A Fast Nonconvex Nonsmooth Minimization Method for Image Restoration and Reconstruction

Mila Nikolova *Senior Member, IEEE*, Michael K. Ng, 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 study a fast nonconvex nonsmooth minimization method for image restoration and reconstruction. Our theoretical results show that the solution of the nonconvex nonsmooth minimization problem can be composed of constant regions surrounded by closed contours and neat edges. The main aim of this paper is to develop a fast minimization algorithm to solve the nonconvex nonsmooth minimization problem. Our experimental results show that the effectiveness and efficiency of the proposed method.

Index Terms

image restoration, image reconstruction, nonconvex nonsmooth regularization, total variation, fast Fourier transform

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 [30], [10], [44], [3], [27]. 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 $\mathbf{f} \in \mathbb{R}^p$, $p = mn$ as:

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

Mila Nikolova is with Centre de Mathématiques et de Leurs Applications, ENS de Cachan, 61 av. President Wilson, 94235 Cachan Cedex, France. (e-mail: nikolova@cmla.ens-cachan.fr)

The Corresponding Author. Michael K. Ng is 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). The research of this author is supported in part by Hong Kong Research Grants Council Grants and HKBU FRGs.

Chi-Pan Tam is with the Centre for Mathematical Imaging and Vision and Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong. (e-mail:cptam@math.hkbu.edu.hk)

where $\mathbf{n}$ 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. Usually 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 to data production model (1) and satisfies some prior requirements. A flexible means to define such a solution is regularization, e.g., [4], [13], [17], [1], where $\hat{\mathbf{f}}$ is a minimizer of a cost function $J(\mathbf{f})$ 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 trade-off between these two terms. The most usual choice for Θ is

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

In a statistical setting, this Θ corresponds to assume that $\mathbf{n}$ is white Gaussian noise. Our computational method is designed for Θ as given in (3), or for any smooth and convex function Θ. Recently, data terms of the form $\Theta(\mathbf{v}) = \|\mathbf{v}\|_1$ were shown to be useful if some data entries have to be preserved, which is appreciable for instance if $\mathbf{n}$ is impulse noise [35], [11], [23] or to image decomposition [2] Our method is straightforward to extend to this situation, see for instance [20].

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 the set of all pixels of the image,

$$I = \{1, \cdots, p\},$$

$\mathrm{D}_i : \mathbb{R}^p \to \mathbb{R}^s$, for $s \geq 1$ is a linear operator yielding a vector containing the differences between pixel i and its neighbors, and φ is called a potential function (PF). Each D_i can be seen as an $s \times p$ matrix. If $\{\mathrm{D}_i : i \in I\}$ correspond to the forward discretization of the gradient operator, we have $\mathrm{D}_i \in \mathbb{R}^{2\times p}$ and in particular,

$$\mathrm{D}_i^1\mathbf{f} = \begin{cases} \mathbf{f}_{i+1} - \mathbf{f}_i & \text{if} \quad i \notin \{n, 2n \cdots, mn\} \\ 0 & \text{else} \end{cases} \tag{5}$$

$$\mathrm{D}_i^2\mathbf{f} = \begin{cases} \mathbf{f}_{i+n} - \mathbf{f}_i & \text{if} \quad i \notin \{(m-1)n+1, (m-1)n+2, \cdots, mn\} \\ 0 & \text{else} \end{cases} \tag{6}$$

In this case,

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

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

For any $s \geq 1$ and for any $i \in I$, we define the $s \times p$ matrix D_i as given below:

$$\mathrm{D}_i = \begin{bmatrix} \mathrm{D}_i^1 \\ \cdots \\ \mathrm{D}_i^s \end{bmatrix} \tag{7}$$

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

$$\Phi(\mathbf{f}) = \sum_{i=1}^{r} \varphi(|\mathbf{d}_i^T \mathbf{f}|) \tag{8}$$

where the superscript $.^T$ stands for transpose, and $\mathbf{d}_i \in \mathbb{R}^p$, for $i = 1, \cdots, r$, usually yield the first-order differences between each pixel and its 4 or 8 adjacent neighbors. Note that this kind of regularizer are customary in Markov Random Field modeling, see e.g. the classical survey paper [4]. Even though rotation invariance is not well defined in the discrete setting, it is usually observed that the use of the ℓ_2 norm of the discrete gradient, i.e. $\|\mathrm{D}_i \mathbf{f}\|_2$ instead of $|\mathbf{d}_i^T \mathbf{f}|$ yields image restorations of better quality.

Various potential functions (PFs) φ have been used in the literature, a complete review can be found for instance in [5]. An important requirement is that φ allows the recovery of large differences $\|\mathrm{D}_i \mathbf{f}\|_2$ at the right 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 [46]. Since the pioneering work of Geman & Geman [17], different non-convex functions φ have been considered either in a statistical or in a variational framework [1], [4], [18], [15], [16], [26], [29], [40]. In order to avoid the numerical intricacies arising with nonconvex regularization, an important effort was done to derive *convex* edge-preserving PFs [19], [25], [43], [45], [8], [12]. Nevertheless, *nonconvex nonsmooth regularization* offers much richer possibilities to restore high quality images with neat edges: for regularizer functions of the form (8) a theoretical explanation was provided in [36], [37] while numerical examples can be found in numerous articles, see e.g. [15], [16], [32], [42], [38],

This papers provides two main contributions. The theoretical one is to prove that the minimizers of cost functions of the form (3)-(4), where $\varphi(\|.\|_2)$ is nonconvex and nonsmooth, are composed of constant regions surrounded by closed contours and neat edges (Section 2). The practical contribution is quite challenging: we derive fast algorithms to approximate faithfully the global minimizer of these nonconvex and nonsmooth cost functions (Section 3). Our

	φ	φ''	T	M	$\varphi''(t^{\pm}),\ t\in M$
(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)	$\begin{cases} t^{\alpha},\ \alpha\in(0,1) & t<1 \\ t^{\gamma},\ \gamma\in(0,\alpha\min\{1,\frac{1-\alpha}{\alpha}\}) & t\geq 1 \end{cases}$	$\varphi''(t)=\begin{cases} \alpha(\alpha-1)t^{\alpha-2} & t<1 \\ \gamma(\gamma-1)t^{\gamma-2} & t\geq 1 \end{cases}$	∞	$\{1\}$	$\varphi''(1^-)=\alpha(\alpha-1)$ $\varphi''(1^+)=\gamma(\gamma-1)$
(f5)	$\begin{cases} t^{\alpha},\ \alpha\in(0,1) & t<1 \\ 1 & \text{else} \end{cases}$	$\varphi''(t)=\begin{cases} \alpha(\alpha-1)t^{\alpha-2} & t<1 \\ 0 & \text{else} \end{cases}$	1	$\{1\}$	$\varphi''(1^-)=\alpha(\alpha-1)$ $\varphi''(1^+)=0$

TABLE I

SEVERAL PFS YIELDING NONSMOOTH AND NONCONVEX REGULARIZATION AND SATISFYING H2, H3 AND H4. NOTE THAT FOR (F4) AND (F5) WE HAVE $\varphi''(1^+)>\varphi''(1^-)$.

experimental results (Section 4) show clearly the effectiveness and efficiency of the proposed numerical schemes. Concluding remarks and perspectives are sketched in Section 5.

II. ESSENTIAL PROPERTIES OF MINIMIZERS OF $\boldsymbol{J}$

In this section, we study the properties of minimizers of $\boldsymbol{J}$ defined according to (3)-(4) under customary, weak assumptions. We will assume the following:

H1: $\ker(H)\cap\ker(\mathrm{D}_i)=\{0\}$, $\forall i\in I$;

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

H3: $\varphi:\mathbb{R}_+\to\mathbb{R}_+$ 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 $\varphi'(t^-)>\varphi'(t^+)$;

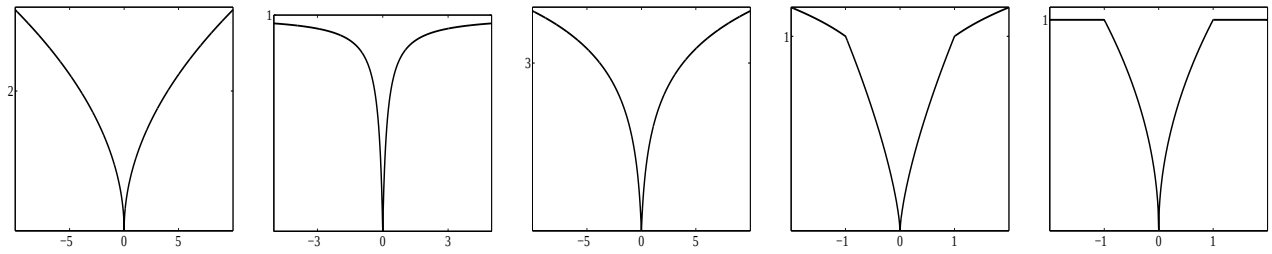
H4: $\varphi''(0)<0$, for all $t\in\mathbb{R}_+\setminus M$ we have φ'' increasing with $\varphi''(t)\leq 0$, and $\lim_{t\to\infty}\varphi''(t)=0$, while if $t\in M$, $\varphi''(t^-)$ and $\varphi''(t^+)$ are finite and $\varphi''(t^-)<\varphi''(t^+)$.

Examples of PFs $\varphi:\mathbb{R}_+\to\mathbb{R}$ satisfying these assumptions are given in Table I and illustrated in Fig. 1. In particular, φ'' on $RR_+\setminus M$ for all these functions in plotted in Fig. 2. The constant T and the set M are defined in (10) and in H3. Note that by H2, all terms of Φ such that $\|\mathrm{D}_i\mathbf{f}\|_2=0$ are non-differentiable.

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

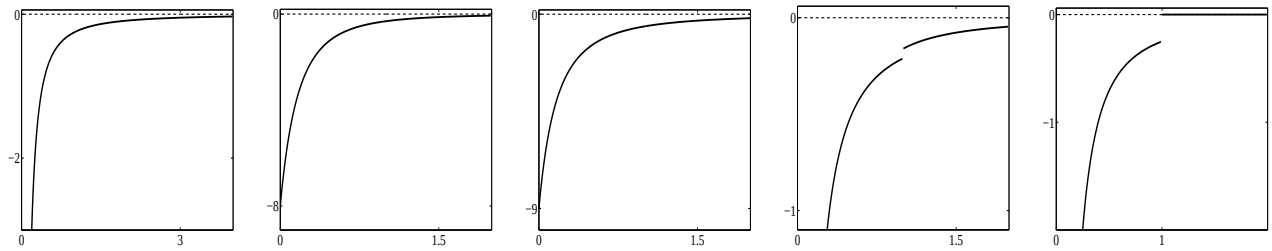
Theorem 1: Let J be of the form (2) for $\beta>0$ and all assumptions H1, H2, H3 and H4 hold. Given $g\in\mathbb{R}^q$, let $\widehat{\mathbf{f}}$ be any (local) minimizer of J. Then we have:

$$\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2>0\quad\Rightarrow\quad\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2\notin M, \tag{9}$$



(f1): $t \mapsto \varphi(|t|)$ (f2): $t \mapsto \varphi(|t|)$ (f3): $t \mapsto \varphi(|t|)$ (f4): $t \mapsto \varphi(|t|)$ (f5): $t \mapsto \varphi(|t|)$

Fig. 1. The potential functions from Table I applied to $|t|$.



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

Fig. 2. $\varphi'' : \mathbb{R}_+ \setminus M \to \mathbb{R}$ for all the potential functions in Table I.

where M is the set described in H3.

Given a PF φ as described by H2, H3 and H4, we define $T \in (0, \infty]$ by

$$t \geq T \quad \Rightarrow \quad \varphi''(T) = 0 \ . \tag{10}$$

The lemma stated below is of great use in the proof of our main result, namely Theorem 2. The multifunction considered is illustrated on Fig. 3.

Lemma 1: Let $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ satisfy the assumptions H2, H3 and H4 in page 4. Consider the multifunction ξ as defined below:

$$\begin{aligned} \xi \quad &: \quad [0,T] \mapsto \text{the subsets of } \mathbb{R} \ , \\ \xi(t) \quad &= \quad \begin{cases} \dfrac{\varphi''(t)}{t} & \text{if} \quad t \notin M \ , \\ \left[\dfrac{\varphi''(t^-)}{t}, \dfrac{\varphi''(t^+)}{t} \right] & \text{if} \quad t \in M \ , \end{cases} \end{aligned} \tag{11}$$

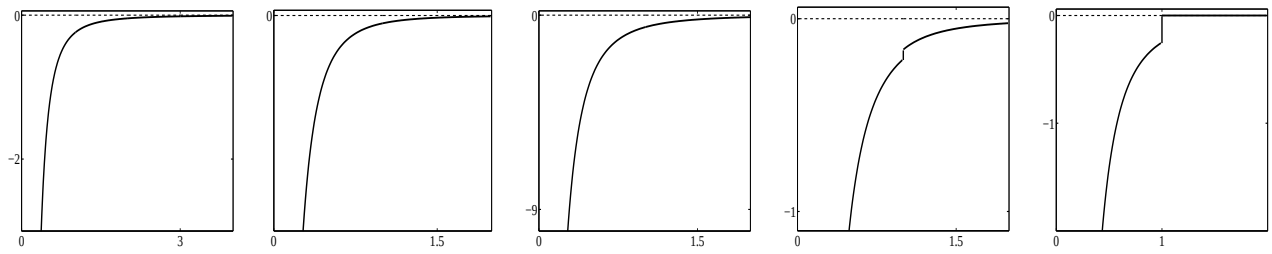
where M is the finite set described in H3 and T is given in (10). Then

(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 $c \to \theta_c \in (0, T)$ increases when c decreases on $(0, \infty)$.



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

Fig. 3. The multifunction ξ in (11) for each PF in Table I. Remind that for (f5) we have $T = 1$

The proof of this 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 cost-function J, namely that minimizers $\widehat{\mathbf{f}}$ are composed of constant regions surrounded by closed contours and neat edges higher than θ. The solution $\widehat{\mathbf{f}}$ is hence 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 can be found in Appendix C.

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

$$\text{either} \quad \|\mathrm{D}_i\widehat{\mathbf{f}}\|_2 = 0 \quad \text{or} \quad \|\mathrm{D}_i\widehat{\mathbf{f}}\|_2 \geq \theta, \quad \forall i \in I, \tag{12}$$

where $0 < \theta < T$ and the inequality above is strict if $\theta \in M$. The alternative (12) holds always true for the unique θ that solves

$$\xi(\theta) = -\frac{2\,\alpha\,\|H\widehat{\mathbf{f}}\|_2^2}{\beta K}, \quad K = \#\{i \in I : \|\mathrm{D}_i\widehat{\mathbf{f}}\|_2 > 0\}, \tag{13}$$

where ξ is the multifunction in (11) and $\alpha \stackrel{\text{def}}{=} \max\Big\{1,\, \big(\min\{\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2 > 0\}\big)^{-3}\Big\}$.

(i) If in addition we have: rank$(H) = p$ <u>or</u> φ strictly increasing on $\mathbb{R}_+$, then (12) holds true for the unique θ that solves

$$\xi(\theta) = -\frac{2\,\alpha\,\|\mathbf{g}\|_2^2}{\beta K}, \quad K = \#\{i \in I : \|\mathrm{D}_i\widehat{\mathbf{f}}\|_2 > 0\}. \tag{14}$$

(ii) Whenever $K \geq 1$, (12) holds as well if we set $K = 1$ in (13) and in (14), for a smaller value of θ.

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 $\mathbf{n} \sim \text{Normal}(0, \sigma^2)$ with $\sigma = 0.8$. The histogram of all $\mathbf{f}$ and $\mathbf{g}$ are shown in Figures 4(a) and 4(b). After this, the solution $\hat{\mathbf{f}}$ is calculated by minimizing $J(\mathbf{f}) = (\mathbf{f} - \mathbf{g})^2 + \varphi(||\mathbf{f}||)$ 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 $|\hat{\mathbf{f}}_2| > \theta$ for $\theta \approx 0.47$, and the global minimizer $\hat{\mathbf{f}}$ is found by searching the minimizer over the $x - y$ grid with the step size 0.01 in the $x-$ and $y-$ range between -10 and 10. The empirical histogram

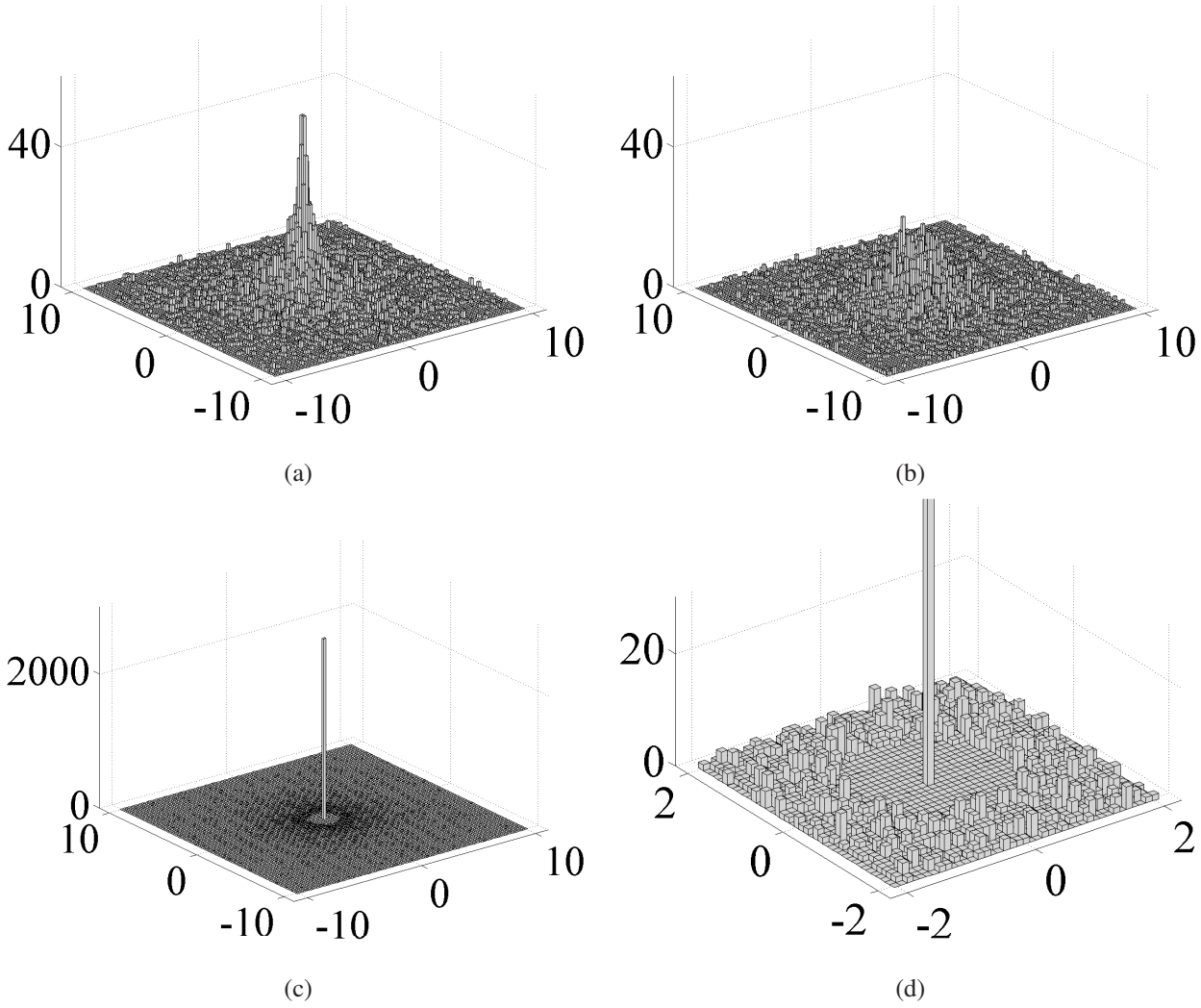


Fig. 4. (a) The samples of $\mathbf{f}$; (b) the observed samples of $\mathbf{g}$ by adding the noise to $\mathbf{f}$; (c) the restored samples and (d) the zoomed version of the restored samples.

of all solutions $\hat{\mathbf{f}}$, shown in Figure 4(c) and its zoomed version in Figure 4(d), these results illustrates the theorem, and we have $\hat{\mathbf{f}} = 0$ in 27% of the trials while the smallest non-zero $|\hat{\mathbf{f}}|$ is $1.18 > \theta$.

III. Minimization Methods

The minimization of nonconvex function J 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 for local minimization. Finally, the matrix H can have numerous nonzero elements beyond the diagonal and is often ill-conditioned. Global minimization of J can be considered using either stochastic algorithms or continuation-based deterministic relaxation.

Stochastic algorithms can be used to minimize nonconvex function is simulated annealing (SA) or Metropolis annealing. Since [24], 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 huge when H is not an identity, where in our problems, H may be a blurring operator or a reconstruction matrix which are not an identity. Therefore, it is not suitable to handle image restoration and reconstruction problems. Recently, Robini et al. [41] studied some 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 cost functional to speed up the convergence. However, the number of new parameters required in acceleration techniques must be handled properly. According to [48] , 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) algorithm [6] leads to minimizers $\hat{\mathbf{f}}$ having a lower (hence better) energy than SA [7]. However, this approach cannot properly address the nonsmoothness of J. Recently, Robini et al. [42] introduced a new class of hybrid algorithms that combines simulated annealing with deterministic continuation. Numerical experiments have shown that this approach outperforms standard simulated annealing. Nevertheless, stochastic algorithms cannot yield solutions that incorporate one of the main properties of nonsmooth regularization, namely the presence of constant regions in the restored image.

In [38], a nonsmooth GNC continuation method is proposed to solve the nonconvex nonsmooth minimization problem. Consider a sequence

$$\varepsilon_0 = 0 < \varepsilon_1 ... < \varepsilon_n = 1,$$

i.e. $\lim_{n\to\infty} \varepsilon_n = 1$, φ is approximated by a sequence of φ_ε s.t. φ_0 is convex and $\varphi_1 = \varphi$. Therefore the function J can be approximate by a sequence of function J_ε correspondingly:

$$J_\varepsilon(\mathbf{f}) = \|H\mathbf{f} - \mathbf{g}\|_2^2 + \beta \sum \varphi_\varepsilon(\|\mathrm{D}_i\mathbf{f}\|_2). \tag{15}$$

By splitting φ_ε into two terms, where the first one is a smooth but nonconvex term ψ_ε and the second one is a convex but nonsmooth total variational type regularization term $|t|$, i.e $\varphi_\varepsilon(t) = \psi_\varepsilon(t) + \alpha_\varepsilon|t|$ where $\psi_\varepsilon(t) = \varphi_\varepsilon(t) - \varphi_\varepsilon{}'(0^+)|t|$ and $\alpha_\varepsilon = \varphi_\varepsilon{}'(0^+)$, $J_\varepsilon(\mathbf{f})$ can be further rewritten as:

$$J_\varepsilon(\mathbf{f}) = \|H\mathbf{f} - \mathbf{g}\|_2^2 + \beta \sum \psi_\varepsilon(\|\mathrm{D}_i\mathbf{f}\|_2) + \beta\alpha_\varepsilon \sum \|\mathrm{D}_i\mathbf{f}\|_2 \tag{16}$$

which contains a data fidelity term with ℓ_2 norm, a nonconvex term and a nonsmooth TV regularization term. The main aim of this paper is to use a nonsmooth GNC continuation method for (15).

We remark in [38] that $\psi_\varepsilon(\|\mathrm{D}_i\mathbf{f}\|_2)$ and $\|\mathrm{D}_i\mathbf{f}\|_2$ are approximated by $\psi_\varepsilon(\mathbf{d}_i^T\mathbf{f})$ and $|\mathbf{d}_i\mathbf{f}|$ respectively, where $\mathbf{d}_i^T : \mathbb{R}^p \to \mathbb{R}$, for $i = 1, \cdots, r$, are linear operators. Usually they are the first-order differences between each pixel and its 4 or 8 adjacent neighbors. In such case, the problem can be reformulated using linear programming techniques. More precisely, this nonconvex nonsmooth minimization problem is given as the minimization of a nonconvex differentiable objective function under a set of linear constraints and it can be solved by the modified

primal-dual interior point method, which guarantees the descent direction at each step. Experimental results showed that this method can provide better performance compared to a simulated annealing method with significantly less computational cost. In this paper, the use of the ℓ_2 norm of the discrete gradient, i.e. $\|\mathrm{D}_i\mathbf{f}\|_2$ instead of $|\mathbf{d}_i^T\mathbf{f}|$ yields image restorations of better quality.

A. The proposed algorithm

Our idea is to tackle the difficulty for minimizing the function consists of both a nonconvex term and a nonsmooth term, we use variable-splitting and penalty technique to separate them with respect to two different variables. Based on the idea of [21], an auxiliary variable $\mathbf{u}$ is used to transfer the nonsmooth TV term from $\mathbf{f}$. A fitting term is then added to J_ε to ensure the closeness of $\mathbf{f}$ and $\mathbf{u}$. Therefore, we have

$$J_\varepsilon(\mathbf{f},\mathbf{u}) = \|H\mathbf{f}-\mathbf{g}\|_2^2 + \beta\sum\psi_\varepsilon(\|\mathrm{D}_i\mathbf{f}\|_2) + \omega\|\mathbf{f}-\mathbf{u}\|_2^2 + \beta\alpha_\varepsilon\sum\|\mathrm{D}_i\mathbf{u}\|_2 \tag{17}$$

where ω is the regularization parameter for the fitting term $\|\mathbf{f}-\mathbf{u}\|_2^2$. By fixing the variable $\mathbf{u}$, J_ε is a smooth function with respect to $\mathbf{f}$ so that it can be minimized by gradient-based methods. By fixing the variable $\mathbf{f}$, minimizing J_ε is equal to a TV denoising problem which can be solved by existing method efficiently [9]. We will demonstrate that the computational cost of this proposed method is greatly reduced compared with that of interior point method while the qualities of restored images are competitive by the experimental results in Section 4. We remark that another alternating minimization algorithm for TV restoration is proposed in [47]. This algorithm also based on variable-splitting and penalty technique to transfer the nonsmooth term out of J with respect to $\mathbf{f}$. The difference between the method in [21] and this method is that it fits $\mathbf{f}$ by $\mathrm{D}_i\mathbf{u}$ rather than $\mathbf{u}$.

Let us state the algorithm in detail. For each ε, starting from an initial guess $\mathbf{u}^{(0)}$, this method computes a sequence of iterates:

$$\mathbf{f}^{(1)},\mathbf{u}^{(1)},\mathbf{f}^{(2)},\mathbf{u}^{(2)}...,\mathbf{f}^{(i)},\mathbf{u}^{(i)},...$$

such that

$$\begin{aligned} S_\varepsilon^h(\mathbf{u}^{(i-1)}) &= \mathbf{f}^{(i)} = \arg\min_{\mathbf{f}} J_\varepsilon^h(\mathbf{f},\mathbf{u}^{(i-1)}) \qquad (18)\\ &\quad \text{where } J_\varepsilon^h(\mathbf{f},\mathbf{u}) = \|H\mathbf{f}-\mathbf{g}\|_2^2 + \beta\sum\psi_\varepsilon(\|\mathrm{D}_i\mathbf{f}\|_2) + \omega\|\mathbf{f}-\mathbf{u}\|_2^2 \\ S_\varepsilon^{tv}(\mathbf{f}^{(i)}) &= \mathbf{u}^{(i)} = \arg\min_{\mathbf{u}} J_\varepsilon^{tv}(\mathbf{f}^{(i)},\mathbf{u}) \qquad (19)\\ &\quad \text{where } J_\varepsilon^{tv}(\mathbf{f},\mathbf{u}) = \omega\|\mathbf{f}-\mathbf{u}\|_2^2 + \beta\alpha_\varepsilon TV(\mathbf{u}) \end{aligned}$$

for $i = 1, 2, \cdots$. The first step of this algorithm is to minimize J_ε^h. For $\varepsilon = 0$, we have

$$\sum\psi_\varepsilon(\|\mathrm{D}_i\mathbf{f}\|_2) = 0,\ \forall\mathbf{f},$$

minimizing J_ε^h is exactly the same as the total-variation deblurring problem in [21]. Therefore, we can follow the framework of [21] to minimize J_ε^h. For $\varepsilon > 0$, since all the terms in J_ε^h are differentiable, we can find out the gradient vector and the Hessian of J_ε^h to tackle the minimization problem:

$$\nabla_{\mathbf{f}} J_\varepsilon^h = 2H^T(H\mathbf{f}-\mathbf{g}) + \beta\psi_\varepsilon{}'\sum(\|\mathrm{D}_i\mathbf{f}\|_2)\cdot\partial(\|\mathrm{D}_i\mathbf{f}\|_2) + 2\omega(\mathbf{f}-\mathbf{u}^{(i-1)}) \tag{20}$$

where $\partial(\|\mathrm{D}_i\mathbf{f}\|_2)$ is the subgradient of $\|\mathrm{D}_i\mathbf{f}\|_2$, i.e.

$$\partial(\|\mathrm{D}_i\mathbf{f}\|_2) = \begin{cases} -\mathrm{div}(\mathrm{D}_i\mathbf{f}/\|\mathrm{D}_i\mathbf{f}\|_2), & \text{if} \|\mathrm{D}_i\mathbf{f}\|_2 \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

and

$$\nabla^2_{\mathbf{f}} J^h_\varepsilon = 2H^T H + 2\omega I + \beta \frac{\partial^2}{\partial \mathbf{f}^2} \psi_\varepsilon(\|\mathrm{D}_i\mathbf{f}\|)_2.$$

For the Hessian matrix of quasi-Newton method, $\nabla^2_{\mathbf{f}} J^h_\varepsilon$ may be not positive definite since involving the term $\frac{\partial^2}{\partial \mathbf{f}^2}\psi_\varepsilon(\|\mathrm{D}_i\mathbf{f}\|)_2$. In order to ensure the convergence of the algorithm, we simply use the positive definite part of the Hessian matrix in the optimization procedure. Because of 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 Quasi-Newton method can then be used to solve J^h_ε [39]. We remark that in image restoration H is usually a blurring matrix generated by a symmetric point spread function. The computational cost of the quasi-Newton method is dominated by three fast transforms for performing matrix-vector multiplication with H [31]. The computational cost for each fast transform is only $O(n^2 log 2n)$ for a $n^2 \times n^2$ blurring matrix [31]. After finding out the descent direction $\Delta\mathbf{f}^{(j)}$ by solving

$$(2H^T H + 2\omega I)\Delta\mathbf{f}^{(j-1)} = \nabla_{\mathbf{f}} J^h_\varepsilon,$$

the step size τ of the method is then determined by line search method in order to find out $\mathbf{f}^{(j)} = \mathbf{f}^{(j-1)} + \tau\Delta\mathbf{f}^{(j-1)}$. Three different conditions have been considered, including Armijo rule, Goldstein rule and a fixed τ [39]. However, the experimental results suggest us to use a fixed τ as it requires less computational cost compared with the other two line search methods. The second step of this method is to apply an exact TV denoising problem scheme to $\mathbf{f}^{(i)}$. Since the function J^{tv}_ε is identically the same as that in [21], we employ the Chambolle projection algorithm [9] to solve this problem. For ε_k, we simply choose a linear increase ε_k in our experiments as suggested in [38]. Let us summarize the proposed algorithm as follows:

- Let $\varepsilon_0 = 0 < \varepsilon_1 < \cdots < \varepsilon_n = 1$. Start from $\varepsilon(0)$. Initialize $\mathbf{u}^{(0)}$.
- For $k = 0 \rightarrow n$, $\varepsilon = \varepsilon_k$

  $j = 1$

  While residual $> tol$ do

  If $k = 1$

  Minimize $J^h_\varepsilon(\mathbf{f}, \mathbf{u}^{(j-1)})$ based on [21];

  Otherwise

  Solve $(2H^T H + 2\omega I)\Delta\mathbf{f}^{(j-1)} = \nabla_{\mathbf{f}} J^h_\varepsilon$ for $\nabla\mathbf{f}^{(j-1)}$;

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

  End If;

  Minimize J^{tv}_ε for $\mathbf{u}^{(j)}$;

  $j = j + 1$;

End While

$\mathbf{u}^{(0)} = \mathbf{u}^{(j-1)}$;

$\varepsilon(k+1) = \varepsilon(k) + \Delta\varepsilon$;

- End For

In the next section, we will test the performance of this proposed algorithm.

IV. Experimental Results

In this section, we present the experimental results to demonstrate the efficiency of the proposed method. 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 stopping criterion of our proposed method is that the relative difference between the successive iterate of the restored image should satisfy the following inequality:

$$\frac{\|\mathbf{f}^{(i+1)} - \mathbf{f}^{(i)}\|_2}{\|\mathbf{f}^{(i+1)}\|_2} < 10^{-4} \tag{21}$$

The PF used in all the illustrations is:

$$\varphi(t) = \frac{\alpha|t|}{1+\alpha|t|}, \;\; \varphi_\varepsilon(t) = \frac{\alpha|t|}{1+\varepsilon\alpha|t|}, 0 \le \varepsilon \le 1 \tag{22}$$

where it has been tested in [38]. We compare the proposed method to the modified interior point method in [38]. Conjugate gradient method is used to solve inner linear systems arising from outer iterations, and the stopping criterion of the conjugate gradient method is that the residual should be less than 10^{-5}. We remark our proposed method does not require to solve any inner linear systems.

Six images are tested, which are all gray images with intensity values ranging from 0 to 1. The first image is an artificial circles image of size 64×64 , and the second one is cameraman of size 256×256. These two images have been tested in [38]. The third and fourth images are F16 and tank images 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 two dimensional gaussian function

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

, which is truncated such that the function has a support of 7×7. The fifth image is the modified Shepp-Logan image of size 50×50 The sixth 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 methods were insensitive to all of the initial guesses. Therefore, we only demonstrate the results which the initial guesses are the observed images.

TABLE II

RESTORED SNRS AND CPU TIMES FOR DIFFERENT IMAGES.

	Interior point method		The proposed method	
Image	Computational time	SNR	Computational time	SNR
Circle	32.44	19.85	3.73	19.29
Cameraman	674.95	22.82	19.78	23.50
F16	3617.12	27.45	55.16	27.79
Tank	3467.75	31.00	39.75	30.73

According to the potential function, we can set $\alpha_\varepsilon = 0.5$. We tested different values of β and ω 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 to find out the restored image with the highest SNR.

A. Test of blurred images

Figures 5-8(a) show the original images. Figures 5-8(b) show their corresponding images with blur and noise as described in the above settings respectively. Figures 5-8(c) and 5-8(d) show the images restored by the modified interior point method and the proposed method respectively. For simplicity, we use the same set of parameters for restoring images in Figures 5-8(c) and 5-8(c), and also in Figures 5-8(d) and 5-8(d). We see from the figures that the images restored by the modified interior point method and the proposed method 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 the proposed method is significantly lower than the interior point method. These results demonstrate the proposed method is quite efficient for restoring images.

B. Test of Radon transform images

Radon transform is a two-dimensional 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 [49].

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 [38], we set the image to be of size 50×50. Figure 9(a) is the original modified Shepp-Logan image. The image is transformed along the angles from 0 to 180 of the increasing of 6 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 Figure 9(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

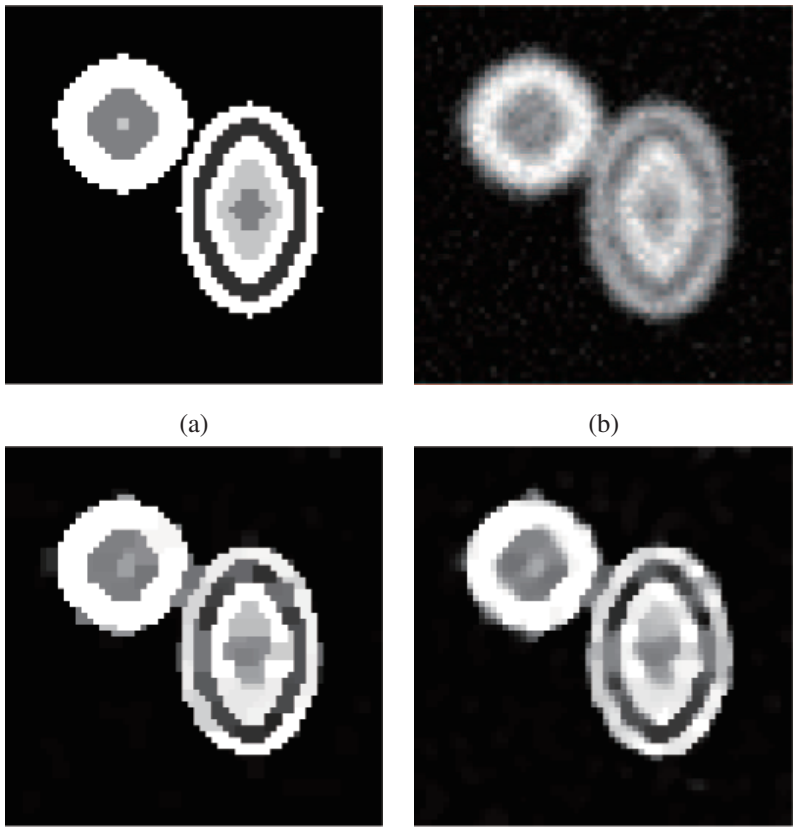


Fig. 5. (a) The original circle image; (b) the observed image; (c) the image restored by the interior point method with $\beta = 0.015$; (d) the image restored by the proposed method with $\omega = 0.5$ and $\beta = 0.003$.

use of the back-projection operator B [22]. 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 $(x^2 + y^2)^{-\frac{1}{2}}$ which can be used to construct the convolution matrix H. Now we solve

$$J_\varepsilon(\mathbf{f}, \mathbf{u}) = \|H\mathbf{f} - B\mathbf{g}\|_2^2 + \beta \sum \psi_\varepsilon(\|\mathrm{D}_i\mathbf{f}\|_2) + \omega\|\mathbf{f} - \mathbf{u}\|_2^2.$$

The back-projected Radon transform $B\mathbf{g}$ is shown in Figure 9(c). Figure 9(d) shows the resulting image that is reconstructed from Figure9(b) by the interior point method. Figure 9(e) shows the resulting image reconstructed from Figure 9(c) by the proposed method. Both of the methods provide high quality restored images. However, as conjugate gradient method is required to solve the linear system, the interior point method took more time (1860 seconds) to reconstruct the image. The proposed method only takes 0.5 seconds to restore about the same quality of the reconstructed image. The SNRs of the reconstructed image by the interior point method and the proposed method are 41.96dB and 42.52dB.

In the next experiment, we test a larger image. Figure 10(a) shows the original modified Shepp-Logan image of size 1000×1000. The corresponding back-projected Radon transform image is shown in Figure 10(b). Figure

(a) (b)

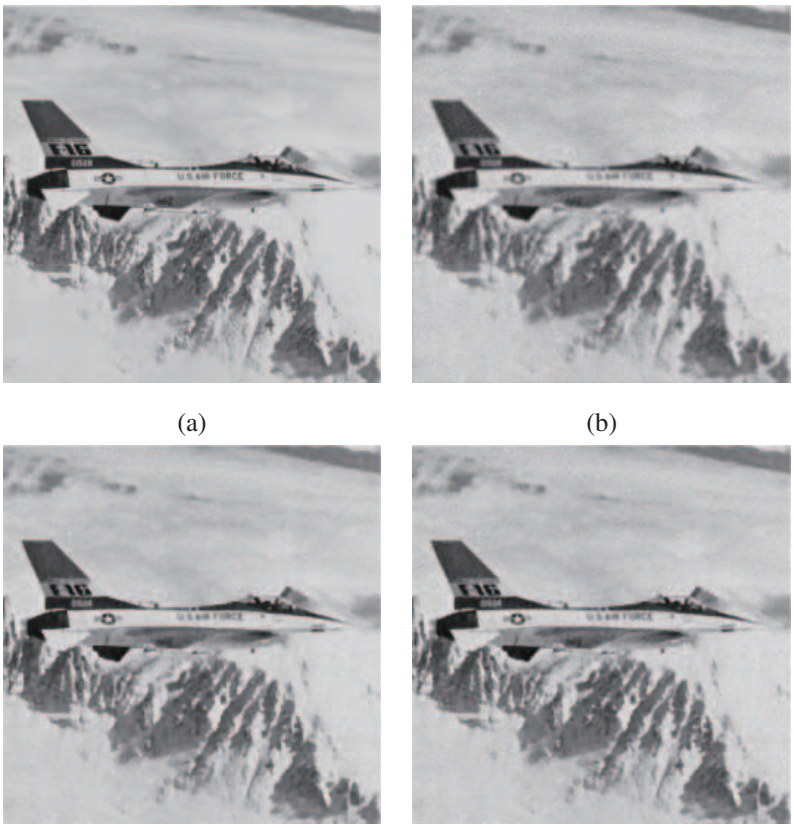
(c) (d)

Fig. 6. (a) The original cameraman image; (b) the observed image; (c) the image restored by interior point method with $\beta = 0.04$; (d) the image restored by the proposed method with $\omega = 2.0$ and $\beta = 0.007$.

10(c) shows the reconstructed image of the proposed method. The computational time required is 277 seconds, and the SNR is 46.78dB. However, the result of the interior point method is not obtained to this case as the required computational time takes more than 2 hours.

V. Concluding Remarks

Nonconvex regularization has advantages over convex regularization for restoring images with neat edges, see [38] and the section 2 of this paper. The main aim of this paper is to study a fast nonconvex and nonsmooth regularization on the ℓ_2 norm of the discrete gradient of the image minimization method for image restoration and reconstruction. Our theoretical results show that the solution of the nonconvex and nonsmooth regularization on the ℓ_2 norm of the discrete gradient of the image minimization problem can be composed of constant regions surrounded by closed contours and neat edges. We have developed a fast minimization algorithm to solve the nonconvex total variation minimization problem. Our experimental results show that the effectiveness and efficiency of the proposed method.



(a) (b)

(c) (d)

Fig. 7. (a) The original F16 image; (b) the observed image; (c) the image restored by interior point method with $\beta = 0.04$; (d) the image restored by the proposed method with $\omega = 2.0$ and $\beta = 0.007$.

REFERENCES

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

[2] J.-F. Aujol, G. Gilboa, T. Chan and S. Osher, "Structure-Texture Image Decomposition - Modeling, Algorithms, and Parameter Selection", *International Journal of Computer Vision*, vol. 67, n. 1, pp 111-136, 2006.

[3] M. R. Banham and A. K. Katsaggelos, "Digital image restoration," *IEEE Signal Processing Magazine,* vol. 14, pp. 24-41, 1997.

[4] J. E. Besag, "Digital image processing: towards Bayesian image analysis," *J. Appl. Stat.,* vol. 16, pp. 395-407, 1989.

[5] M. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications to early vision," *International Journal of Computer Vision,* vol. 19, pp. 57-91,1996.

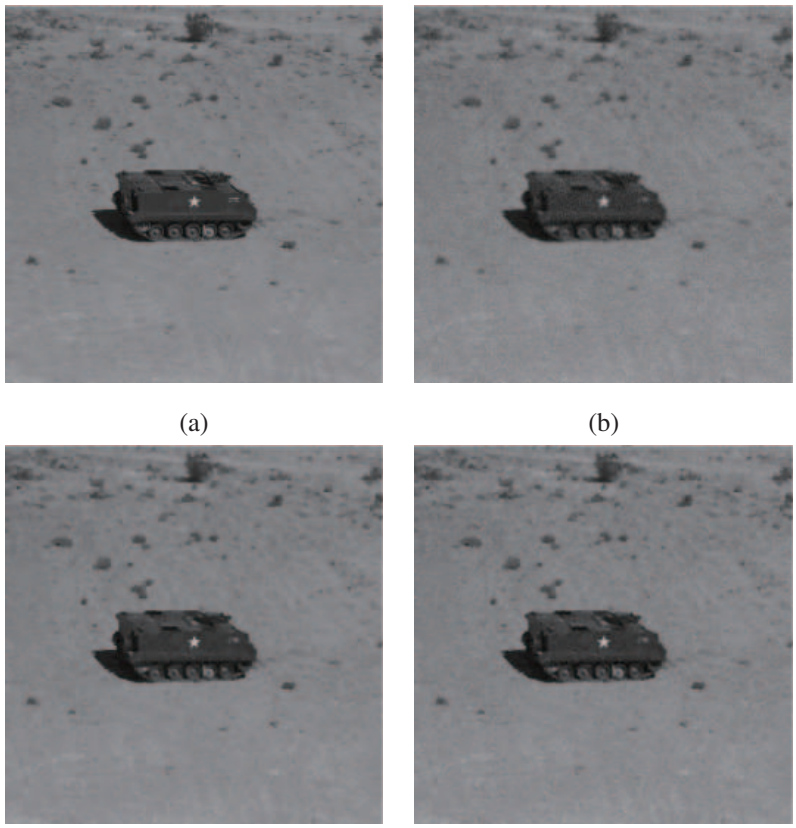
[6] A. Blake and A. Zisserman, *Visual Reconstruction,* Cambridge, MA: MIT Press, 1987.

[7] A. Blake, "Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction," *IEEE Trans. Pattern Anal. Machine Intell,.* vol. 11, pp. 2-12, Jan. 1989.

[8] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Processing,* vol. 2, pp. 296-310, 1993.

[9] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imaging Vision,* vol. 20, pp.89-97, 2004.

[10] C. L. Chan, A. K. Katsaggelos, and A. V. Sahakian, "Image sequence filtering in quantum-limited noise with applications to low-dose fluoroscopy," *IEEE Trans. Medical Imaging*, vol. 12, pp. 610-621, Sep. 1993.



(a) (b)

(c) (d)

Fig. 8. (a) The original tank image; (b) the observed image; (c) the image restored by interior point method with $\beta = 0.04$; (d) the image restored by the proposed method with $\omega = 2.0$ and $\beta = 0.007$.

[11] T. Chan and S. Esedoglu, "Aspects of total variation regularized ℓ_1 function approximation, " *Technical Report,* University of California at Los Angeles, CAM report, 2004.

[12] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Processing,* vol. 6, pp. 298-311, 1997.

[13] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structure and problems," *IEEE Trans. Acoustics Speech and Signal Processing,* vol. ASSP-37, pp. 2024-2036, Dec. 1989.

[14] 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, pp. 1881-1902, 2006.

[15] D. Geman and G.Reynolds, "Constrained restoration and recovery of discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.,* PAMI-14, pp. 367-383, 1992.

[16] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Processing,* vol.4, No.7, July, 1995.

[17] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributinons, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. PAMI-6, pp. 721-741, Nov. 1984.

[18] S. Geman and D. E. McClure, "Statistical methods for tomographic image reconstruction", *Proc. of the 46-th Session of the ISI, Bulletin of the ISI,* vol. 52, pp. 22-26, 1987.

[19] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imag.,* vol. 9,

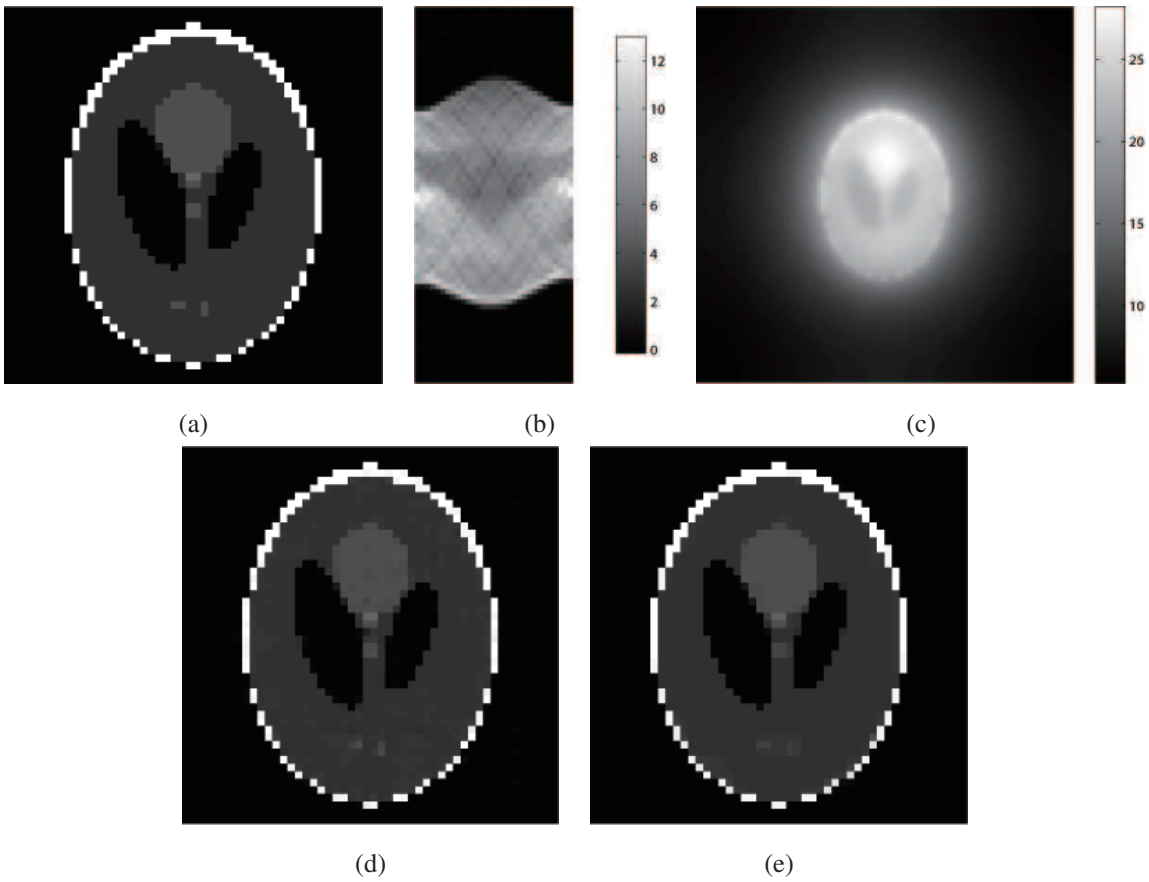


Fig. 9. (a) The original modified Shepp-Logan image with size 50×50; (b) the obtained image after Radon transform along the angles from 0 to 180 with the increasing of 6 degrees; (c) the back-projected Radon transform image; (d) the image restored from Figure 9(b) by the interior point method with $\beta = 0.8$; (e) the image restored from Figure 9(c) by the proposed method with $\omega = 0.4$ and $\beta = 0.002$.

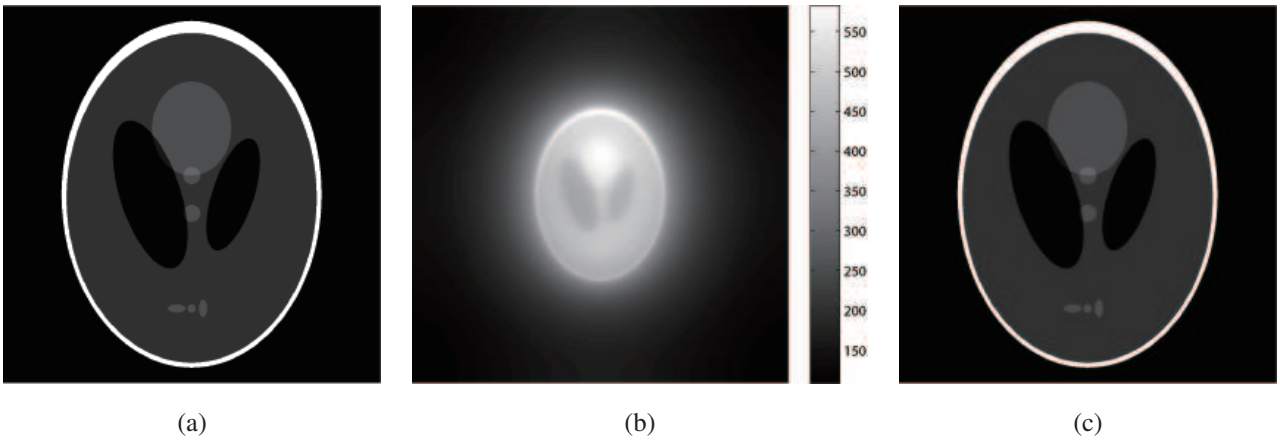


Fig. 10. (a) The original modified Shepp-Logan image with size 1000×1000; (b) the back-projected Radon transform image; (c) the image restored by the proposed method with $\omega = 0.4$ and $\beta = 0.002$.

pp. 84-93, Mar. 1990.

[20] X. Guo, F. Li and M. Ng, "A fast l1-TV algorithm for image restoration", SIAM Journal on Scientific Computing, 2009.

[21] Y. Huang, M. K. Ng, and Y. Wen, "A fast total variation minimization method for image restoration," *SIAM Journal on Multiscale Modeling and Simulation,* vol. 7, pp. 774-795, 2008.

[22] A. Jain, *Fundamentals of Digital Image Processing,* Prentice-Hall, Englewood Cliffs, NJ, 1989.

[23] T. K̀arkkäinen, K. Kunisch, K. Majava, "Denoising of smooth images using l^1-fitting," *Computing*, vol. 74, pp. 353 - 376, 2005.

[24] S. Kirkpatrick, C. Gelatt and M. Vecchi, "Optimization by simulated annealing,", *Science*, vol. 220, p.3, 1983.

[25] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs priors," *IEEE Trans. Med. Imag.,* vol. 9, pp. 439-446, 1990.

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

[27] Z. K. Liu and J. Y. Xiao, "Restoration of blurred TV pictures caused by uniform linear motion," *Comput. Vision, Graphics, Image Proc.,* vol. 44, pp. 30-34, 1988.

[28] D. Luenberger, *Linear and Nonlinear Programming,* Kluwer Academic Publishers, 2003.

[29] D. Mumford and J. Shan, "Boundary detection by minimizing functionals," *Proceedings of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing,* pp. 22-26, 1985.

[30] M. R. Nagel, "Introduction to evaluation of motion-degraded images," *Proc. of NASA Electronics Research Center Seminar,* Cambridge, MA, Dec. 3-5, 1968.

[31] 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, pp. 851-866, 1999.

[32] M. Nikolova, "Markovian reconstruction using a GNC approach," *IEEE Trans. on Image Processing,* vol. 8, n. 9, pp. 1204-1220, 1999.

[33] M. Nikolova and M. Ng, "Analysis of half-quadratic minimization methods for signal and image recovery", *SIAM Journal on Scientific Computing,* 27, pp. 937-966, 2006.

[34] M. Nikolova, "Minimizers of cost-functions involving non-smooth data-fidelity terms. Application to the processing of outliers," *SIAM Journal on Numerical Analysis,* vol. 40, pp. 965-994, 2002.

[35] M. Nikolova, "A variational approach to remove outliers and impulse noise,"*J. of Mathematical Imaging and Vision,* vol. 20, pp. 99-120, 2004.

[36] M. Nikolova, "Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares," *SIAM Journal on Multiscale Modeling and Simulation*, vol. 4, n. 3, pp. 960-991, 2005.

[37] M. Nikolova, "Analytical bounds on the minimizers of (nonconvex) regularized least-squares," *AIMS Journal on Inverse Problems and Imaging*, vol. 1, n. 4, pp. 661-677, 2007.

[38] M. Nikolova, M. K. Ng, S. Zhang, and W. Ching, "Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization," *SIAM J. Imaging Sciences,* vol. 1, pp. 2-25, 2008.

[39] J. Nocedal and S. Wright, *Numerical Optimization,* Springer, 1999.

[40] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Machine Intell.,* PAMI-12, pp. 629-639, 1990.

[41] M. Robini, T. Rastello and I. Magnin, "Simulated annealing, acceleration techniques, and image restoration," *IEEE Trans. on Image Processing*, vol. 8, pp. 1374-1387, 1999.

[42] M. Robini, A. Lachal, and I. Magnin, "A stochastic continuation approach to piecewise constant reconstruction", *IEEE Trans. on Image Processing*, 16 (2007), pp. 2576–2589.

[43] L. Rudin, S. Osher and C. Fatemi, "Nonlinear total variaion based noise removal algorithm," *Physica,* vol. 60, pp. 259-268, 1992.

[44] C. H. Slump, "Real-time image restoration in diagnostic X-ray imaging, the effects on quantum noise," *Proc, 11th IAPR Int. Conf. on Pattern Recognition,* vol. II, Conference B: Pattern Recognition Methodology and Systems, pp. 693-696, 1992.

[45] R. Stevenson and E. Delp, "Fitting curves with discontinuities," *Proc. of the 1st Int. Workshop on Robust Comput. Vision,* Seattle, WA, pp. 127-136, 1990.

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

[47] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *Accepted by SIAM Journal on Imaging Sciences,* 2008.

[48] E. Wasserstrom, "Numerical solutions by the continuation method," *SIAM Rev.,* vol. 15, pp. 89-119, Jan. 1973.

[49] M. Wintermark, M. Sesay, E. Barbier, K. Borbly, W. P. Dillon, J. D. Eastwood, T. C. Glenn, C. B. Grandin, S. Pedraza, J. F. Soustiel, T. Nariai, G. Zaharchuk, J. M. Caill, V. Dousset, H. Yonas, "Comparative overview of brain perfusion imaging techniques," *Stroke,* vol. 32, pp. 294-314, 2005.

APPENDIX

A. *Proof of Theorem 1*

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

$$\widehat{I}_0 = \{i \in I : \|\mathrm{D}_i\widehat{\mathbf{f}}\|_2 = 0\} \quad \text{and} \quad \widehat{I}_1 = I \setminus \widehat{I}_0, \tag{23}$$

as well as the vector subspace $K(\widehat{I}_0)$ given below:

$$\begin{aligned} K(\widehat{I}_0) &\stackrel{\text{def}}{=} \big\{v \in \mathbb{R}^p : \|\mathrm{D}_i v\|_2 = 0, \forall i \in \widehat{I}_0\big\} \\ &\equiv \big\{v \in \mathbb{R}^p : \mathrm{D}_i^k v = 0, 1 \leq k \leq s, \forall i \in \widehat{I}_0\big\}. \end{aligned} \tag{24}$$

If $\widehat{I}_1 = \varnothing$, then $\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2 = 0$, $\forall i \in I$, hence (9) is trivially true. In what follows we consider that $\widehat{I}_1 \neq \varnothing$. Note that if $\widehat{I}_0 = \varnothing$, then $K(\widehat{I}_0) = \mathbb{R}^p$.

Given a point $v \in \mathbb{R}^p$ and a constant $\rho > 0$, we denote by $B(v, \rho)$ the *open* ball in $\mathbb{R}^p$ centered at v of radius ρ. We will exhibit $\rho > 0$ such that the function $\mathcal{J}$ given below

$$\mathcal{J}(\mathbf{f}) = \|H\mathbf{f} - \mathbf{g}\|_2^2 + \beta \sum_{i \in \widehat{I}_1} \varphi(\|\mathrm{D}_i\mathbf{f}\|_2), \quad \text{for} \quad \mathbf{f} \in K(\widehat{I}_0) \cap B(\widehat{\mathbf{f}}, \rho)$$

is the restriction of J on $K(\widehat{I}_0) \cap B(\widehat{\mathbf{f}}, \rho)$. Put[1]

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

By (23) we see that $\rho > 0$. For every $v \in B(\widehat{\mathbf{f}}, \rho)$ we can hence write down that

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

which shows that $\widehat{\mathbf{f}}$ and $\widehat{\mathbf{f}} + v$ share the same set $\widehat{I}_1$. Since $\widehat{\mathbf{f}} \in K(\widehat{I}_0)$, we have

$$v \in K(\widehat{I}_0) \quad \Rightarrow \quad \|\mathrm{D}_i(\widehat{\mathbf{f}} + v)\|_2 = \|\mathrm{D}_i(\widehat{\mathbf{f}})\|_2 = 0, \quad \forall i \in \widehat{I}_0$$

and then

$$v \in K(\widehat{I}_0) \cap B(0, \rho) \quad \Rightarrow \quad \varphi(\|\mathrm{D}_i(\widehat{\mathbf{f}} + v)\|_2) = \varphi(\|\mathrm{D}_i(\widehat{\mathbf{f}})\|_2) = \varphi(0), \quad \forall i \in \widehat{I}_0.$$

It follows that

$$v \in K(\widehat{I}_0) \cap B(0, \rho) \quad \Rightarrow \quad \mathcal{J}(\widehat{\mathbf{f}} + v) = J(\widehat{\mathbf{f}} + v), \tag{25}$$

where $K(\widehat{I}_0)$ is defined in (24). Thus $\mathcal{J}$ is the restriction of J on $K(\widehat{I}_0) \cap B(\widehat{\mathbf{f}}, \rho)$ which entails that it has a (local) minimum at $\widehat{\mathbf{f}}$.

[1]Remind that for any matrix $A \in \mathbb{R}^{k,\ell}$ we have $\|A\|_2 = \max_{\|v\|_2 \leq 1} \|Av\|_2$ where $v \in \mathbb{R}^\ell$, see e.g. [28]. Hence $\|Av\|_2 \leq \|A\|_2\|v\|_2$.

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

$$d_v^- \mathcal{J}(\widehat{\mathbf{f}}) \leq 0 \leq d_v^+ \mathcal{J}(\widehat{\mathbf{f}}), \quad \forall v \in K(\widehat{I}_0) \setminus \{0\}, \tag{26}$$

must hold.

Let us denote

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

where M is the set in H3, p. 4. For any $i \in \widehat{I}_1 \setminus \widehat{I}_M$ and $\forall v \in \mathbb{R}^p$ we have[3]

$$\langle \nabla \varphi(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2), v \rangle = \varphi'(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2) \frac{\langle \mathrm{D}_i v, \mathrm{D}_i \widehat{\mathbf{f}} \rangle}{\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2} \tag{27}$$

and in particular $\langle \nabla \varphi(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2), \widehat{\mathbf{f}} \rangle = \varphi'(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2) \|\mathrm{D}_i \widehat{\mathbf{f}}\|_2$. Note that the gradient operator ∇ is considered with respect to $\widehat{\mathbf{f}}$.

Using H2 and H3 it is easy to find that for all $i \in \widehat{I}_M$ we have [4]

$$d_{\widehat{\mathbf{f}}}^+ \varphi(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2) = \varphi'(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2^+) \;\text{ and }\; d_{\widehat{\mathbf{f}}}^- \varphi(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2) = \varphi'(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2^-),$$

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

Considering the necessary condition (26) for $v = \widehat{\mathbf{f}} \in K(\widehat{I}_0)$ yields

$$\begin{aligned} 2\widehat{\mathbf{f}}^T H^T (H\widehat{\mathbf{f}} - \mathbf{g}) + \beta \sum_{i \in \widehat{I}_0 \setminus \widehat{I}_M} \varphi'(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2) \|\mathrm{D}_i \widehat{\mathbf{f}}\|_2 + \beta \sum_{i \in \widehat{I}_M} \varphi'(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2^-) \leq 0 \leq \\ \leq 2\widehat{\mathbf{f}}^T H^T (\widehat{\mathbf{f}} - \mathbf{g}) + \beta \sum_{i \in \widehat{I}_0 \setminus \widehat{I}_M} \varphi'(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2) \|\mathrm{D}_i \widehat{\mathbf{f}}\|_2 + \beta \sum_{i \in \widehat{I}_M} \varphi'(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2^+) \end{aligned}$$

This is equivalent to

$$\beta \sum_{i \in \widehat{I}_M} \varphi'(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2^-) \leq -2\widehat{\mathbf{f}}^T H^T (H\widehat{\mathbf{f}} - g) - \beta \sum_{i \in \widehat{I}_0 \setminus \widehat{I}_M} \varphi'(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2) \|\mathrm{D}_i \widehat{\mathbf{f}}\|_2 \leq \beta \sum_{i \in \widehat{I}_M} \varphi'(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2^+).$$

[2]Let us remind that these are defined by

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

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

[3]More precisely, $\forall i \in \widehat{I}_1 \setminus \widehat{I}_M$ we have

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

[4]In detail, the calculation is

$$\begin{aligned} d_{\widehat{\mathbf{f}}}^+ \varphi(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2) &= \lim_{t \searrow 0} \frac{\varphi\big(\|\mathrm{D}_i(\widehat{\mathbf{f}} + t\widehat{\mathbf{f}})\|_2\big) - \varphi(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2)}{t} = \lim_{t \searrow 0} \frac{\varphi\big((1+t)\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2\big) - \varphi(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2)}{t} = \varphi'(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2^+) \; ; \\ d_{\widehat{\mathbf{f}}}^- \varphi(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2) &= \lim_{t \searrow 0} \frac{\varphi\big(\|\mathrm{D}_i(\widehat{\mathbf{f}} - t\widehat{\mathbf{f}})\|_2\big) - \varphi(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2)}{-t} = \lim_{t \searrow 0} \frac{\varphi\big((1-t)\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2\big) - \varphi(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2)}{-t} = \varphi'(\|\mathrm{D}_i \widehat{\mathbf{f}}\|_2^-) \; . \end{aligned}$$

In particular, we must have

$$\sum_{i\in\widehat{I}_M}\varphi'(\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2^-)\leq\sum_{i\in\widehat{I}_M}\varphi'(\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2^+).$$

This is clearly impossible since by H3 we have $\sum_{i\in\widehat{I}_M}\varphi'(\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2^-)>\sum_{i\in\widehat{I}_M}\varphi'(\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2^+)$. It follows that

$$\widehat{I}_M=\varnothing\ \text{ and }\ \widehat{I}_0\setminus\widehat{I}_M=\widehat{I}_0.$$

The proof is complete.

B. Proof of Lemma 1

By H4, $\varphi''(t)<0$ for all $t\in[0,T)\setminus M$ while $\forall t\in M$, φ'' has finite left and right (negative) limits. Noticing also that $0\notin M$ shows that ξ in (11) is well defined and strictly negative on $[0,T)$.

By the same assumption,

$$\begin{aligned}\varepsilon>0,\ (t,t+\varepsilon)\in(\mathbb{R}_+\setminus M)^2\quad&\Rightarrow\quad 0\leq|\varphi''(t+\varepsilon)|\leq|\varphi''(t)|\ ,\\ t\in M\quad&\Rightarrow\quad 0<|\varphi''(t^+)|<|\varphi''(t^-)|<\infty.\end{aligned}$$

On the other hand,

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

Then

$$\begin{aligned}\varepsilon>0,\ (t,t+\varepsilon)\in(\mathbb{R}_+\setminus M)^2\quad&\Rightarrow\quad 0\leq|\xi(t+\varepsilon)|=\frac{|\varphi''(t+\varepsilon)|}{t+\varepsilon}\leq\frac{|\varphi''(t)|}{t}=|\xi(t)|\ ,\\ t\in M\quad&\Rightarrow\quad|\xi(t)|=\left[\frac{|\varphi''(t^-)|}{t},\frac{|\varphi''(t^+)|}{t}\right],\end{aligned}\tag{28}$$

and in particular,

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

Note that the multifunction in (28) is strictly decreasing at t. If $t=T\in M$, we have $|\varphi''(T^+)|=0$ and $|\varphi''(T^-)|>0$. It follows that $|\xi(.)|$ is strictly decreasing on $[0,T]$

Using that $\varphi''(t)=-|\varphi''(t)|$ on $\mathbb{R}_+\setminus M$, and that $\varphi''(t^-)=-|\varphi''(t^-)|$ and $\varphi''(t^+)=-|\varphi''(t^+)|$ on M shows that $\xi(t)=-|\xi(t)|$, $\forall t\in\mathbb{R}_+$. In particular, we deduce that ξ is strictly increasing on $[0,T]$.

The proof of statement (i) is complete.

Noticing that $\lim_{t\to\infty}\dfrac{1}{t}\searrow 0$ and using H4 yet again shows that

$$\lim_{t\to\infty}\xi(t)=0.$$

It is obvious that

$$\lim_{t\searrow 0}\xi(t)=-\infty.$$

Combining the last two limits with statement (i) leads immediately to (ii).

Statement (iii) is a straightforward consequence of (i) and (ii).

C. Proof of Theorem 2

With $\widehat{\mathbf{f}}$, we associate the subsets $\widehat{I}_0$ and $\widehat{I}_1$, as in (23), as well as the subspace $K(\widehat{I}_0)$ as introduced in (24). If $\widehat{I}_1 = \varnothing$, then $\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2 = 0$, $\forall i \in I$ in which case (12) is trivially satisfied. In what follows we consider that $\widehat{I}_1 \neq \varnothing$.

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

$$\langle \nabla^2\mathcal{J}(\widehat{\mathbf{f}})v, v\rangle \geq 0, \quad \forall v \in K(\widehat{I}_0). \tag{29}$$

Next we derive the expression of $\langle \nabla^2\mathcal{J}(\widehat{\mathbf{f}})v, v\rangle$. Using (27), we get

$$\langle \nabla^2\varphi(\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2)v, v\rangle = \varphi''(\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2)\left(\frac{\langle \mathrm{D}_i\widehat{\mathbf{f}}, \mathrm{D}_i v\rangle}{\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2}\right)^2 + \varphi'(\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2)\frac{\|\mathrm{D}_i v\|_2^2\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2^2 - \langle \mathrm{D}_i\widehat{\mathbf{f}}, \mathrm{D}_i v\rangle^2}{\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2^3}$$

The necessary condition (29) in detail reads

$$\begin{aligned}\langle \nabla^2\mathcal{J}(\widehat{\mathbf{f}})v, v\rangle &= 2\|Hv\|_2^2 + \beta\sum_{i\in\widehat{I}_1}\varphi''(\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2)\frac{\langle \mathrm{D}_i\widehat{\mathbf{f}}, \mathrm{D}_i v\rangle^2}{\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2^2} \\ &+ \beta\sum_{i\in\widehat{I}_1}\varphi'(\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2)\frac{\|\mathrm{D}_i v\|_2^2\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2^2 - \langle \mathrm{D}_i\widehat{\mathbf{f}}, \mathrm{D}_i v\rangle^2}{\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2^3} \geq 0, \ \forall v \in K(\widehat{I}_0).\end{aligned} \tag{30}$$

It might be useful to remind that $\mathrm{D}_i\widehat{\mathbf{f}} \in \mathbb{R}^s$ as well as that $\mathrm{D}_i v \in \mathbb{R}^s$, hence $\langle \mathrm{D}_i\widehat{\mathbf{f}}, \mathrm{D}_i v\rangle \in \mathbb{R}$.

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

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

$$\|\mathrm{D}_j\widehat{\mathbf{f}}\|_2 \leq \|\mathrm{D}_i\widehat{\mathbf{f}}\|_2, \quad \forall i \in \widehat{I}_1 \tag{31}$$

and set

$$\kappa \stackrel{\text{def}}{=} \|\mathrm{D}_j\widehat{\mathbf{f}}\|_2 \ . \tag{32}$$

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

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

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

$$\langle \mathrm{D}_i\widehat{\mathbf{f}}, \mathrm{D}_i\hat{v}\rangle = \|\mathrm{D}_i\widehat{\mathbf{f}}\|\,\|\mathrm{D}_i\hat{v}\|,$$

so the numerator in the last term in (30) is null:

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

Furthermore

$$\langle \mathrm{D}_i\widehat{\mathbf{f}}, \mathrm{D}_i\hat{v}\rangle^2 = \|\mathrm{D}_i\widehat{\mathbf{f}}\|\,\left(\kappa^{-1}\|\mathrm{D}_i\widehat{\mathbf{f}}\|\right)^3 \geq \|\mathrm{D}_i\widehat{\mathbf{f}}\|, \quad \forall i \in \widehat{I}_1.$$

Let us also define

$$\alpha \stackrel{\text{def}}{=} \max\left\{1,\, \kappa^{-3}\right\}, \tag{33}$$

where κ is defined in (32). Using that $\varphi''(\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2) \leq 0$, $\forall i \in \widehat{I}_1$, we get

$$\begin{aligned}
\langle \nabla^2 \mathcal{J}(\widehat{\mathbf{f}})\hat{v}, \hat{v}\rangle &\leq 2\,\kappa^{-3}\,\|H\widehat{\mathbf{f}}\|_2^2 + \beta \sum_{i\in\widehat{I}_1} \frac{\varphi''(\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2)}{\|\mathrm{D}_i\widehat{\mathbf{f}}\|_2} \\
&\leq 2\,\alpha\,\|H\widehat{\mathbf{f}}\|_2^2 + \beta\#\widehat{I}_1 \frac{\varphi''(\|\mathrm{D}_j\widehat{\mathbf{f}}\|_2)}{\|\mathrm{D}_j\widehat{\mathbf{f}}\|_2} \qquad (34) \\
&= 2\,\alpha\,\|H\widehat{\mathbf{f}}\|_2^2 + \beta\#\widehat{I}_1 \xi(\|\mathrm{D}_j\widehat{\mathbf{f}}\|_2)\ , \qquad (35)
\end{aligned}$$

where the element $\|\mathrm{D}_j\widehat{\mathbf{f}}\|_2$ is defined in (31) and ξ is the multifunction in (11) considered in Lemma 1.

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

$$\xi(\theta) = -\frac{2\,\alpha\,\|H\widehat{\mathbf{f}}\|_2^2}{\beta\#\widehat{I}_1}, \tag{36}$$

where α is defined in (33). By Lemma 1, this θ is well defined and unique. Suppose that

$$0 < \|\mathrm{D}_j\widehat{\mathbf{f}}\| = \kappa < \theta.$$

According to Lemma 1

$$-\infty = \lim_{t\searrow 0} \xi(t) < \xi(\|\mathrm{D}_j\widehat{\mathbf{f}}\|) < \xi(\theta) < 0. \tag{37}$$

This, jointly with (35) yields

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

The obtained result clearly contradicts the necessary condition stated in (30). It follows that

$$\|\mathrm{D}_j\widehat{\mathbf{f}}\|_2 \geq \theta, \tag{38}$$

where the inequality is strict if $\theta \in M$. Hence the result for θ as in (13).

Next we focus on the conditions given in (i). Combining [37, Theorem 2.3.] with Theorem 1 immediately yields that for any (local) minimizer $\widehat{\mathbf{f}}$ we have

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

Inserting this result into (34) yields

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

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

$$\xi(\theta) = -\frac{2\,\alpha\,\|\mathbf{g}\|_2^2}{\beta\#\widehat{I}_1}. \tag{39}$$

Then we apply the same reasoning that led us to (38) above. This proves statement (ii).

Statement (iii) comes from the observation that

$$\beta\#\widehat{I}_1 \frac{\varphi''(\|\mathrm{D}_j\widehat{\mathbf{f}}\|_2)}{\|\mathrm{D}_j\widehat{\mathbf{f}}\|_2} \leq \beta \frac{\varphi''(\|\mathrm{D}_j\widehat{\mathbf{f}}\|_2)}{\|\mathrm{D}_j\widehat{\mathbf{f}}\|_2},$$

where the equality takes place only if $\#I_1 = 1$. In such a case we choose θ that solves (36) or (39) where $\#I_1$ is replaced by 1. The resultant θ is smaller according to Lemma 1(iii).