4. THE CONSTANT T, THE SET M AND THE NOTATION $\ddot{\varphi}$ ARE DEFINED IN (7) AND IN H3. NOTE THAT FOR (f4) AND (f5) WE HAVE $\varphi'(1^+) < \varphi i'(1^-)$ AND $\ddot{\varphi}(1^+) > \ddot{\varphi}(1^-)$

	$\varphi\|_{\mathbb{R}_+^*}$	$\varphi''\|_{\mathbb{R}_+^* \setminus M}$	T	M	$\ddot{\varphi}\|_{\mathbb{R}_+^*}(t^\pm),\ t \in M$
(f1)	$t^\alpha, \quad \alpha \in (0,1)$	$\alpha(\alpha-1)t^{\alpha-2}$	∞	$\varnothing$	
(f2)	$\dfrac{\alpha t}{1+\alpha t}$	$\dfrac{-2\alpha^2}{(1+\alpha t)^3}$	∞	$\varnothing$	
(f3)	$\log(\alpha t + 1)$	$\dfrac{-\alpha^2}{(1+\alpha t)^2}$	∞	$\varnothing$	
(f4)	$\begin{cases} t^\alpha,\ \alpha \in (0,1) & t < 1 \\ t^\gamma,\ \gamma \in (0,\alpha\min\{1,\frac{1-\alpha}{\alpha}\}) & t \geq 1 \end{cases}$	$\begin{cases} \alpha(\alpha-1)t^{\alpha-2} & t < 1 \\ \gamma(\gamma-1)t^{\gamma-2} & t \geq 1 \end{cases}$	∞	$\{1\}$	$\ddot{\varphi}(1^-) = \alpha(\alpha-1)$ $\ddot{\varphi}(1^+) = \gamma(\gamma-1)$
(f5)	$\begin{cases} t^\alpha,\ \alpha \in (0,1) & t < 1 \\ 1 & \text{else} \end{cases}$	$\begin{cases} \alpha(\alpha-1)t^{\alpha-2} & t < 1 \\ 0 & \text{else} \end{cases}$	1	$\{1\}$	$\ddot{\varphi}(1^-) = \alpha(\alpha-1)$ $\ddot{\varphi}(1^+) = 0$

The first theorem addresses PFs φ as those given in (f4) and (f5), Table I. Its proof can be found in Appendix A.

Theorem 1: Let J be of the form (2) along with (3) and (4) for $\beta > 0$ and all assumptions H1, H2, H3 and H4 hold. Given $\mathbf{g} \in \mathbb{R}^q$, let $\hat{\mathbf{f}}$ be any (local) minimizer of J. Then we have

$$\|D_i\hat{\mathbf{f}}\|_2 > 0 \Rightarrow \|D_i\hat{\mathbf{f}}\|_2 \notin M \quad (6)$$

where M is the set described in H3.

Given a PF φ satisfying H2, H3 and H4, we define $T \in \mathbb{R}_+^*$ by

$$t \geq T \Rightarrow \varphi''(T) = 0. \quad (7)$$

The next lemma is of great use to prove our main result, namely Theorem 2.

Lemma 1: Let φ satisfy the assumptions H2, H3, and H4. Consider the multifunction ξ in the following:

$$\xi: \ ]0,T] \mapsto \text{the subsets of } \mathbb{R},$$

$$\xi(t) = \begin{cases} \dfrac{\varphi''(t)}{t}, & \text{if } t \notin M \\ \left[\dfrac{\ddot{\varphi}(t^-)}{t}, \dfrac{\ddot{\varphi}(t^+)}{t}\right], & \text{if } t \in M \end{cases} \quad (8)$$

where M is the finite set in H3, $\ddot{\varphi}$ is defined in H3 and T is given in (7). Then

i) $\xi(t) < 0$ for all $t \in]0,T]$ and ξ is strictly increasing on $]0,T]$;

ii) $\forall\ c \in]0,\infty[$, there is a unique $\theta_c \in]0,T[$ such that $\xi(\theta_c) + c = 0$.

iii) The function $c \mapsto \theta_c \in]0,T[$ increases when c decreases on $]0,\infty[$.

The multifunction in (8) is illustrated on Fig. 3. The proof of the lemma can be found in Appendix B. Using these preliminary results, we can state a spectacular property of the (local or global) minimizers of the energy J, namely that if $\{D_i\}$ are discrete gradients, the minimizers $\hat{\mathbf{f}}$ are composed of constant regions surrounded by closed contours and neat edges higher than a bound $\theta > 0$. The solution $\hat{\mathbf{f}}$ is then a *segmented image, for any operator H* in (2)–(3). Note that different bounds θ can be derived under slightly different conditions. The proof of Theorem 2 is outlined in Appendix C.

Theorem 2: Let J be of the form (2) along with (3) and (4) for $\beta > 0$, and all assumptions H1, H2, H3, and H4 hold. Given $\mathbf{g} \in \mathbb{R}^q$, let $\hat{\mathbf{f}}$ be any (local) minimizer of J. Then we have

$$\text{either} \ \ \|D_i\hat{\mathbf{f}}\|_2 = 0 \ \ \text{or} \ \ \|D_i\hat{\mathbf{f}}\|_2 \geq \theta, \quad \forall i \in I \quad (9)$$

where $0 < \theta < T$ (see (7) for the definition of T) and the previously shown inequality is strict if $\theta \in M$. Put
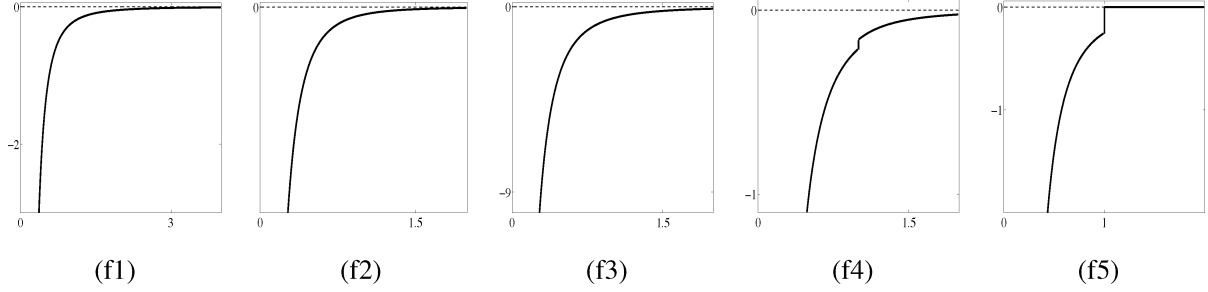
$$K = \#\{i \in I : \|D_i\hat{\mathbf{f}}\|_2 > 0\}.$$

(f1)   (f2)   (f3)   (f4)   (f5)

Fig. 3.   Multifunction ξ in (8) for each PF in Table I. Remind that for (f5) we have $T = 1$.

(a)   (b)
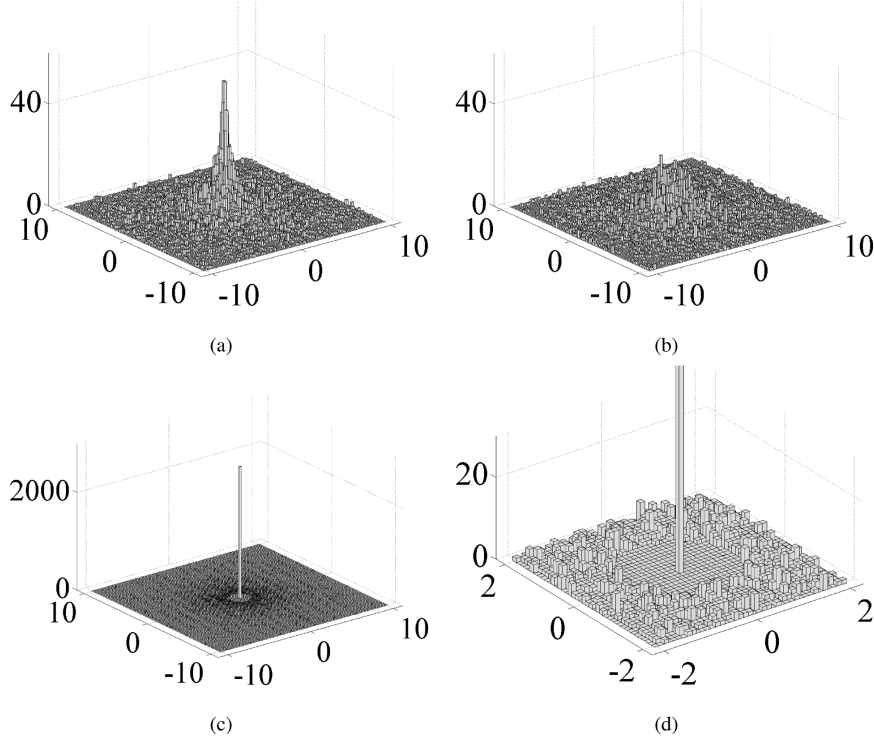
(c)   (d)

Fig. 4.   Histograms. (a) Samples of $\mathbf{f}$. (b) Observed samples of $\mathbf{g}$ where noise is added to $\mathbf{f}$. (c) Restored samples $\hat{\mathbf{f}}$. (d) Zoom of (c) near the origin.

If $K \geq 1$, the alternative (9) holds true for the unique θ that solves

$$\xi(\theta) = -\frac{2\,\alpha\,\|H\hat{\mathbf{f}}\|_2^2}{\beta K} \qquad (10)$$

where ξ is the multifunction in (8) and $\alpha \overset{\text{def}}{=} \max\{1, (\min\{\|\mathrm{D}_i\hat{\mathbf{f}}\|_2 > 0\})^{-3}\}$.

   i) If in addition we have: $\mathrm{rank}(H) = p$ or φ strictly increasing on $\mathbb{R}_+$, then (9) holds true for the unique θ that solves

$$\xi(\theta) = -\frac{2\,\alpha\,\|\mathbf{g}\|_2^2}{\beta K}. \qquad (11)$$

The proof of the theorem reveals that the bound θ is underestimated. This fact is observed in the following example.

Example 1: In order to illustrate Theorem 2, we consider a simple example where the underlying $\mathbf{f}$ is a point in $\mathbb{R}^2$, D and H are the identity matrices and $\varphi(t) = |t|^{1/2}$. We realized 10 000 independent trials where an original $\mathbf{f} \in \mathbb{R}^2$ is sampled from $p(\mathbf{f}) \propto \exp(-\lambda\varphi(\|\mathbf{f}\|_2))$ for $\lambda = 2$ and then $\mathbf{g} = \mathbf{f} + \mathbf{n}$ for $\mathbf{n} \sim \mathrm{Normal}(0, \sigma^2)$ with $\sigma = 0.8$. The histogram of all $\mathbf{f}$

and $\mathbf{g}$ are shown in Fig. 4(a) and (b). After this, the solution $\hat{\mathbf{f}}$ is calculated by minimizing $J(\mathbf{f}) = \|\mathbf{f} - \mathbf{g}\|^2 + \varphi(\|\mathbf{f}\|_2)$ for $\beta = 2\sigma^2\lambda$. For every $\mathbf{g} \neq 0$ the function J has two local minimizers, $\hat{\mathbf{f}}_1 = (0, 0)$ and $\hat{\mathbf{f}}_2$ satisfying $\|\hat{\mathbf{f}}_2\|_2 > \theta$ for $\theta \approx 0.47$; the global minimizer $\hat{\mathbf{f}}$ is found by exhaustive search over the x-y grid with the step size 0.01 in the $x-$ and $y-$ range between -10 and 10. The empirical histogram of all solutions $\hat{\mathbf{f}}$ is shown in Fig. 4(c) and zoom of $\hat{\mathbf{f}}$ in Fig. 4(d). These results illustrate the statement of the theorem: we have $\hat{\mathbf{f}} = 0$ in 27% of the trials while the smallest nonzero $\|\hat{\mathbf{f}}\|_2$ is $1.18 > \theta$.

III. MINIMIZATION METHODS

The minimization of nonconvex nonsmooth energy J given by (2), (3), and (4) involves three major difficulties that drastically restrict the methods that can be envisaged. Because of the nonconvexity of φ, J may exhibit a large number of local minima which are not global. In addition, J is usually nonsmooth at the minimizers and, thus, usual gradient-based methods are inappropriate even for local minimization. Finally, the matrix H can have numerous nonzero elements beyond the diagonal and is often ill-conditioned. Given that our problem is high dimensional (p is typically more than 26×10^4), global

minimization of J can be considered using either stochastic algorithms or continuation-based deterministic methods.

A. Comments on Preexisting Methods

Since [21], asymptotically convergent global minimization of nonconvex functions has been conducted using stochastic schemes, such as simulated annealing or Metropolis annealing. However, the computational cost of such algorithms is prohibitive when H is not a diagonal matrix, whereas in general image restoration and reconstruction problems H has a large number of non zero coefficients outside the diagonal and can also be a dense matrix. Robini *et al.* in [35] studied inexpensive acceleration techniques that do not alter the theoretical convergence properties of annealing algorithms. They employed restriction of the state space to a locally bounded image space and increasing concave transform of the energy to speed up the convergence. However, the numerous new parameters required for acceleration are very tricky to handle. Recently, Robini *et al.* [36] introduced a new class of hybrid algorithms that combine simulated annealing with deterministic continuation. Numerical experiments have shown that this approach outperforms standard simulated annealing. Nevertheless, it still requires quite a high computational effort. Moreover, stochastic algorithms cannot yield solutions that incorporate one of the main properties of nonsmooth regularization, namely (9) stated in Theorem 2 (e.g., recovering of truly constant regions).

According to [41], the idea of continuation is a good deterministic alternative to deal with nonconvex energies J. Even though there is no guarantee for global convergence, extensive experiments have shown that for a finite number of iterations the *graduated nonconvexity (GNC)* approach [7] leads to minimizers $\hat{\mathbf{f}}$ having a lower (hence, better) energy than simulated annealing, see for instance [8]. Extensions of the GNC approach were done, e.g., in [4] and [28]. Some conditions to improve the convergence of GNC for a broad class of nonconvex energies where H can also be a dense or ill-conditioned matrix have been investigated in [27]. However, in these GNC methods the energy J is tracked using a sequence of *smooth* approximates of J, so they cannot properly address the nonsmoothness of J in our problem.

In [33], a *nonsmooth* GNC continuation method is inaugurated to solve a nonconvex nonsmooth minimization problem where J is of the form (2), (3) and (5).

B. Our Approach

Here our goal is to conceive nonsmooth GNC schemes for J of the form given by (2), (3) and (4). The main difficulty with respect to [33] is that now the PF φ is applied to $\|\mathrm{D}_i\mathbf{f}\|_2$.

Consider a sequence

$$\varepsilon_0 = 0 < \varepsilon_1 < \cdots < \varepsilon_k < \cdots < \varepsilon_n = 1. \quad (12)$$

We approach φ by a sequence of $\varphi_\varepsilon : \mathbb{R} \to \mathbb{R}_+$ such that φ_0 is convex and φ_ε monotonously reaches φ when ε goes from 0 to 1 in (12), with $\varphi_1 = \varphi$ and φ_ε *is nonsmooth at 0 for any* $\varepsilon \in [0,1]$. (To simplify the notations, we write ε for ε_k whenever

this is clear from the context.) Correspondingly, our energy J is approximated by a sequence J_ε as given in the following:

$$J_\varepsilon(\mathbf{f}) = \|H\mathbf{f} - \mathbf{g}\|_2^2 + \beta \sum_{i \in I} \varphi_\varepsilon(\|\mathrm{D}_i\mathbf{f}\|_2), \quad 0 \le \varepsilon \le 1. \quad (13)$$

Thus, J_0 is convex (and nonsmooth), J_ε monotonously goes to J when ε increases and we have $J_1 = J$.

For computational reasons, we assume two *additional* hypotheses.

H5: $\varphi'(0^+) > 0$ and $\ddot{\varphi}(0^+) \stackrel{\text{def}}{=} \lim_{t \searrow 0} \varphi''(t) < 0$ are finite.
Given the result (6) in Theorem 1, we consider that

H6: the set M in H3 is empty.

Using H1–H4 together with H5–H6, we can choose $\varphi_\varepsilon : \mathbb{R} \to \mathbb{R}$ such that

(a) $\varphi_\varepsilon(0) = 0, \varphi_\varepsilon$ is symmetric and continuous on $\mathbb{R}$, and $\mathcal{C}^2$ − smooth on $\mathbb{R}_+^*, \quad \forall \varepsilon \in [0,1]$

(b) for $\varepsilon = 0$ we have $\varphi_0(t) = \varphi(0^+)|t|$ while for $\varepsilon = 1, \varphi_1 = \varphi$; $\quad (14)$

(c) $\varphi_\varepsilon'(0^+) > 0$ is finite $\quad \forall \varepsilon \in [0,1]$;

(d) $\varphi_\varepsilon''(t) < 0, \quad \forall t \in \mathbb{R} \setminus \{0\}$
and
$\ddot{\varphi}(0^+) \stackrel{\text{def}}{=} \lim_{t \searrow 0} \varphi_\varepsilon''(t) < 0$ is finite $\quad \forall \varepsilon \in]0,1]$.

We can further rewrite φ_ε as follows:

$$\varphi_\varepsilon(t) = \psi_\varepsilon(t) + \alpha_\varepsilon|t| \quad \text{where} \quad \alpha_\varepsilon = \varphi_\varepsilon'(0^+). \quad (15)$$

The next lemma is needed to derive the numerical methods. Its proof is given in Appendix D.

Lemma 2: Let φ_ε satisfy all conditions in (14). Consider the function $\psi_\varepsilon(t) : \mathbb{R} \to \mathbb{R}$

$$\psi_\varepsilon(t) = \varphi_\varepsilon(t) - \varphi_\varepsilon'(0^+)|t|. \quad (16)$$

Then we have the following:
a) $\forall t \in \mathbb{R}$, we have $\psi_0(t) = 0$ and $\psi_1(t) = \varphi(t) - \varphi'(0^+)|t|$;
b) ψ_ε is $\mathcal{C}^1$ on $\mathbb{R}$ with $\psi_\varepsilon'(0) = 0, \quad \forall \varepsilon \in [0,1]$;
c) ψ_ε is $\mathcal{C}^2$ on $\mathbb{R}$ and $\psi_\varepsilon''(-t) = \psi_\varepsilon''(t) < 0, \quad \forall t \in \mathbb{R}, \quad \forall \varepsilon \in]0,1]$.

If φ satisfies H1–H4 (p. 4) and H5–H6, an easy way to construct φ_ε is

$$\varphi_\varepsilon(t) = \varepsilon\varphi(t) + \varphi'(0^+)(1 - \varepsilon)|t|. \quad (17)$$

One can check that φ_ε satisfies all conditions in (14) and, hence, Lemma 2 holds true. By Lemma 2, we see that φ_ε in (15) is composed of two terms: the first one ψ_ε is $\mathcal{C}^2$-smooth and concave whereas the second one $\alpha_\varepsilon|t|$ is convex and nonsmooth at zero.

Decomposing J_ε in (13) according to (15) yields

$$J_\varepsilon(\mathbf{f}) = \|H\mathbf{f} - \mathbf{g}\|_2^2 + \beta\Psi_\varepsilon(\mathbf{f}) + \beta\alpha_\varepsilon \sum_{i \in I} \|\mathrm{D}_i\mathbf{f}\|_2$$

where $\quad \Psi_\varepsilon(\mathbf{f}) = \sum_{i \in I} \psi_\varepsilon(\|\mathrm{D}_i\mathbf{f}\|_2). \quad (18)$

Remark 2: By using the results in Lemma 2, the function: $\mathbf{f} \mapsto \|H\mathbf{f} - \mathbf{g}\|_2^2 + \beta\Psi_\varepsilon(\mathbf{f})$ is *twice differentiable[1] and nonconvex* whereas the function: $\mathbf{f} \mapsto \sum_{i \in I}\|\mathrm{D}_i\mathbf{f}\|_2$ is *convex and nondifferentiable.*

In the usual case when $\{\mathrm{D}_i\}$ are discrete gradients, the term $\sum_{i \in I}\|\mathrm{D}_i\mathbf{f}\|_2$ amounts to the TV regularization: by a slight restriction of language, in what follows we call it *the* TV *term* and write

$$\mathrm{TV}(\mathbf{f}) \stackrel{\text{def}}{=} \sum_{i \in I}\|\mathrm{D}_i\mathbf{f}\|_2. \tag{19}$$

Note that in [33] that $\{\|\mathrm{D}_i\mathbf{f}\|_2\}$ are approximated by $\{|\mathbf{d}_i^T\mathbf{f}|\}$ where $\mathbf{d}_i^T : \mathbb{R}^p \to \mathbb{R}$, for $i \in I$, are linear operators (e.g., the first-order differences between each pixel and its four or eight adjacent neighbors). Then the energy minimization can be reformulated as a nonconvex smooth optimization problem under linear constraints and solved using linear programming techniques.[2] Actually, the energy J_ε in (18) cannot be reformulated in a similar way because of the ℓ_2 norm in TV term (19).

Our approach to tackle the difficulties for minimizing the function J_ε in (18) is to apply variable-splitting and penalty techniques to separate the nonconvex term and the nonsmooth term using additional variables. In the following, we propose and compare two different numerical schemes to minimize J_ε in (18) for every $\varepsilon \in [0, 1]$. The minimizer $\hat{\mathbf{f}}$ of J_1 provides the sought-after approximation of the global minimizer of J.

C. Numerical Scheme Based Upon Fitting to $\mathbf{f}$

Following the idea of [19], an auxiliary variable $\mathbf{u} \in \mathbb{R}^p$ is used to transfer the nonsmooth TV term from $\mathbf{f}$. More precisely, a quadratic fitting term, weighted by $\omega > 0$, is added to J_{ε_k} in (18) in order to ensure the closeness of $\mathbf{f}$ and $\mathbf{u}$. The resultant augmented energy $\mathcal{J}_{\varepsilon_k} : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ reads

$$\mathcal{J}_{\varepsilon_k}(\mathbf{f}, \mathbf{u}) = \|H\mathbf{f} - \mathbf{g}\|_2^2 + \beta\Psi_{\varepsilon_k}(\mathbf{f}) \\ + \beta\alpha_{\varepsilon_k}\mathrm{TV}(\mathbf{u}) + \omega\|\mathbf{f} - \mathbf{u}\|_2^2. \tag{20}$$

In the algorithm, ω can be gradually increasing in the iterations in order to force that $\mathbf{f}$ is close $\mathbf{u}$. By fixing the variable $\mathbf{u}$, $\mathcal{J}_{\varepsilon_k}(., \mathbf{u})$ is a twice differentiable function with respect to $\mathbf{f}$ so that it can be minimized by gradient-based methods. For $\mathbf{f}$ fixed, minimizing $\mathcal{J}_{\varepsilon_k}(\mathbf{f}, .)$ with respect to $\mathbf{u}$ amounts to a TV denoising problem which can be solved efficiently by already existing method such as the Chambolle's algorithm [9].

Let us state the proposed method in detail. For each ε_k [cf. (12)], we initialize with $\mathbf{u}^{(0,k)} \stackrel{\text{def}}{=} \mathbf{u}_{\varepsilon_{k-1}}$ where $\mathbf{u}_{\varepsilon_{k-1}}$ results from the minimization of $\mathcal{J}_{\varepsilon_{k-1}}$. Then the proposed method computes a sequence of iterates $\mathbf{f}^{(1,k)}, \mathbf{u}^{(1,k)}, \mathbf{f}^{(2,k)}, \mathbf{u}^{(2,k)} \dots, \mathbf{f}^{(j,k)}, \mathbf{u}^{(j,k)}, \cdots,$ for each $k = 0, 1, \dots, n$, according to

$$\mathbf{f}^{(j,k)} = \arg\min_{\mathbf{f} \in \mathbb{R}^p} \mathcal{J}_{\varepsilon_k}\left(\mathbf{f}, \mathbf{u}^{(j-1,k)}\right)$$

[1]Note that by Lemma 2(ii) we have $\nabla\psi_\varepsilon(\|\mathrm{D}_i\mathbf{f}\|_2) = 0$ if $\|\mathrm{D}_i\mathbf{f}\|_2 = 0$.

[2] The experimental results in [33] showed that the resultant method provides better performance with significantly smaller computational cost, compared to a simulated annealing method.

$$= \arg\min_{\mathbf{f} \in \mathbb{R}^p}\left\{\|H\mathbf{f} - \mathbf{g}\|_2^2 + \beta\Psi_{\varepsilon_k}(\mathbf{f}) \right. \\ \left. + \omega\|\mathbf{f} - \mathbf{u}^{(j-1,k)}\|_2^2\right\}$$

$$\mathbf{u}^{(j,k)} = \arg\min_{\mathbf{u} \in \mathbb{R}^p} \mathcal{J}_{\varepsilon_k}\left(\mathbf{f}^{(j,k)}, \mathbf{u}\right) \tag{21}$$

$$= \arg\min_{\mathbf{u} \in \mathbb{R}^p}\left\{\omega\|\mathbf{f}^{(j,k)} - \mathbf{u}\|_2^2 + \beta\alpha_{\varepsilon_k}\mathrm{TV}(\mathbf{u})\right\}. \tag{22}$$

By Lemma 2(i), for $\varepsilon_0 = 0$, we have $\Psi_0(\mathbf{f}) = 0$ for all $\mathbf{f}$, so minimizing $\mathcal{J}_0$ in the previously mentioned two steps is exactly the same as the TV deblurring problem in [19].

1) Computation of $\mathbf{f}^{(j,k)}$ According to (21): For $\varepsilon_0 = 0$ (i.e., $k = 0$), the finding of $\mathbf{f}^{(j,0)}$ amounts to minimize the convex quadratic function

$$\mathcal{J}_0\left(\mathbf{f}, \mathbf{u}^{(j-1,0)}\right) = \|H\mathbf{f} - \mathbf{g}\|_2^2 + \omega\left\|\mathbf{f} - \mathbf{u}^{(j-1,0)}\right\|_2^2. \tag{23}$$

It is equivalent to solving a linear system

$$(H^T H + \omega I)\mathbf{f}^{(j,0)} = H^T\mathbf{g} + \omega\mathbf{u}^{(j-1,0)}. \tag{24}$$

For $\varepsilon_k > 0$ (i.e., $k > 0$), the Quasi-Newton method [34, Ch.8] can be used to solve (21). Since all the terms in $\mathcal{J}_{\varepsilon_k}(., \mathbf{u}^{(j-1,k)})$ are twice differentiable (see Remark 2), we can find out the corresponding gradient vector $\nabla_{\mathbf{f}}\mathcal{J}_{\varepsilon_k} \stackrel{\text{def}}{=} \nabla_{\mathbf{f}}\mathcal{J}_{\varepsilon_k}(\mathbf{f}, \mathbf{u}^{(j-1,k)})$ and the Hessian $\nabla_{\mathbf{f}}^2\mathcal{J}_{\varepsilon_k} \stackrel{\text{def}}{=} \nabla_{\mathbf{f}}^2\mathcal{J}_{\varepsilon_k}(\mathbf{f}, \mathbf{u}^{(j-1,k)})$ of $\mathcal{J}_{\varepsilon_k}(., \mathbf{u}^{(j-1,k)})$ to tackle the minimization problem

$$\nabla_{\mathbf{f}}\mathcal{J}_{\varepsilon_k} = 2H^T(H\mathbf{f} - \mathbf{g}) + \beta\nabla\Psi_{\varepsilon_k}(\mathbf{f}) \\ + 2\omega\left(\mathbf{f} - \mathbf{u}^{(j-1,k)}\right) \tag{25}$$

and

$$\nabla_{\mathbf{f}}^2\mathcal{J}_{\varepsilon_k} = 2H^T H + 2\omega I + \beta\nabla_{\mathbf{f}}^2\Psi_{\varepsilon_k}(\mathbf{f}) \tag{26}$$

where $\nabla_{\mathbf{f}}^2\Psi_{\varepsilon_k}(\mathbf{f}) \stackrel{\text{def}}{=} \nabla_{\mathbf{f}}^2\Psi_{\varepsilon_k}(\mathbf{f}, \mathbf{u}^{(j-1,k)})$ of $\Psi_{\varepsilon_k}(., \mathbf{u}^{(j-1,k)})$. Since $\nabla_{\mathbf{f}}^2\Psi_{\varepsilon_k}(\mathbf{f})$ is negative definite [see Lemma 2(iii)], the Hessian $\nabla_{\mathbf{f}}^2\mathcal{J}_{\varepsilon_k}$ may be not positive definite. This may prevent the Quasi-Newton method from convergence as as the resultant search direction may not be a descent direction. In order to ensure the descent direction, we simply use the positive definite part of the Hessian matrix in the optimization procedure. Such procedure can guarantee that the proposed algorithm is a descent method for the minimization problem. Thanks to the term $2\omega I$, the coefficient matrix $2H^T H + 2\omega I$ is always positive definite even if $H^T H$ is singular. The solution can be updated by

$$\mathbf{f}^{(j,k)} = \mathbf{f}^{(j-1,k)} + \tau\Delta\mathbf{f}^{(j-1,k)}$$

where $\tau > 0$ is the step-size and $\Delta\mathbf{f}^{(j-1,k)}$ is found by solving

$$(2H^T H + 2\omega I)\Delta\mathbf{f}^{(j-1,k)} = -\nabla_{\mathbf{f}}\mathcal{J}_{\varepsilon_k}. \tag{27}$$

We remark in image restoration that H is usually a blurring matrix generated by a symmetric point spread function. The computational cost of the method is dominated by three fast discrete transforms in solving the linear system in (24) or (27), see [26]. The computational cost for each fast transform is only $O(p\log p)$ for a $p \times p$ blurring matrix H [26].

Three different strategies to determine the step-size τ were tested: Armijo rule, Goldstein rule and a fixed τ ([34], Ch. 3). By observing experimental results, we found out that the numerical schemes based upon these three rules converged to the same solutions, while using the first two rules required heavy additional computation cost. Therefore, we fixed $\tau = 1$ for all of our experiments.

2) Computation of $\mathbf{u}^{(j,k)}$ According to (22): The second step of this method is to apply an exact TV denoising problem scheme to $\mathbf{f}^{(j,k)}$. Since the function $\mathbf{u} \mapsto \mathcal{J}_{\varepsilon_k}(\mathbf{u}, \mathbf{f}^{(j,k)})$ is identically the same as that in [19], we employ the Chambolle's projection algorithm [9] to solve this problem.

3) Algorithm I: For each ε_k, we simply choose a linear increase ε_k [cf. (12)] in our experiments as suggested in [33]. The full algorithm is given in the following.

Set $\varepsilon_0 = 0$ and $\Delta\varepsilon = 1/n$, and initialize $\mathbf{u}^{(0,0)}$.

For $k = 0 \to n$

Set $j = 1$, initial value of ω, and relative-change $= \text{tol} + 1$

While relative-change $> \text{tol}$ do

If $k = 0$

Solve $(H^T H + \omega I)\mathbf{f}^{(j,k)} = H^T \mathbf{g} + \omega \mathbf{u}^{(j-1,k)}$;

Otherwise

Solve $(2H^T H + 2\omega I)\Delta\mathbf{f}^{(j-1,k)} = -\nabla_{\mathbf{f}}\mathcal{J}_{\varepsilon_k}$;

Update $\mathbf{f}^{(j,k)} = \mathbf{f}^{(j-1,k)} + \tau\Delta\mathbf{f}^{(j-1,k)}$;

End If;

Minimize $\mathbf{u} \to \mathcal{J}_{\varepsilon_k}(\mathbf{f}^{(j,k)}, \mathbf{u})$ to obtain $\mathbf{u}^{(j,k)}$ using the Chambolle's method [9];

Compute relative-change $= \|\mathbf{u}^{(j,k)} - \mathbf{u}^{(j-1,k)}\|_2 / \|\mathbf{u}^{(j,k)}\|_2$

Increase ω and set $j = j + 1$;

End While

Set $\mathbf{u}^{(0,k+1)} = \mathbf{u}^{(j,k)}$ (for the initial guess of the next outer loop);

Set $\varepsilon_{k+1} = \varepsilon_k + \Delta\varepsilon$;

End For

D. Numerical Scheme Based on Fitting to $D\mathbf{f}$

Here we derive a different approach to minimize (18). It is based on variable-splitting and penalty technique to transfer the nonsmooth term out of $\mathcal{J}_{\varepsilon_k}$ in such a way that the TV denoising step is avoided with an aditional s-dimensional shrinkage operation, as proposed in [40]. To this end, we consider an another augmented energy $\mathcal{J}_{\varepsilon_k} : \mathbb{R}^p \times \mathbb{R}^{sp} \to \mathbb{R}$ which involves a fitting of the auxiliary variable $\mathbf{u} \in \mathbb{R}^{sp}$ to $D\mathbf{f}$

$$\mathcal{J}_{\varepsilon_k}(\mathbf{f}, \mathbf{u}) = \|H\mathbf{f} - \mathbf{g}\|_2^2 + \beta\Psi_{\varepsilon_k}(\mathbf{f}) + \omega\|D\mathbf{f} - \mathbf{u}\|_2^2 + \beta\alpha_{\varepsilon_k}\sum_{i \in I}\|\mathbf{u}_i\|_2 \quad (28)$$

where $\omega > 0$ and $\mathbf{u}_i \in \mathbb{R}^s$, $\forall i \in I$. For $\mathbf{f}$ fixed, $\mathcal{J}_{\varepsilon_k}(\mathbf{f}, .)$ is convex and nondifferentiable. Given $\mathbf{u}$, the function $\mathbf{f} \mapsto \mathcal{J}_{\varepsilon_k}(\mathbf{f}, \mathbf{u})$ is twice differentiable (see Remark 2) and nonconvex so that it can be minimized by gradient-based methods similar to Algorithm I, see Section III-C. The computational steps analogous to (22) and (21) are given as follows:

$$\mathbf{u}^{(j,k)} = \arg\min_{\mathbf{u} \in \mathbb{R}^{sp}} \mathcal{J}_{\varepsilon_k}\left(\mathbf{f}^{(j-1,k)}, \mathbf{u}\right)$$

$$= \arg\min_{\mathbf{u} \in \mathbb{R}^{sp}}\left\{\sum_{i \in I}\left(\omega\|D_i\mathbf{f}^{(j-1,k)} - \mathbf{u}_i\|_2^2\right.\right.$$

$$\left.\left. + \beta\alpha_\varepsilon\|\mathbf{u}_i\|_2\right)\right\} \quad (29)$$

$$\mathbf{f}^{(j,k)} = \arg\min_{\mathbf{f} \in \mathbb{R}^p} \mathcal{J}_\varepsilon\left(\mathbf{f}, \mathbf{u}^{(j,k)}\right)$$

$$= \arg\min_{\mathbf{f} \in \mathbb{R}^p}\left\{\|H\mathbf{f} - \mathbf{g}\|_2^2 + \beta\Psi_\varepsilon(\mathbf{f})\right.$$

$$\left. + \omega\|D\mathbf{f} - \mathbf{u}^{(j,k)}\|_2^2\right\}. \quad (30)$$

In this case, we initialize with $\mathbf{f}^{(0,k)} \stackrel{\text{def}}{=} \mathbf{f}_{\varepsilon_{k-1}}$ where $\mathbf{f}_{\varepsilon_{k-1}}$ results from the minimization of $\mathcal{J}_{\varepsilon_{k-1}}$.

1) Computation of $\mathbf{u}^{(j,k)}$ According to (29): Solving (29) amounts to solve p independent problems

$$\mathbf{u}_i^{(j,k)} = \arg\min_{\mathbf{u} \in \mathbb{R}^s}\left\{\omega\left\|D_i\mathbf{f}^{(j-1,k)} - \mathbf{u}_i\right\|_2^2 + \beta\alpha_\varepsilon\|\mathbf{u}_i\|_2\right\},$$

$$\forall i \in I. \quad (31)$$

As shown in ([40], pp. 251–252), each one of the problems in (31) can be solved efficiently using s-dimensional shrinkage

$$\mathbf{u}_i^{(j,k)} = \frac{D_i\mathbf{f}^{(j-1,k)}}{\|D_i\mathbf{f}^{(j-1,k)}\|}\max\left\{\left\|D_i\mathbf{f}^{(j-1,k)}\right\|_2 - \frac{\beta\alpha_{\varepsilon_k}}{2\omega}, 0\right\},$$

$$\forall i \in I. \quad (32)$$

2) Computation of $\mathbf{f}^{(j,k)}$ According to (30): The task is similar to the computation of $\mathbf{f}^{(j,k)}$ in Algorithm I.

For $\varepsilon_0 = 0$ ($k = 0$), the finding of $\mathbf{f}^{(j,0)}$ amounts to minimize the convex quadratic function in (23). For $\varepsilon_k > 0$, the Quasi-Newton method can be used to solve (30). Since all the terms in $\mathcal{J}_\varepsilon(., \mathbf{u}^{(j-1,k)})$ are twice differentiable (see Remark 2), we can find out the corresponding gradient vector [cf. (25)] and the Hessian matrix [cf. (26)] to tackle the minimization problem. Again, in order to ensure the descent direction, we simply use the positive definite part of the Hessian matrix $2H^T H + 2\omega\sum_{i \in I}D_i^T D_i$ in the optimization procedure. Such matrix is nonsingular by H1 and it can be diagonalized by the discrete transform matrix proposed in [26]. Now the solution can be updated by $\mathbf{f}^{(j,k)} = \mathbf{f}^{(j-1,k)} + \tau\Delta\mathbf{f}^{(j-1,k)}$ where $\Delta\mathbf{f}^{(j-1,k)}$ is solved by

$$\left(2H^T H + 2\omega\sum_{i \in I}D_i^T D_i\right)\Delta\mathbf{f}^{(j-1,k)} = -\nabla_{\mathbf{f}}\mathcal{J}_{\varepsilon_k} \quad (33)$$

efficiently in $O(p\log p)$ for an image with p pixels. By observing experimental results for different rules of step size, we also fixed $\tau = 1$ for all of our experiments
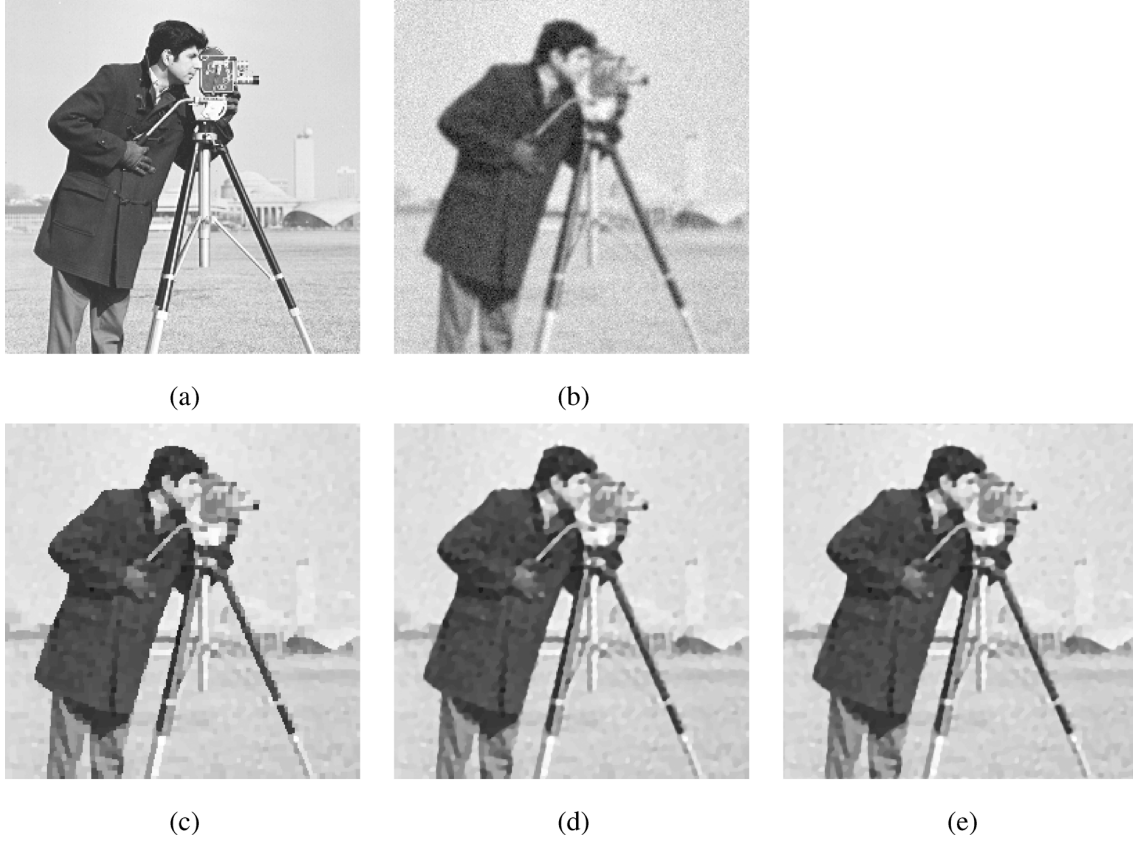
3) Algorithm II:

Fig. 5. (a) Original Cameraman image. (b) Observed image. (c) Image restored by the method in [33] with $\beta = 0.03$. (d)–(e) Images restored by Algorithms I and II, respectively, with $\beta = 0.03$.

- Set $\varepsilon_0 = 0$ and $\Delta\varepsilon = 1/n$, and initialize $\mathbf{f}^{(0,0)}$.

For $k = 0 \rightarrow n$

Set $j = 1$, initial value of ω, and relative-change $= \text{tol} + 1$

While relative-change $> \text{tol}$ do

Obtain $\mathbf{u}^{(j,k)}$ by computing the formula in (32);

If $k = 0$

Solve
$$(H^T H + \omega \sum_{i \in I} \mathrm{D}_i^T \mathrm{D}_i)\mathbf{f}^{(j,k)} = H^T \mathbf{g} + \omega \sum_{i \in I} \mathrm{D}_i^T \mathbf{u}^{(j,k)};$$

Otherwise

Solve $(2H^T H + 2\omega \sum_{i \in I} \mathrm{D}_i^T \mathrm{D}_i)\Delta\mathbf{f}^{(j-1,k)} = -\nabla_{\mathbf{f}}\mathcal{J}_{\varepsilon_k}$;

Update $\mathbf{f}^{(j,k)} = \mathbf{f}^{(j-1,k)} + \Delta\mathbf{f}^{(j-1,k)}$;

End If;

Compute relative-change $= \|\mathbf{f}^{(j,k)} - \mathbf{f}^{(j-1,k)}\|_2 / \|\mathbf{f}^{(j,k)}\|_2$;

Increase ω and set $j = j + 1$;

End While

Set $\mathbf{f}^{(0,k+1)} = \mathbf{f}^{(j,k)}$ (for the initial guess of the next outer loop);

Update $\varepsilon_{k+1} = \varepsilon_k + \Delta\varepsilon$;

End For

IV. EXPERIMENTAL RESULTS

In this section, we present the experimental results to demonstrate the efficiency of the algorithms proposed in Sections III-C-3 and III-D-3 and compare their performance. The codes are available at http://www.math.hkbu.hk/~mng/imaging-software.html. Signal to noise ratio (SNR) is used to measure the quality of the restored images while CPU time is also used to compare the efficiency of the restoration method. The parameter tol is set to be 10^{-4} in the two proposed methods. The step size used in the Chambolle's method is 0.25. The initial value of ω is set to be 1.1, and its value is updated by 1.8ω at each iteration. The PF used in all the illustrations was also tested in [33]

$$\varphi(t) = \frac{\alpha|t|}{1 + \alpha|t|}, \quad \varphi_\varepsilon(t) = \frac{\alpha|t|}{1 + \varepsilon\alpha|t|}, \quad 0 \le \varepsilon \le 1. \quad (34)$$

Note that φ satisfies H2–H6 and that φ_ε satisfies (14). We compare the two proposed methods with the method in [33] where the outer interior point method and the inner conjugate gradient method are employed. Now we adopt an augmented approach to minimize the nonsmooth nonconvex problem in (18) and consider two schemes described in Sections III-C and III-D. The
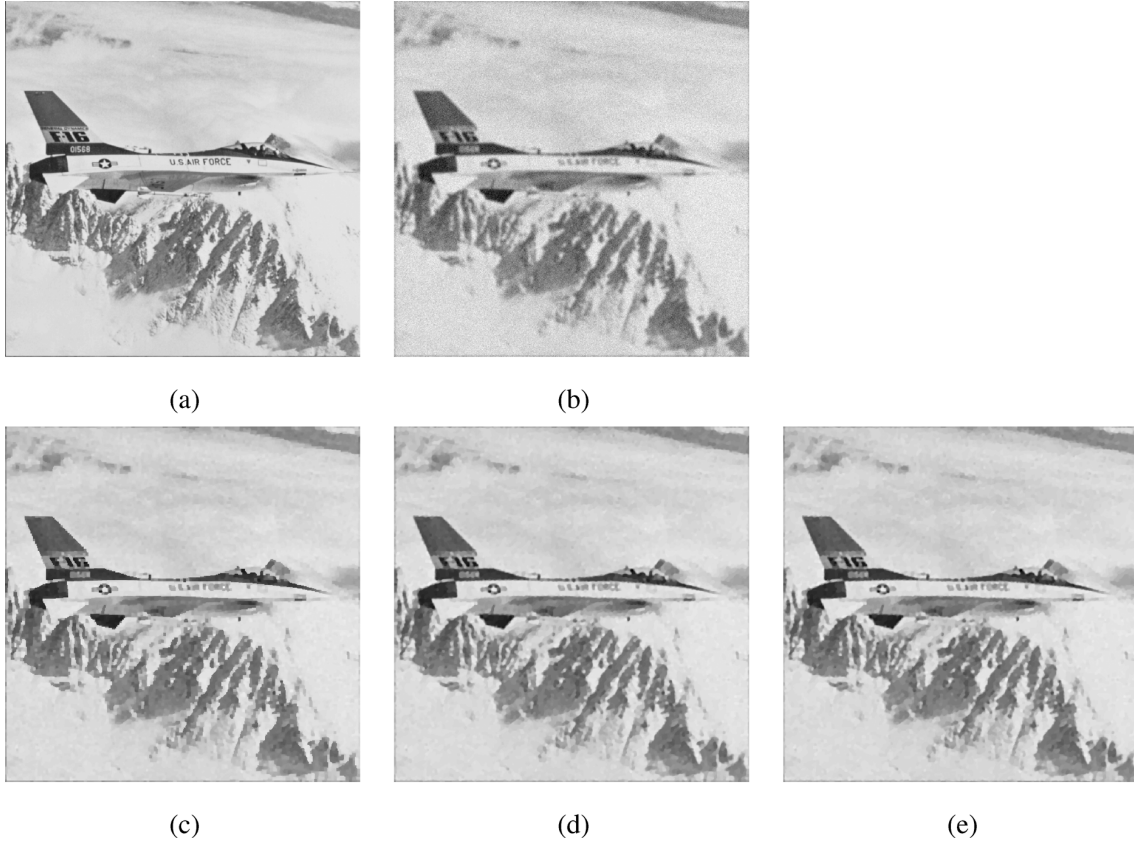
Fig. 6. (a) Original F16 image. (b) Observed image. (c) Image restored by the method in [33] with $\beta = 0.04$. (d)–(e) Images restored by Algorithms I and II, respectively, with $\beta = 0.04$.

linear systems in (27) and (33) are solved by fast discrete transforms in the two proposed methods. All the computational tasks are performed using MATLAB on a computer with Corel(TM)2 CPU with 2.66 GHz and 1.98 GB of RAM.

Five gray-value images are tested with intensity ranging from 0 to 1. The first image is Cameraman of size 256×256. The second and third images are F16 and Tank of size 512×512. To generate the observed images, we added Gaussian noise with the standard deviation of 0.05 with blurring. The blurring function is chosen to be a 2-D truncated Gaussian function

$$h(s,t) = \exp\left(\frac{-s^2 - t^2}{2\sigma^2}\right), \quad \text{for} -3 \le s, t \le 3$$

with $\sigma = 1.5$. The fourth image is the modified Shepp–Logan image of size 50×50. The fifth image is the modified Shepp–Logan image of size 1000×1000. We use this large image to demonstrate the efficiency of the proposed method. To generate the observed images for these two images, we added Gaussian noise with the standard deviation of 0.05. Radon transform is used to construct the degradation matrix H. These images are further transformed by back-projection so that H can be reformulated as a convolution operator.

Different initial guesses have been considered, including the observed image, the least squares solution and a flat image (all the pixel values are 0.5). From our experimental results, both of the proposed methods were insensitive to all of the initial

guesses. Therefore, we only demonstrate the results which the initial guesses are the observed images. For φ_ε in (34) we have

$$\alpha_\varepsilon = \varphi_\varepsilon(0^+) = \varphi(0^+) = \alpha.$$

In the experiments we used $\alpha = \alpha_\varepsilon = 0.5$. We tested different values of β in order to find out the restored image with the highest SNR among the tested values. Similarly, we also tested different values of the regularization parameter in the modified interior point method [33] to find out the restored image with the highest SNR.

A. Test of Blurred and Noisy Images

Figs. 5–7(a) show the original images. Figs. 5–7(b) show their corresponding images with blur and noise as described in the previously mentioned settings, respectively. Figs. 5–7(c) and 5–7(d)–(e) show the images restored by the method proposed in [33], and Algorithms I and II, respectively. For simplicity, we use the same set of parameters in Figs. 5–7(c) for restoring images in Figs. 5–7(d)–(e) by the proposed methods. We see from the figures that the images restored by the method proposed in [33] and the two proposed methods are visually about the same. In Table II, we show their SNR results, and find that they are about the same. However, the computational time (in seconds) required by Algorithms I and II is significantly lower than the
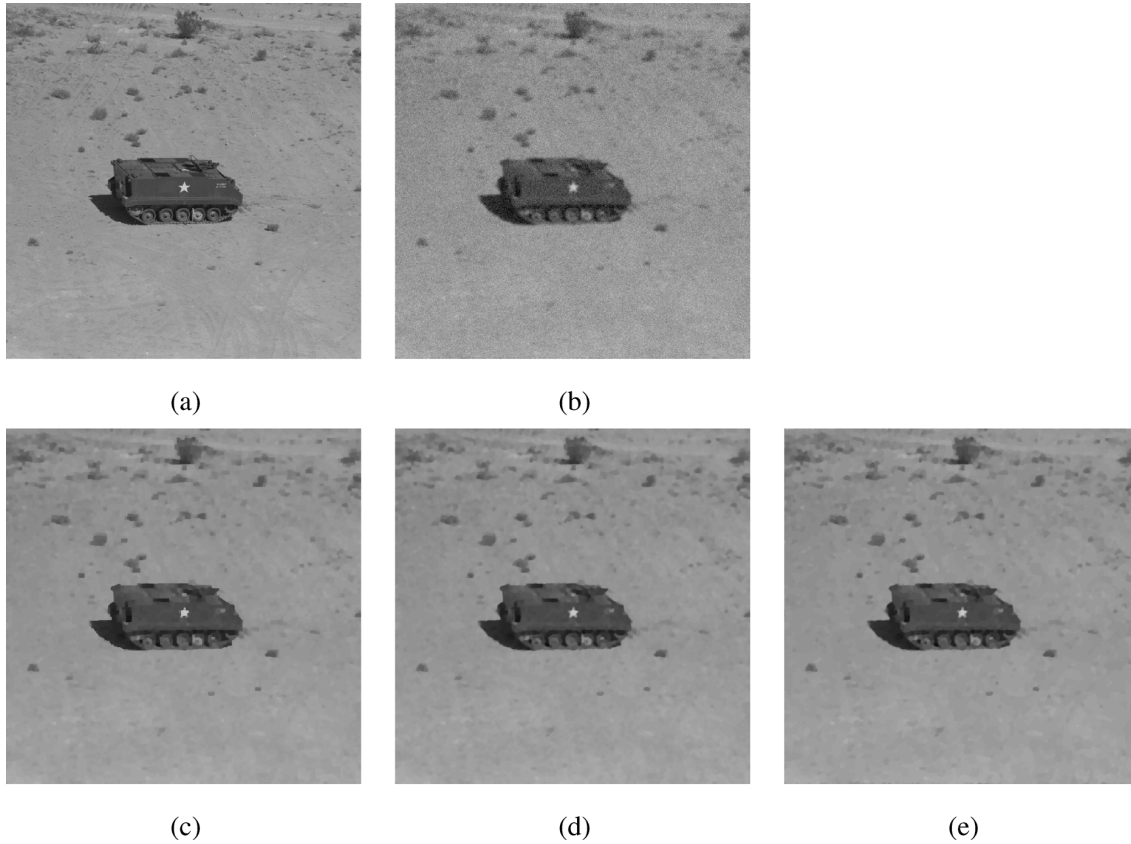
Fig. 7. (a) Original Tank image. (b) Observed image. (c) Image restored by the method in [33] with $\beta = 0.07$. (d)–(e) Images restored by Algorithms I and II, respectively, with $\beta = 0.07$.

TABLE II
RESTORED SNRs AND CPU TIMES FOR DIFFERENT IMAGES

Image	The method [33]		Algorithm I		Algorithm II	
	Computational time (seconds)	SNR	Computational time (seconds)	SNR	Computational time (seconds)	SNR
Cameraman	743.11	22.87	17.14	23.49	12.80	23.49
F16	3864.14	27.13	63.03	28.05	28.03	28.05
Tank	3015.09	31.17	39.16	31.30	15.77	31.21

interior point method. These results demonstrate the proposed methods are quite efficient for restoring images.

Next, we test the performance of the two proposed methods under different levels of noise and blur. The tested Gaussian noises are with the standard deviation of 0.01, 0.05, 0.10, and 0.20. The tested blurring functions are chosen to be truncated 2-D Gaussian function $h(s,t) = \exp((-s^2 - t^2)/(2\sigma^2))$ with (i) $\sigma = 1$ and the support being equal to 5×5; (ii) $\sigma = 1.5$ and the support being equal to 7×7; (iii) $\sigma = 2$ and the support being equal to 9×9 and (iv) $\sigma = 2.5$ and the support being equal to 11×11. In the experiments, we first find good regularization parameters for each noise case when the 7×7 blur is tested. For instance, the regularization parameters are 0.005, 0.03, 0.10, and 0.30 for the noise levels 0.01, 0.05, 0.10, and 0.20, respectively. Then we apply the same regularization parameter for other blurs with the same noise level. In Table III,

we show their SNR results, and find that the SNRs of the two proposed methods are about the same. However, the computational time required by Algorithm II is less than the half of the computational time required by Algorithm I. We remark that Algorithm II employs the shrinkage computation in one step [see (32)] whereas Algorithm I requires to solve TV denoising problem by the Chambolle's algorithm in an iterative manner. Therefore, Algorithm I takes more computational time than that by Algorithm II.

B. Test of Radon Transform Images

Radon transform is a 2-D integral transform that integrate the function along straight lines. Images can be reconstructed by the inverse of the transform. Those resulting images are widely used for guiding medical treatment decisions [42].

TABLE III
RESTORED SNRs AND CPU TIMES FOR DIFFERENT NOISE AND BLUR LEVELS

Noise	Blur	Algorithm I		Algorithm II	
		Computational time (seconds)	SNR	Computational time (seconds)	SNR
0.01	5×5	25.16	28.57	10.11	28.59
	7×7	36.02	25.30	13.58	25.33
	9×9	43.11	23.60	13.09	23.61
	11×11	47.69	22.76	17.31	22.77
0.05	5×5	29.31	25.17	12.00	25.16
	7×7	34.38	23.47	12.80	23.47
	9×9	42.11	22.44	14.23	22.46
	11×11	42.23	21.89	13.67	22.89
0.10	5×5	31.77	23.61	12.20	23.59
	7×7	41.23	22.55	13.70	22.53
	9×9	44.09	21.83	14.05	21.80
	11×11	49.61	21.19	13.11	21.21
0.20	5×5	39.97	21.96	18.05	21.97
	7×7	44.14	21.41	15.34	21.39
	9×9	41.22	20.90	14.19	20.88
	11×11	45.70	20.41	14.92	20.43

In this subsection, the reconstruction of the images transformed by the Radon transform using our proposed method is presented. The modified Shepp–Logan image is applied to illustrate the efficiency of our algorithm. Following the example in [33], we set the image to be of size 50×50. Fig. 8(a) is the original modified Shepp–Logan image. The image is transformed along the angles from 0 to 180 of the increasing of six degrees to create the size of 75×31 Radon transform of the original image. The noise from normal distribution with mean zero and standard deviation 0.05 is added to this transformed image to generate the observed image in Fig. 8(b). The degradation matrix H in this example is the discrete Radon transform matrix and cannot be reformulated as a convolution operator. In order to restore the image and maintain the efficiency of the proposed method, we make use of the back-projection operator B [20]. It can be shown that the back-projected Radon transform is an image of $\mathbf{f}$ blurred by the point spread function of the form $(x^2 + y^2)^{-(1/2)}$ which can be used to construct the convolution matrix H. Now we solve

$$\mathcal{J}_\varepsilon(\mathbf{f}, \mathbf{u}) = \|H\mathbf{f} - B\mathbf{g}\|_2^2 + \beta \sum_{i \in I} \psi_\varepsilon(\|\mathrm{D}_i\mathbf{f}\|_2) + \omega\|\mathbf{f} - \mathbf{u}\|_2^2.$$

and

$$\mathcal{J}_\varepsilon(\mathbf{f}, \mathbf{u}) = \|H\mathbf{f} - B\mathbf{g}\|_2^2 + \beta \sum_{i \in I} \psi_\varepsilon(\|\mathrm{D}_i\mathbf{f}\|_2) + \omega\|\mathrm{D}_i\mathbf{f} - \mathbf{u}_i\|_2^2$$

in Algorithms I and II, respectively [cf. (18)]. The back-projected Radon transform $B\mathbf{g}$ is shown in Fig. 8(c). Fig. 8(d) shows the resulting image that is reconstructed from Fig. 8(b)

by the method [33]. Fig. 8(e) shows the resulting image reconstructed from Fig. 8(c) by Algorithm I. Both of the methods provide high quality restored images. However, as conjugate gradient method is required to solve the linear system, the method [33] took more time (1860 s) to reconstruct the image. The proposed method only takes 2.2 s to restore about the same quality of the reconstructed image. The SNRs of the reconstructed image by the method [33] and the proposed method are 41.96 dB and 43.68 dB, respectively.

In the next experiment, we test a larger image. Fig. 9(a) shows the original modified Shepp–Logan image of size 1000×1000. The corresponding back-projected Radon transform image is shown in Fig. 9(b). Fig. 9(c) shows the reconstructed image of the proposed method. The computational time required is about 480 s, and the SNR is 49.50 dB. When Algorithm II is used, we find that the same quality of reconstructed image is obtained within 250 s. However, the result of the method [33] is not obtained to this case as the required computational time takes more than 2 h.

V. CONCLUDING REMARKS

In this paper, we considered image reconstruction and image restoration using nonconvex and nonsmooth regularization on the ℓ_2 norm of the discrete gradient of the image. Our theoretical results show that the solutions of the corresponding minimization problem are images composed of constant regions surrounded by closed contours and neat edges. From a practical side, the main goal was to conceive fast numerical schemes to solve this difficult minimization problem. We developed two very fast minimization algorithms. Extended experiments have
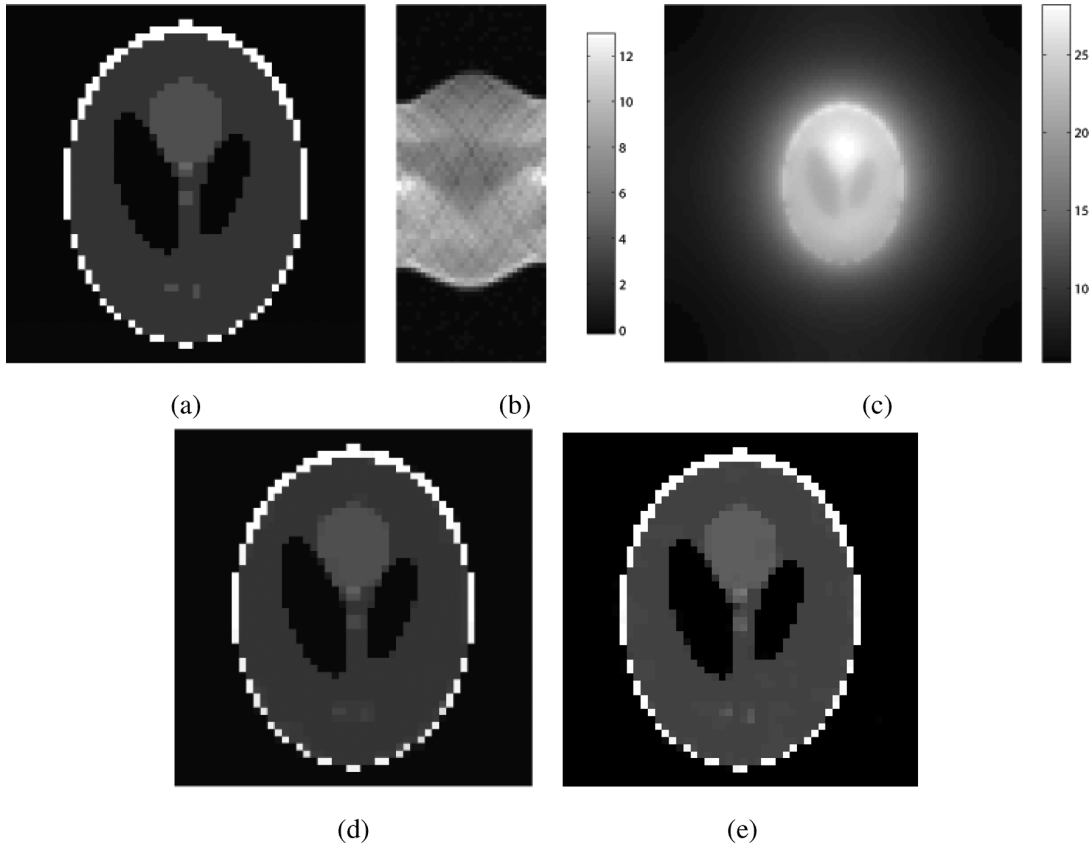
Fig. 8. (a) Original modified Shepp–Logan image with size 50×50. (b) Obtained image after Radon transform along the angles from 0 to 180 with the increasing of six degrees. (c) Back-projected Radon transform image. (d) Image restored from Fig. 8(b) by the method [33] with $\beta = 0.8$. (e) Image restored from Fig. 8(c) by Algorithm I with $\beta = 0.04$.
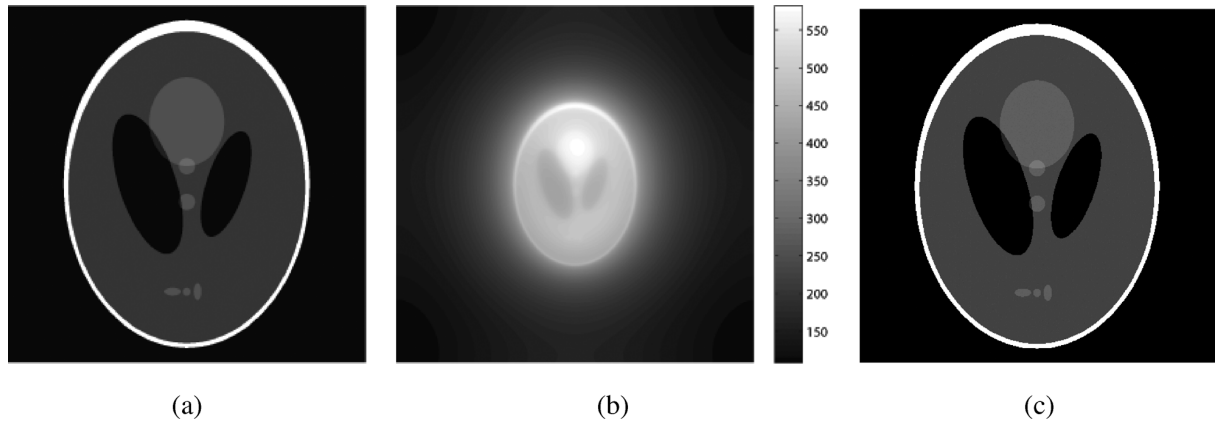

Fig. 9. (a) Original modified Shepp–Logan image with size 1000×1000. (b) Back-projected Radon transform image. (c) Image restored by Method I with $\beta = 0.04$.

shown the effectiveness and efficiency of the proposed numerical schemes.

As for the future research work, we plan to study how to select the regularization parameter for image restoration. We remark in [24] that we have developed a fast total variation image restoration method with an automatic selection of regularization parameter scheme to restore blurred and noisy images. The method exploits the generalized cross-validation technique to determine inexpensively how much regularization to use in each restoration step. We would like to extend this approach to work for non-

convex nonsmooth regularization methods for image restoration and reconstruction problems.

APPENDIX

A. Proof of Theorem 1

With $\hat{\mathbf{f}}$ we associate the following subsets:

$$\hat{I}_0 = \{i \in I : \|D_i\hat{\mathbf{f}}\|_2 = 0\} \quad \text{and} \quad \hat{I}_1 = I \setminus \hat{I}_0 \qquad (35)$$

as well as the vector subspace $K(\hat{I}_0)$ given in the following:

$$
\begin{aligned}
K(\hat{I}_0) &\overset{\text{def}}{=} \{v \in \mathbb{R}^p : \|\mathrm{D}_i v\|_2 = 0, \ \forall i \in \hat{I}_0\} \\
&\equiv \{v \in \mathbb{R}^p : \mathrm{D}_i^k v = 0, 1 \le k \le s, \ \forall i \in \hat{I}_0\}. \quad (36)
\end{aligned}
$$

If $\hat{I}_1 = \varnothing$, then $\|\mathrm{D}_i \hat{\mathbf{f}}\|_2 = 0$, $\forall i \in I$, hence, (6) is trivially true. In what follows, we consider that $\hat{I}_1 \neq \varnothing$. Note that if $\hat{I}_0 = \varnothing$, then $K(\hat{I}_0) = \mathbb{R}^p$.

Given $v \in \mathbb{R}^p$ and $\rho > 0$, we denote by $B(v, \rho) \subset \mathbb{R}^p$ the *open* ball centered at v of radius ρ. Put

$$
\rho \overset{\text{def}}{=} \min_{i \in \hat{I}_1} \|\mathrm{D}_i \hat{\mathbf{f}}\|_2 \frac{1}{\max_{i \in I} \|\mathrm{D}_i\|_2}.
$$

By (35) we see that $\rho > 0$. For every $v \in B(\hat{\mathbf{f}}, \rho)$ we can, hence, write down that

$$
\begin{aligned}
i \in \hat{I}_1 \Rightarrow \|\mathrm{D}_i(\hat{\mathbf{f}} + v)\|_2 \\
\ge \|\mathrm{D}_i \hat{\mathbf{f}}\|_2 - \|\mathrm{D}_i v\|_2 \\
\ge \min_{i \in \hat{I}_1} \|\mathrm{D}_i \hat{\mathbf{f}}\|_2 - \max_{i \in I} \|\mathrm{D}_i\|_2 \|v\|_2 \\
= (\rho - \|v\|_2) \max_{i \in I} \|\mathrm{D}_i\|_2 > 0.
\end{aligned}
$$

Since $\hat{\mathbf{f}} \in K(\hat{I}_0)$, we have

$$
\begin{aligned}
&v \in K(\hat{I}_0) \cap B(0, \rho) \\
&\Rightarrow \begin{cases} \varphi(\|\mathrm{D}_i(\hat{\mathbf{f}} + v)\|_2) = \varphi(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2) = \varphi(0), & \forall i \in \hat{I}_0 \\ \varphi(\|\mathrm{D}_i(\hat{\mathbf{f}} + v)\|_2) > 0, & \forall i \in \hat{I}_1. \end{cases}
\end{aligned}
$$

It follows that:

$$
\begin{aligned}
&v \in K(\hat{I}_0) \cap B(0, \rho) \Rightarrow \mathcal{J}(\hat{\mathbf{f}} + v) \\
&= J(\hat{\mathbf{f}} + v) \\
&\text{where } \mathcal{J}(\mathbf{f}) = \|H\mathbf{f} - \mathbf{g}\|_2^2 + \beta \sum_{i \in \hat{I}_1} \varphi(\|\mathrm{D}_i \mathbf{f}\|_2). \quad (37)
\end{aligned}
$$

Thus, $\mathcal{J}$ is the restriction of J on $K(\hat{I}_0) \cap B(\hat{\mathbf{f}}, \rho)$ which entails that $\mathcal{J}$ has a (local) minimum at $\hat{\mathbf{f}}$.

Let $d_v^+ J(\mathbf{f})$ and $d_v^- J(\mathbf{f})$ denote the right-side and the left-side derivatives of J at $\mathbf{f}$ in the direction of v, respectively.

Let us remind that these are defined by

$$
d_v^+ J(\mathbf{f}) = \lim_{t \searrow 0} \frac{J(\mathbf{f} + tv) - J(\mathbf{f})}{t} \quad \text{and}
$$
$$
d_v^- J(\mathbf{f}) = \lim_{t \searrow 0} \frac{J(\mathbf{f} - tv) - J(\mathbf{f})}{-t}
$$

If J is differentiable at $\mathbf{f}$, then $d_v^+ J(\mathbf{f}) = d_v^- J(\mathbf{f}) = \langle \nabla J(\mathbf{f}), v \rangle$.

Then the first-order necessary condition for $\mathcal{J} : K(\hat{I}_0) \cap B(\hat{\mathbf{f}}, \rho) \mapsto \mathbb{R}$ to have a local minimum at $\hat{\mathbf{f}}$, namely

$$
d_v^- \mathcal{J}(\hat{\mathbf{f}}) \le 0 \le d_v^+ \mathcal{J}(\hat{\mathbf{f}}), \qquad \forall v \in K(\hat{I}_0) \setminus \{0\} \quad (38)
$$

must hold. Let us denote

$$
\hat{I}_M = \{i \in \hat{I}_1 : \|\mathrm{D}_i \hat{\mathbf{f}}\|_2 = t \text{ for } t \in M\}
$$

where M is the set in H3, p. 4. For any $i \in \hat{I}_1 \setminus \hat{I}_M$ and $\forall v \in \mathbb{R}^p$ we have

$$
\langle \nabla \varphi(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2), v \rangle = \varphi'(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2) \frac{\langle \mathrm{D}_i v, \mathrm{D}_i \hat{\mathbf{f}} \rangle}{\|\mathrm{D}_i \hat{\mathbf{f}}\|_2} \quad (39)
$$

and in particular $\langle \nabla \varphi(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2), \hat{\mathbf{f}} \rangle = \varphi'(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2) \|\mathrm{D}_i \hat{\mathbf{f}}\|_2$.

More precisely, $\forall i \in \hat{I}_1 \setminus \hat{I}_M$ we have

$$
\begin{aligned}
\langle \nabla \|\mathrm{D}_i \hat{\mathbf{f}}\|_2, v \rangle &= \left. \frac{d}{dt} \sqrt{\sum_{k=1}^s \left(\mathrm{D}_i^k(\hat{\mathbf{f}} + tv) \right)^2} \right|_{t=0} \\
&= \frac{\sum_{k=1}^s \mathrm{D}_i^k \hat{\mathbf{f}} \, \mathrm{D}_i^k v}{\|\mathrm{D}_i \hat{\mathbf{f}}\|_2} = \frac{\langle \mathrm{D}_i \hat{\mathbf{f}}, \mathrm{D}_i v \rangle}{\|\mathrm{D}_i \hat{\mathbf{f}}\|_2}.
\end{aligned}
$$

Note that the gradient operator ∇ is considered with respect to $\hat{\mathbf{f}}$. Using H2 and H3, for all $i \in \hat{I}_M$ we have

$$
\begin{aligned}
d_{\hat{\mathbf{f}}}^+ \varphi(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2) &= \varphi'(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2^+) \quad \text{and} \\
d_{\hat{\mathbf{f}}}^- \varphi(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2) &= \varphi'(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2^-)
\end{aligned}
$$

where again the side derivatives are considered with respect to $\hat{\mathbf{f}}$.

In detail, by setting $\tau = t \|\mathrm{D}_i \hat{\mathbf{f}}\|_2 > 0$, the calculation is

$$
\begin{aligned}
d_{\hat{\mathbf{f}}}^{\pm} \varphi(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2) &= \lim_{t \searrow 0} \frac{\varphi(\|\mathrm{D}_i(\hat{\mathbf{f}} \pm t\hat{\mathbf{f}})\|_2) - \varphi(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2)}{\pm t} \\
&= \lim_{t \searrow 0} \frac{\varphi((1 \pm t)\|\mathrm{D}_i \hat{\mathbf{f}}\|_2) - \varphi(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2)}{\pm t} \\
&= \lim_{\tau \searrow 0} \frac{\varphi(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2 \pm \tau) - \varphi(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2)}{\pm \tau} \\
&= \varphi'\left(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2^{\pm}\right).
\end{aligned}
$$

Considering the necessary condition (38) for $v = \hat{\mathbf{f}} \in K(\hat{I}_0)$ yields

$$
\begin{aligned}
&d_{\hat{\mathbf{f}}}^- \mathcal{J}(\hat{\mathbf{f}}) \\
&= 2\hat{\mathbf{f}}^T H^T(H\hat{\mathbf{f}} - \mathbf{g}) + \beta \sum_{i \in \hat{I}_1 \setminus \hat{I}_M} \varphi'(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2) \|\mathrm{D}_i \hat{\mathbf{f}}\|_2 \\
&\quad + \beta \sum_{i \in \hat{I}_M} \varphi'(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2^-) \le 0 \\
&\le 2\hat{\mathbf{f}}^T H^T(\hat{\mathbf{f}} - \mathbf{g}) + \beta \sum_{i \in \hat{I}_1 \setminus \hat{I}_M} \varphi'(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2) \|\mathrm{D}_i \hat{\mathbf{f}}\|_2 \\
&\quad + \beta \sum_{i \in \hat{I}_M} \varphi'(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2^+) = d_{\hat{\mathbf{f}}}^+ \mathcal{J}(\hat{\mathbf{f}}).
\end{aligned}
$$

This is equivalent to

$$
\begin{aligned}
&\beta \sum_{i \in \hat{I}_M} \varphi'(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2^-) \\
&\le -2\hat{\mathbf{f}}^T H^T(H\hat{\mathbf{f}} - g) - \beta \sum_{i \in \hat{I}_1 \setminus \hat{I}_M} \varphi'(\|\mathrm{D}_i \hat{\mathbf{f}}\|_2) \|\mathrm{D}_i \hat{\mathbf{f}}\|_2
\end{aligned}
$$

$$\leq \beta \sum_{i \in \hat{I}_M} \varphi'(\|D_i \hat{\mathbf{f}}\|_2^+).$$

In particular, we must have

$$\sum_{i \in \hat{I}_M} \varphi'(\|D_i \hat{\mathbf{f}}\|_2^-) \leq \sum_{i \in \hat{I}_M} \varphi'(\|D_i \hat{\mathbf{f}}\|_2^+).$$

This is clearl impossible since by H3 we have $\sum_{i \in \hat{I}_M} \varphi'(\|D_i \hat{\mathbf{f}}\|_2^-) > \sum_{i \in \hat{I}_M} \varphi'(\|D_i \hat{\mathbf{f}}\|_2^+)$. It follows that:

$$\hat{I}_M = \varnothing \quad \text{and} \quad \hat{I}_1 \setminus \hat{I}_M = \hat{I}_1.$$

The proof is complete.

B. Proof of Lemma 1

By H4 and (7), $\varphi''(t) < 0$ for all $t \in]0, T] \setminus M$ while $\forall t \in M$, φ'' has finite left and right (negative) limits. Noticing that $0 \notin M$, ξ in (8) is well defined and strictly negative on $]0, T]$. We have

$$\varepsilon > 0, (t, t+\varepsilon) \in (\,]0, T] \setminus M)^2$$
$$\Rightarrow \varphi''(t) \leq \varphi''(t+\varepsilon) \leq 0.$$

On the other hand

$$0 < \frac{1}{t+\varepsilon} < \frac{1}{t}, \quad \forall t > 0, \quad \forall \varepsilon > 0.$$

Then

$$\varepsilon > 0, (t, t+\varepsilon) \in (\,]0, T] \setminus M)^2$$
$$\Rightarrow \xi(t) = \frac{\varphi''(t)}{t} < \frac{\varphi''(t+\varepsilon)}{t+\varepsilon} = \xi(t+\varepsilon) \leq 0.$$

If $t = T \in M$, we have $\ddot{\varphi}(T^+) = 0$ and $\ddot{\varphi}(T^-) < 0$. It follows that $\xi(\cdot)$ in (8) is strictly increasing on $]0, T]$. The proof of statement (i) is complete.

Using H4 yet again shows that

$$\lim_{t \to \infty} \xi(t) = 0 \quad \text{and} \quad \lim_{t \searrow 0} \xi(t) = -\infty.$$

Combining this with statement (i) leads immediately to (ii). Statement (iii) is a straightforward consequence of (i) and (ii).

C. Proof of Theorem 2

With $\hat{\mathbf{f}}$, we associate the subsets $\hat{I}_0$ and $\hat{I}_1$, as in (35), as well as the subspace $K(\hat{I}_0)$ as introduced in (36). If $\hat{I}_1 = \varnothing$, then $\|D_i \hat{\mathbf{f}}\|_2 = 0$, $\forall i \in I$ in which case (9) is trivially satisfied. In what follows, we consider that $\hat{I}_1 \neq \varnothing$.

¿From Theorem 1 we know that $\mathcal{J} : K(\hat{I}_0) \cap B(\hat{\mathbf{f}}, \rho) \mapsto \mathbb{R}$, as given in (37), is $\mathcal{C}^2$-continuous on a neighborhood of $\hat{\mathbf{f}} \in K(\hat{I}_0)$. Hence, $\nabla \mathcal{J}(\hat{\mathbf{f}})$ and $\nabla^2 \mathcal{J}(\hat{\mathbf{f}})$ are well defined in the usual sense. The second-order necessary condition for a (local) minimum of $\mathcal{J}$ at $\hat{\mathbf{f}}$ must also hold and reads

$$\langle \nabla^2 \mathcal{J}(\hat{\mathbf{f}}) v, v \rangle \geq 0, \qquad \forall v \in K(\hat{I}_0). \tag{40}$$

Next we derive the expression of $\langle \nabla^2 \mathcal{J}(\hat{\mathbf{f}}) v, v \rangle$. Using (39), for any $i \in \hat{I}_1$ we get

$$\langle \nabla^2 \varphi(\|D_i \hat{\mathbf{f}}\|_2) v, v \rangle = \varphi''(\|D_i \hat{\mathbf{f}}\|_2) \left(\frac{\langle D_i \hat{\mathbf{f}}, D_i v \rangle}{\|D_i \hat{\mathbf{f}}\|_2} \right)^2$$
$$+ \varphi'(\|D_i \hat{\mathbf{f}}\|_2) \frac{\|D_i v\|_2^2 \|D_i \hat{\mathbf{f}}\|_2^2 - \langle D_i \hat{\mathbf{f}}, D_i v \rangle^2}{\|D_i \hat{\mathbf{f}}\|_2^3}.$$

The necessary condition (40) in detail reads

$$\langle \nabla^2 \mathcal{J}(\hat{\mathbf{f}}) v, v \rangle$$
$$= 2\|Hv\|_2^2 + \beta \sum_{i \in \hat{I}_1} \varphi''(\|D_i \hat{\mathbf{f}}\|_2) \frac{\langle D_i \hat{\mathbf{f}}, D_i v \rangle^2}{\|D_i \hat{\mathbf{f}}\|_2^2}$$
$$+ \beta \sum_{i \in \hat{I}_1} \varphi'(\|D_i \hat{\mathbf{f}}\|_2)$$
$$\times \frac{\|D_i v\|_2^2 \|D_i \hat{\mathbf{f}}\|_2^2 - \langle D_i \hat{\mathbf{f}}, D_i v \rangle^2}{\|D_i \hat{\mathbf{f}}\|_2^3} \geq 0,$$
$$\forall v \in K(\hat{I}_0). \tag{41}$$

The core of the proof is conducted by contradiction. We will exhibit a $\theta > 0$ and a direction v such that $\langle \nabla^2 \mathcal{J}(\hat{\mathbf{f}}) v, v \rangle < 0$ if there exists $j \in \hat{I}_1$ such that $0 < \|D_j \hat{\mathbf{f}}\| < \theta$.

Let $j \in \hat{I}_1$ be such that

$$\|D_j \hat{\mathbf{f}}\|_2 \leq \|D_i \hat{\mathbf{f}}\|_2, \qquad \forall i \in \hat{I}_1 \tag{42}$$

and set

$$\kappa \stackrel{\text{def}}{=} \|D_j \hat{\mathbf{f}}\|_2. \tag{43}$$

Then $\kappa > 0$ by the definition of $\hat{I}_1$ and κ is finite since $\#\hat{I}_1 \leq p$. Consider $\hat{v}$ given by

$$\hat{v}[i] = \hat{\mathbf{f}}[i] \, \kappa^{-3/2}, \qquad \forall i \in I.$$

Clearly $\hat{v} \in K(I_0)$. Using Schwarz inequality

$$\langle D_i \hat{\mathbf{f}}, D_i \hat{v} \rangle = \|D_i \hat{\mathbf{f}}\| \, \|D_i \hat{v}\|,$$

so the numerator in the last term in (41) is null

$$\langle \nabla^2 \mathcal{J}(\hat{\mathbf{f}}) \hat{v}, \hat{v} \rangle = 2\|H\hat{v}\|_2^2 + \beta \sum_{i \in \hat{I}_1} \varphi''(\|D_i \hat{\mathbf{f}}\|_2) \frac{\langle D_i \hat{\mathbf{f}}, D_i \hat{v} \rangle^2}{\|D_i \hat{\mathbf{f}}\|_2^2}.$$

Furthermore

$$\langle D_i \hat{\mathbf{f}}, D_i \hat{v} \rangle^2 = \|D_i \hat{\mathbf{f}}\|(\kappa^{-1} \|D_i \hat{\mathbf{f}}\|)^3 \geq \|D_i \hat{\mathbf{f}}\|, \qquad \forall i \in \hat{I}_1.$$

Let us also define

$$\alpha \stackrel{\text{def}}{=} \max\{1, \, \kappa^{-3}\} \tag{44}$$

where κ is defined in (43). Using that $\varphi''(\|D_i\hat{\mathbf{f}}\|_2) \leq 0$, $\forall i \in \hat{I}_1$, we get

$$\langle \nabla^2 \mathcal{J}(\hat{\mathbf{f}})\hat{v}, \hat{v} \rangle \leq 2\,\kappa^{-3}\,\|H\hat{\mathbf{f}}\|_2^2 + \beta \sum_{i \in \hat{I}_1} \frac{\varphi''(\|D_i\hat{\mathbf{f}}\|_2)}{\|D_i\hat{\mathbf{f}}\|_2}$$

$$\leq 2\,\alpha\,\|H\hat{\mathbf{f}}\|_2^2 + \beta\#\hat{I}_1 \frac{\varphi''(\|D_j\hat{\mathbf{f}}\|_2)}{\|D_j\hat{\mathbf{f}}\|_2} \quad (45)$$

$$= 2\,\alpha\,\|H\hat{\mathbf{f}}\|_2^2 + \beta\#\hat{I}_1\xi(\|D_j\hat{\mathbf{f}}\|_2) \quad (46)$$

where the element $\|D_j\hat{\mathbf{f}}\|_2$ is defined in (42) and ξ is the multifunction in (8) considered in Lemma 1.

Define $\theta \in (0, T)$ to solve

$$\xi(\theta) = -\frac{2\,\alpha\,\|H\hat{\mathbf{f}}\|_2^2}{\beta\#\hat{I}_1} \quad (47)$$

where α is defined in (44). By Lemma 1, this θ is well defined and unique. Suppose that

$$0 < \|D_j\hat{\mathbf{f}}\| = \kappa < \theta.$$

According to Lemma 1

$$-\infty = \lim_{t \searrow 0} \xi(t) < \xi(\|D_j\hat{\mathbf{f}}\|) < \xi(\theta) < 0. \quad (48)$$

This, jointly with (46) yields

$$\langle \nabla^2 \mathcal{J}(\hat{\mathbf{f}})\hat{v}, \hat{v} \rangle \leq 2\alpha\|H\hat{\mathbf{f}}\|_2^2 + \beta\#\hat{I}_1\xi(\kappa)$$
$$< 2\alpha\|H\hat{\mathbf{f}}\|_2^2 + \beta\#\hat{I}_1\xi(\theta) = 0.$$

The obtained result clearly contradicts the necessary condition stated in (41). It follows that:

$$\|D_j\hat{\mathbf{f}}\|_2 \geq \theta \quad (49)$$

where the inequality is strict if $\theta \in M$. Hence, the result for θ as in (10).

Next we focus on the conditions given in (i). Combining ([32], Th. 2.3) with Theorem 1 immediately yields that for any (local) minimizer $\hat{\mathbf{f}}$ we have

$$\|H\hat{\mathbf{f}}\|_2 \leq \|\mathbf{g}\|_2.$$

Inserting this result into (45) yields

$$\langle \nabla^2 \mathcal{J}(\hat{\mathbf{f}})\hat{v}, \hat{v} \rangle \leq 2\,\alpha\,\|\mathbf{g}\|_2^2 + \beta\#\hat{I}_1 \frac{\varphi''(\|D_j\hat{\mathbf{f}}\|_2)}{\|D_j\hat{\mathbf{f}}\|_2}.$$

In this case, we consider $\theta \in (0, T)$ that solves

$$\xi(\theta) = -\frac{2\,\alpha\,\|\mathbf{g}\|_2^2}{\beta\#\hat{I}_1}. \quad (50)$$

Then we apply the same reasoning that led us to (49). This proves statement (i).

D. Proof of Lemma 2

Note that by (a) and (16), $\psi_\varepsilon(t)$ is symmetric with $\psi_\varepsilon(0) = 0$, $\forall\varepsilon \in [0, 1]$. Statement (i) follows directly from (b) and (16). By (a), ψ_ε is $\mathcal{C}^2$ on $\mathbb{R} \setminus \{0\}$. Since ψ_ε is symmetric, we have

$$-\psi_\varepsilon'(-t) = \psi_\varepsilon'(t) = \varphi_\varepsilon'(t) - \varphi_\varepsilon'(0^+)\text{sign}(t),$$
$$\forall t \in \mathbb{R} \setminus \{0\} \quad (51)$$

$$\psi_\varepsilon''(-t) = \psi_\varepsilon''(t) = \varphi_\varepsilon''(t),$$
$$\forall t \in \mathbb{R} \setminus \{0\}. \quad (52)$$

Since by (a) φ_ε is continuous on $\mathbb{R}$ and $\varphi_\varepsilon(0) = 0$, we can write down

$$\varphi_\varepsilon'(0^+) = \lim_{t \searrow 0} \frac{\varphi_\varepsilon(t)}{t} = \lim_{t \searrow 0} \varphi_\varepsilon'(t) \quad \text{and}$$

$$\varphi_\varepsilon'(0^-) = \lim_{t \nearrow 0} \frac{\varphi_\varepsilon(t)}{t} = \lim_{t \nearrow 0} \varphi_\varepsilon'(t).$$

By the symmetry of φ_ε, we have $\varphi_\varepsilon'(0^+) = -\varphi_\varepsilon'(0^-)$. Using (51)

$$\psi_\varepsilon'(0^+) = \lim_{t \searrow 0} \varphi_\varepsilon'(t) - \varphi_\varepsilon'(0^+) = 0 \quad \text{and}$$

$$\psi_\varepsilon'(0^-) = \lim_{t \nearrow 0} \varphi_\varepsilon'(t) + \varphi_\varepsilon'(0^+) = 0$$

hence, $\psi_\varepsilon'(0) = 0$. Noticing that ψ_ε is continuous on $\mathbb{R}$, and using Hôpital's rule and the fact that $\psi_\varepsilon(0) = 0$, we can write that

$$\lim_{t \to 0} \psi_\varepsilon'(t) = \lim_{t \to 0} \frac{\psi_\varepsilon(t)}{t} = \psi_\varepsilon'(0) = 0$$

which shows that ψ_ε' is continuous at 0. Hence, (ii).

Using (52), statement (ii), condition (d) and the symmetry of ψ_ε, we have

$$\psi_\varepsilon''(0^-) = \psi_\varepsilon''(0^+) = \lim_{t \searrow 0} \frac{\psi_\varepsilon'(t)}{t}$$
$$= \lim_{t \searrow 0} \psi_\varepsilon''(t) = \ddot{\varphi}_\varepsilon(0^+) = \lim_{t \to 0} \psi_\varepsilon''(t).$$

Consequently, ψ_ε'' is continuous at zero and $\psi_\varepsilon''(0) < 0$. Hence, (iii).

REFERENCES

[1] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing*, 2nd ed. Berlin, Germany: Springer-Verlag, 2006.

[2] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher, "Structure-texture image decomposition—Modeling, algorithms, and parameter selection," *Int. J. Comput. Vis.*, vol. 67, no. 1, pp. 111–136, 2006.

[3] L. Bar, A. Brook, N. Sochen, and N. Kiryati, "Deblurring of color images corrupted by salt- and-pepper noise," *IEEE Trans. Image Process.*, vol. 16, no. 4, pp. 1101–1111, Apr. 2007.

[4] L. Bedini, I. Gerace, and A. Tonazzini, "A GNC algorithm for constrained image reconstruction with continuous-value line process," *Pattern Recognit. Lett.*, vol. 15, no. 9, pp. 907–918, 1994.

[5] J. E. Besag, "Digital image processing: Towards Bayesian image analysis," *J. Appl. Stat.*, vol. 16, no. 3, pp. 395–407, 1989.

[6] M. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications to early vision," *Int. J. Comput. Vis.*, vol. 19, no. 1, pp. 57–91, 1996.

[7] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.

[8] A. Blake, "Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-11, no. 1, pp. 2–12, Jan. 1989.

[9] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, no. 1–2, pp. 89–97, 2004.

[10] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structure and problems," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-37, no. 12, pp. 2024–2036, Dec. 1989.

[11] S. Durand and M. Nikolova, "Denoising of frame coefficients using l1 data-fidelity term and edge-preserving regularization," *SIAM J. Multiscale Model. Sim.*, vol. 6, no. 2, pp. 547–576, 2007.

[12] H. Fu, M. Ng, M. Nikolova, and J. Barlow, "Efficient minimization methods of mixed ℓ_2-ℓ_1 and ℓ_1-ℓ_1 norms for image restoration," *SIAM J. Sci. Comput.*, vol. 27, no. 6, pp. 1881–1902, 2006.

[13] D. Geman and G. Reynolds, "Constrained restoration and recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, Mar. 1992.

[14] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, Jul. 1995.

[15] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributinons, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Jun. 1984.

[16] R. Gonzalez and R. Woods, "Digital image processing," Pearsol Int. Edition3rd ed. 2008.

[17] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imag.*, vol. 9, no. 1, pp. 84–93, Mar. 1990.

[18] X. Guo, F. Li, and M. Ng, "A fast ℓ_1-TV algorithm for image restoration," *SIAM J. Sci. Comput.*, vol. 31, no. 3, pp. 2322–2341, 2009.

[19] Y. Huang, M. K. Ng, and Y. Wen, "A fast total variation minimization method for image restoration," *SIAM J. Multiscale Model. Sim.*, vol. 7, no. 2, pp. 774–795, 2008.

[20] A. Jain, *Fundamentals of Digital Image Processing*. Upper Saddle River, NJ: Prentice-Hall, 1989.

[21] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[22] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs priors," *IEEE Trans. Med. Imag.*, vol. 9, no. 4, pp. 439–446, Dec. 1990.

[23] S. Li, *Markov Random Field Modeling in Computer Vision*, 1st ed. New York: Springer-Verlag, 1995.

[24] H. Liao, F. Li, and M. Ng, "Selection of regularization parameter in total variation image restoration," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 26, pp. 2311–2320, 2009.

[25] D. Mumford and J. Shah, "Boundary detection by minimizing functionals," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 1985, pp. 22–26.

[26] M. K. Ng, R. H. Chan, and W. Tang, "A fast algorithm for deblurring models with Neumann boundary conditions," *SIAM J. Sci. Comput.*, vol. 21, no. 3, pp. 851–866, 1999.

[27] M. Nikolova, "Markovian reconstruction using a GNC approach," *IEEE Trans. Image Process.*, vol. 8, no. 9, pp. 1204–1220, Sep. 1999.

[28] M. Nikolova, J. Idier, and A. Mohammad-Djafari, "Inversion of large-support ill-posed linear operators using a piecewise Gaussian MRF," *IEEE Trans. Image Process.*, vol. 7, no. 4, pp. 571–585, Apr. 1998.

[29] M. Nikolova, "Minimizers of cost-functions involving non-smooth data-fidelity terms. Application to the processing of outliers," *SIAM J. Numer. Anal.*, vol. 40, no. 3, pp. 965–994, 2002.

[30] M. Nikolova, "A variational approach to remove outliers and impulse noise," *J. Math. Imag. Vis.*, vol. 20, no. 1–2, pp. 99–120, 2004.

[31] M. Nikolova, "Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares," *SIAM J. Multiscale Model. Sim.*, vol. 4, no. 3, pp. 960–991, 2005.

[32] M. Nikolova, "Analytical bounds on the minimizers of (nonconvex) regularized least-squares," *AIMS J. Inverse Problems Imag.*, vol. 1, no. 4, pp. 661–677, 2007.

[33] M. Nikolova, M. K. Ng, S. Zhang, and W. Ching, "Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization," *SIAM J. Imag. Sci.*, vol. 1, no. 1, pp. 2–25, 2008.

[34] J. Nocedal and S. Wright, *Numerical Optimization*. New York: Springer-Verlag, 1999.

[35] M. Robini, T. Rastello, and I. Magnin, "Simulated annealing, acceleration techniques, and image restoration," *IEEE Trans. Image Process.*, vol. 8, no. 10, pp. 1374–1387, Oct. 1999.

[36] M. Robini, A. Lachal, and I. Magnin, "A stochastic continuation approach to piecewise constant reconstruction," *IEEE Trans. Image Process.*, vol. 16, no. 10, pp. 2576–2589, Oct. 2007.

[37] L. Rudin, S. Osher, and C. Fatemi, "Nonlinear total variaion based noise removal algorithm," *Physica*, vol. 60, no. 1–4, pp. 259–268, 1992.

[38] R. Stevenson and E. Delp, "Fitting curves with discontinuities," in *Proc. 1st Int. Workshop Robust Comput. Vis.*, Seattle, WA, 1990, pp. 127–136.

[39] A. Tikhonov and V. Arsenin, Solutions of Ill-Posed Problems. Washington, DC, 1977, Winston.

[40] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imag. Sci.*, vol. 1, no. 3, pp. 248–272, 2008.

[41] E. Wasserstrom, "Numerical solutions by the continuation method," *SIAM Rev.*, vol. 15, no. 1, pp. 89–119, 1973.

[42] M. Wintermark, M. Sesay, E. Barbier, K. Borbely, W. P. Dillon, J. D. Eastwood, T. C. Glenn, C. B. Grandin, S. Pedraza, J. F. Soustiel, T. Nariai, G. Zaharchuk, J. M. Caille, V. Dousset, and H. Yonas, "Comparative overview of brain perfusion imaging techniques," *Stroke*, vol. 32, no. 5, pp. 294–314, 2005.

Mila Nikolova (M'07–SM'08) received the Ph.D. degree from the Université de Paris-Sud, France, in 1995.

She got a Habilitation to direct research in 2006. Currently, she is a Research Director with the National Center for Scientific Research (CNRS), Cachan Cedex, France. She performs her duty as a full-time Researcher at the Centre de Mathématiques et de Leurs Applications (CMLA), the 8536 UMR of CNRS, in the Department of Mathematics of ENS de Cachan, France. Her research interests are in mathematical Image and signal reconstruction, inverse problems, regularization and variational methods and the properties of their solutions, and scientific computing.
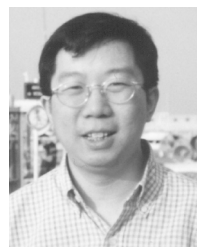
Michael K. Ng is a Professor in the Department of Mathematics at the Hong Kong Baptist University. He was a Research Fellow of Computer Sciences Laboratory at Australian National University from 1995 to 1997 and an Assistant/Associate Professor of the University of Hong Kong from 1997 to 2005 before joining Hong Kong Baptist University. As an applied mathematician, his main research areas include bioinformatics, data mining, operations research, and scientific computing. He has published and edited five books, published more than 200 journal papers. He has reviewed papers for more than 50 international journals. He currently serves on the editorial boards of international journals.
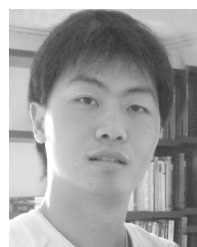
Chi-Pan Tam received the B.Sc. degree in mathematical science and the Ph.D. degree in computational mathematics from the Hong Kong Baptist University, Kowloon Tong, Hong Kong, in 2006 and 2010, respectively.

He is currently a Research Associate in the Department of Mathematics, Hong Kong Baptist University. His research interests include image processing and medical signal processing.