Non-Lipschitz ℓ_p-Regularization and Box Constrained Model for Image Restoration

Xiaojun Chen, Michael K. Ng and Chao Zhang

***Abstract*—Nonsmooth nonconvex regularization has remarkable advantages for the restoration of piecewise constant images. Constrained optimization can improve the image restoration using a priori information. In this paper, we study regularized nonsmooth nonconvex minimization with box constraints for image restoration. We present a computable positive constant θ for using nonconvex nonsmooth regularization, and show that the difference between each pixel and its four adjacent neighbors is either 0 or larger than θ in the recovered image. Moreover, we give an explicit form of θ for the box constrained image restoration model with the non-Lipschitz nonconvex ℓ_p-norm ($0 < p < 1$) regularization. Our theoretical results show that any local minimizer of this imaging restoration problem is composed of constant regions surrounded by closed contours and edges. Numerical examples are presented to validate the theoretical results and show that the proposed model can recover image restoration results very well.**

***Index Terms*—Image restoration, regularization, nonsmooth and nonconvex, non-Lipschitz, box constraints**

I. Introduction

In this paper, we focus on the most common data production model for image restoration or reconstruction where the observed data $g \in R^m$ are related to the underlying $n \times n$ image, rearranged into a vector $f \in R^m$ ($m = n^2$), according to

$$g = Hf + \xi, \tag{1}$$

where $\xi \in R^m$ represents the noise and $H = [h_1, h_2, \ldots, h_m] \in R^{m\times m}$ represents a system matrix. For instance, when a blur is modeled by a point spread function, the matrix H is a block-Toeplitz-Toeplitz-block type matrix, see [1]. It is well-known that the matrix H is typically ill-conditioned. A regularization method should be used in the image restoration and reconstruction process. One usual approach is to determine the recovered image by minimizing a cost function z consisting of a data-fitting term and a regularization term:

$$\min z(f) := \|Hf - g\|^2 + \lambda \sum_{i=1}^{r} \varphi(d_i^T f), \tag{2}$$

where the superscript ^{T} stands for transpose, $r = 2(n-1)n$, $\|\cdot\|$ is the ℓ_2 norm in R^m, λ is a positive regularization parameter, $\varphi : R \to R$ is a potential function, and $d_i^T \in R^m$ is the ith-row of the first-order difference matrix in (7) which is used to define the difference between each pixel and its four adjacent neighbors.

Numerous expressions for potential functions $\varphi(t)$ have been used in the literature:

- Smooth, convex regularization: e.g. Tikhonov regularization [2], $\varphi(t) = t^2$. It has been shown in [1] that an efficient image restoration method based on fast transforms can be developed, and the computational cost is $O(m \log m)$ operations. However, the drawback of the Tikhonov regularization is that image edges cannot be preserved in the restoration process.
- Nonsmooth, convex regularization: e.g. total variation (TV) regularization [3], $\varphi(t) = |t|$. The distinctive feature of TV regularization is that image edges can be preserved. Thus TV regularization is in general more suitable than the Tikhonov regularization for image restoration purpose. We refer readers to [4] for recent developments of TV image restoration.
- Nonsmooth nonconvex regularization [5]–[8]: $\varphi(t)$ is a nonsmooth and nonconvex function, e.g., $\varphi(t) = |t|^p$ ($0 < p < 1$) which is a non-Lipschitz function, $\varphi(t) = \alpha|t|/(1+\alpha|t|)$ ($\alpha > 0$) which is a Lipschitz function. Nonsmooth nonconvex regularization offers a restored image composing of constant regions surrounded by closed contours and neat edges [8]–[10].

In this paper, we use a class of nonsmooth nonconvex potential functions φ, which satisfy the following assumption [8], [9].

Assumption I:

(a) φ is continuous, symmetric on $(-\infty, \infty)$, C^2 on $(0, \infty)$ and $\varphi(0) = 0$ is a strict minimum;
(b) $\varphi'(0^+) > 0$ and $\varphi'(t) \geq 0$ for all $t > 0$;
(c) φ'' is increasing on $(0, \infty)$ with $\varphi''(t) < 0$ and $\lim_{t\to\infty} \varphi''(t) = 0$.

Various existing nonsmooth nonconvex φ satisfy Assumption I, e.g., $\varphi(t) = |t|^p$ ($0 < p < 1$), and $\varphi(t) = \alpha|t|/(1+\alpha|t|)$ ($\alpha > 0$).

Xiaojun Chen is with the Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (maxjchen@polyu.edu.hk). The author's work was supported in part by Hong Kong Research Grant Council grant (PolyU5005/09P).

Michael K. Ng is with The Center for Mathematical Imaging and Vision and Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong (mng@math.hkbu.edu.hk). The research of this author is supported in part by the Hong Kong Research Grants Council Grant and HKBU FRGs.

Chao Zhang is with Department of Applied Mathematics, Beijing Jiaotong University, Beijing 100044, China (zc.njtu@163.com). The research of this author is supported in part by the National Natural Science Foundation of China (11001011), the National Basic Research Program of China (2010CB732501), and the Fundamental Research Funds for the Central Universities (2011JBM129).

In [9], Nikolova has proved that there exists constant $\lambda_0 \geq 0$ such that if $\lambda > \lambda_0$, then any local minimizer f^* of the unconstrained optimization problem (2) with a nonsmooth nonconvex potential function φ satisfying Assumption I offers recovery of neat edges, since the differences $d_i^T f^*$ are either shrunk and form homogeneous regions, or enhanced and form edges. That is, there exist constants $\lambda_0 > 0$ and $\theta > 0$, such that if $\lambda > \lambda_0$, then every local minimizer f^* satisfies

$$\text{either } |d_i^T f^*| = 0 \text{ or } |d_i^T f^*| \geq \theta, \ \forall i \in \{1, 2, \ldots, r\}. \quad (3)$$

In [8], [10], Nikolova et al. studied nonconvex nonsmooth minimization methods for image restoration and reconstruction, and developed fast minimization algorithms to solve the nonconvex nonsmooth minimization problem (2). Their experimental results showed the effectiveness and efficiency of their proposed algorithms. However, the constants λ_0 and θ in [9] are minimizers of two constrained optimization problems which are very difficult to compute. In practice, the values of λ_0 and θ are not known for image restoration problems.

Further adding pertinent constraints to f when minimizing the cost function $z(f)$ can help to restore the image from a priori information. The original image is comprised of nonnegative entries [11], [12]. For example, in image restoration, the pixels of the original image represent light intensities, and in PET, the pixels of the original image represent the number of photon pairs detected by the scanning device. This constraint and other ways of incorporating a priori information have been suggested in various applications, and can lead to substantial improvements in the image restoration and reconstruction [13]–[17].

This paper has the following two new contributions:

(i) We will incorporate with box constraints for image restoration process, and provide edge-preserving properties of a local minimizer f^* of the general box-constrained nonsmooth nonconvex optimization problem

$$\begin{array}{ll} \min & \|Hf - g\|^2 + \lambda \sum_{i=1}^{r} \varphi(d_i^T f) \\ \text{s.t.} & 0 \leq f \leq \kappa e, \end{array} \quad (4)$$

where κ is a positive upper bound parameter and $e = (1, \ldots, 1)^T \in R^m$. We will present easily computable positive constants θ and λ_0 such that any local minimizer f^* of (4) with $\lambda > \lambda_0$ satisfies (3), i.e., the difference between each pixel and its four adjacent neighbors is either 0 or larger than θ in the recovered image.

(ii) We show that the nonsmooth nonconvex and non-Lipschitz regularization term

$$\varphi(t) = |t|^p, \quad 0 < p < 1,$$

commonly used in image processing [18]–[20], can provide valuable edge-preserving properties of local minimizer f^*. In particular, we give an explicit form of θ for the box constrained image restoration model (4) with the non-Lipschitz nonconvex potential function,

$$\begin{array}{ll} \min & \|Hf - g\|^2 + \lambda \sum_{i=1}^{r} |d_i^T f|^p \\ \text{s.t.} & 0 \leq f \leq \kappa e \end{array} \quad (5)$$

for any $\lambda > 0$.

(iii) We propose the smoothing projected gradient (SPG) method [21] to solve the model (5), which is very easy to implement and efficient to solve large-scale nonconvex nonsmooth constrained minimization problem.

These theoretical results show that the solution of imaging restoration problems using nonconvex nonsmooth regularization is composed of constant regions surrounded by closed contours and neat edges. Moreover, these theoretical results can be extended to the following general box constrained problem

$$\begin{array}{ll} \min & \|Hf - g\|^2 + \lambda \sum_{i=1}^{r} \varphi(d_i^T f) \\ \text{s.t.} & \kappa_1 e \leq f \leq \kappa_2 e, \end{array} \quad (6)$$

where $\kappa_1 \in R \cup \{-\infty\}$, $\kappa_2 \in R \cup \{\infty\}$ and $\kappa_1 < \kappa_2$. Problem (6) includes the unconstrained optimization problem considered by Nikolova [9] as a special case.

The outline of the paper is as follows. In Section II, we give easily computable constants θ and λ_0 such that any local minimizer of (4) with $\lambda > \lambda_0$ satisfies (3). In Section III, we provide an explicit form of θ such that any local minimizer of (5) satisfies (3). We propose the SPG method to solve the proposed nonsmooth nonconvex minimization model in Section IV. Our numerical experimental results in this section show the effectiveness of the proposed model as well as the SPG algorithm for solving the model.

Notation: Throughout this paper, $\|\cdot\|$ denotes the ℓ_2 norm. For any set S, $|S|$ denotes the cardinality of S. Any vector $u \in R^n$ is considered as a column vector and $u_S \in R^{|S|}$ denotes the subvector of u whose entries lie in u indexed by S.

II. General Box-constrained Nonsmooth Nonconvex Regularization

In this section, we consider box-constrained nonsmooth nonconvex minimization problem (4) with a general regularization term which satisfies Assumption I. We show that all local minimizers f^* of (4) have edge-preserving properties, which are of both theoretical and practical importance.

Note that the vectors $d_1^T, \ldots, d_r^T$ are the rows of the first-order difference matrix $D \in R^{r \times m}$ defined as follows.

$$D = \begin{pmatrix} D_1 \otimes D_0 \\ D_0 \otimes D_1 \end{pmatrix}, \quad (7)$$

where $D_0 \in R^{n \times n}$ is the identity matrix,

$$D_1 = \begin{pmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{pmatrix} \in R^{(n-1) \times n},$$

and $\otimes$ is the Kronecker product [8]. Each row of D has only two nonzero entries and each column of D has at most four nonzero entries. The nonzero entries are either 1 or -1.

Let $\mathcal{F}^*$ be the set of all local minimizers of problem (4). It is clear that $\mathcal{F}^*$ is nonempty and bounded, since the feasible set

$$\mathcal{F} = \{f \mid 0 \leq f \leq \kappa e\}$$

is bounded, and the objective function of (4) is continuous by Assumption I (a).

Theorem 1: There exist constants $\lambda_0 \geq 0$ and $\theta > 0$ such that if $\lambda > \lambda_0$, then every local minimizer $f^* \in \mathcal{F}^*$ of (4) satisfies

$$\text{either } |d_i^T f^*| = 0 \text{ or } |d_i^T f^*| \geq \min(\theta, \kappa), \ \forall i \in \{1, 2, \ldots, r\}. \tag{8}$$

The proof of Theorem 1 can be found in Appendix. The purpose of Theorem 1 is to show the existence of λ_0 and θ such that (8) holds. Theorem 1 extends Theorem 3.3 of [9] by Nikolova for the unconstrained model (2) to the box constrained model (4). In general the constants λ_0 and θ in Theorem 1 are difficult to compute. Now we provide easily computable constants λ_0 and θ in the following theorem. The proof of Theorem 2 can be found in Appendix.

Theorem 2: Let $f^* \in \mathcal{F}^*$ be a local minimizer of (4), with $\lambda > \lambda_0 := \frac{2\|H^T H\| m}{|\varphi''(0^+)|}$. Let

$$\theta = \inf \left\{ t > 0 \ \mid \ \varphi''(t) = -\frac{2\|H^T H\| m}{\lambda} \right\}. \tag{9}$$

Then for any $i \in \{1, 2, \ldots, r\}$,

$$\text{either} \quad d_i^T f^* = 0 \quad \text{or} \quad |d_i^T f^*| \geq \min(\theta, \kappa). \tag{10}$$

Remark 1: If the potential function $\varphi(t)$, e.g., $\varphi(t) = |t|^p$, $(0 < p < 1)$, satisfying $|\varphi''(0^+)| = +\infty$, then (10) holds for any $\lambda > 0$. Moreover, following the proof of Theorem 2, we know that for a given local minimizer $f^* \in \mathcal{F}^*$, Theorem 2 still holds if we replace λ_0 and θ by $\hat{\lambda}_0$ and $\hat{\theta}$ respectively, where

$$\hat{\lambda}_0 := \frac{2\|H^T H\| \ |I|}{|\varphi''(0^+)|} \quad \text{and}$$
$$\hat{\theta} := \inf\{t > 0 \mid \varphi''(t) = -\frac{2\|H^T H\| \ |I|}{\lambda}\}.$$

Here $I = \{i \in \{1, 2, \ldots, m\} \mid 0 < f_i < \kappa\}$ denotes the inactive set of f, whose explanation in detail can be found at the beginning of Appendix. This enlarges the lower bound for θ in Theorem 1 in the case $|I| << m$.

Theorems 1 and 2 provide interesting theoretical justification that any local minimizer of the box-constrained nonsmooth nonconvex minimization model offers better possibilities of restoring images with neat edges. Using this result, we can consider the restoration of piecewise constant images where the number of the regions and their values are not fixed in advance from noisy data obtained at the output of a linear operator. Moreover, the constant θ in Theorem 2 can be given by solving a single equation if in addition φ'' is strictly increasing. For example, if $\varphi(t) = |t|^p, (0 < p < 1)$, then θ is the solution of $2m\|H^T H\| + \lambda p(p-1)|t|^{p-2} = 0$, that is, θ has an explicit form as

$$\theta = \left(\frac{\lambda p(1-p)}{2m\|H^T H\|} \right)^{\frac{1}{2-p}}.$$

Nikolova [9] first proved the existence of lower bounds of $|d_i^T f^*|$ for unconstrained minimization problem (2). However, to get the lower bound in [9], one has to solve a difficult minimization problem. In general, the lower bound in [9] has no explicit form.

III. The ℓ_p-regularization with Box-constraints

In this section, we will focus on (5) which has a regularization term $\varphi(t) = |t|^p$. Using special properties of $\varphi(t) = |t|^p$, we give an explicit form of the lower bound θ which is bigger than the bound given in Theorem 2.

For $k = 1, 2, \ldots, m$, we set the index sets

$$\mathcal{J}_k = \{ i \in \{1, \ldots, r\} \mid D_{ik} \neq 0\},$$
$$\mathcal{C}_k = \{ j \in \{1, \ldots, m\} \mid j \neq k, \ D_{ij} \neq 0 \text{ for some } i \in \mathcal{J}_k\},$$

where D_{ij} refers the (i, j)th entry of the first-order difference matrix D. We note that $\mathcal{J}_k$ indicates the rows which have nonzero entries at the kth column of D and $\mathcal{C}_k$ indicates the columns which have nonzero entries in such rows except the kth column. Both $\mathcal{J}_k$ and $\mathcal{C}_k$ have at most four index numbers. Moreover, by the structure of the matrix D, it is easy to find that

$$\mathcal{C}_k = \{k-1, k+1, k-n, k+n\} \cap \{1, 2, \ldots, m\},$$

which has indexes within the four adjacent neighbors of the kth pixel.

Let us denote the objective function of (5) by

$$z(f) = \|Hf - g\|^2 + \lambda \sum_{i=1}^{r} |d_i^T f|^p.$$

By the definition of $\|\cdot\|_p^p$, $z(f)$ can be written as

$$z(f) = \|Hf - g\|^2 + \lambda \|Df\|_p^p.$$

Recall that h_k represents the kth column of H. In the following, we provide lower bounds of $|d_i^T f^*|$ for certain indices i, which are defined by constants

$$\alpha_k := \left(\frac{\lambda p}{2\|h_k\| \sqrt{z(f^0)}} \right)^{\frac{1}{1-p}} \quad \text{and} \quad \beta_k = \left(\frac{\lambda p(1-p)}{2h_k^T h_k} \right)^{\frac{1}{2-p}},$$

for all $k = 1, \ldots, m$.

Theorem 3: Let $f^0 \in \mathcal{F}$ be an arbitrarily given feasible point, and f^* be a local minimizer of (5) satisfying $z(f^*) \leq z(f^0)$. If

$$f_k^* \leq \min\{f_i^*, i \in \mathcal{C}_k\} \quad \text{or} \quad f_k^* \geq \max\{f_i^*, i \in \mathcal{C}_k\}, \tag{11}$$

then either there exists $i \in \mathcal{C}_k$ such that $f_i^* - f_k^* = 0$ or

$$|f_i^* - f_k^*| \geq \alpha_k \qquad \text{for all} \quad i \in \mathcal{C}_k.$$

Remark 2: Assume that $f^0 \in \mathcal{F}$ is a good estimation of the original image, which, for example, may be the observed image or an acceptable guess after a certain degree of restoration. Most minimization methods start from such f^0 and reduce the function value $z(f)$ at each step to find a good restored image. Although (5) has many local minimizers, one may be only interested in these local minimizers f^* satisfying $z(f^*) \leq z(f^0)$.

Theorem 4: Let $f^0 \in \mathcal{F}$ be an arbitrarily given feasible point, and f^* be a local minimizer of (5) satisfying $z(f^*) \leq$

$z(f^0)$. Then for any entry f_k^* of f^*, either there exists $i \in \mathcal{C}_k$ such that $f_i^* - f_k^* = 0$, or for all $i \in \mathcal{C}_k$, $|f_k^* - f_i^*| \geq \min\{\alpha_k, \beta_k\}$. In particular, we have

$$|f_k^* - f_i^*| \geq \beta_k, \quad \text{if} \quad 0 < f_k^* < \kappa$$
$$|f_k^* - f_i^*| \geq \alpha_k, \quad \text{if} \quad f_k^* = 0, \text{ or } f_k^* = \kappa.$$

The proofs of Theorem 3 and Theorem 4 can be found in Appendix.

Remark 3: In [22], Chen et al. presented a lower bound theory for the unconstrained ℓ_2-ℓ_p minimization problem:

$$\min \|Hf - g\|^2 + \lambda \|f\|_p^p.$$

In this paper, we derive new lower bounds for the box constrained ℓ_2-ℓ_p minimization problem (5) where the regularization term is $\lambda\|Df\|_p^p$. Our new results can be considered as an extension of the lower bound theory in [22]. Such extension is interesting as the box constrained ℓ_2-ℓ_p minimization problem (5) has important applications in image restoration. The employing of box constraints has been suggested in various image restoration applications, and it can lead to substantial improvements in image restoration [13], [16], [17]. Moreover, deriving new lower bounds for (5) is not trivial, since we have to consider the constraints and the relation among components of any local minimizer.

IV. Numerical Results

In this section, we present the numerical results on seven experiments to validate the theoretical results and show that the proposed model can recover the original image from its degraded image well, especially for piecewise constant images.

We first give a simulation experiment on images of only two pixels to demonstrate the theoretical results in Sections II and III. We then employ the smoothing projected gradient (SPG) method [21] to solve the nonsmooth nonconvex constrained minimization problems. We perform three numerical experiments on the restorations of blurred and noisy images, including both synthetic and true images which are (nearly) piecewise constant.

While many natural images are not piecewise constant regions surrounded by edges. Theoretical results of this paper indicate that the restoration results on gradual changed region would create piecewise constant image. Hence we also test our model on two often used images: Cameraman image and Barbara image, as well as the Books image with gradual shading caused by illumination.

All testing images are transformed to gray level images of intensity values ranging from 0 to 1. Each observed image is then blurred by a two-dimensional Gaussian function, and then added a Gaussian noise with the zero mean and the given standard derivation. In the numerical experiments, the two-dimensional Gaussian function is set to be:

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

which is truncated such that the function has a support of 7×7, and is normalized to be equal to 1.

In the comparison, we consider minimizing $z(f)$ in (2) without constraints, with one-side constraints $f \geq 0$ and two-side constraints $0 \leq f \leq e$, respectively. For the first two cases, we further truncate the solutions onto the feasible region $[0, e]$ to justify the usefulness of the box constraints.

The SPG method deals with nonsmooth nonconvex constrained minimization problem, which combines the smoothing strategy and the classic projected gradient method. It is especially attractive for solving large-scale box-constrained problems. To implement the SPG method, we need construct the smoothing function of the nonsmooth objective function.

In numerical experiments, we use the two potential functions φ defined by

$$\varphi(t) = \frac{0.5|t|}{1 + 0.5|t|} \quad \text{and} \quad \varphi(t) = |t|^p, \ 0 < p < 1.$$

The nonsmooth nonconvex regularization term $\varphi(t)$ involves $|t|$. We first provide a smooth approximation $s_\mu(t)$ for $|t|$ by

$$s_\mu(t) = \begin{cases} |t| & \text{if } |t| > \mu \\ \frac{t^2}{2\mu} + \frac{\mu}{2} & \text{if } |t| \leq \mu, \end{cases}$$

with a smoothing parameter $\mu > 0$. It is easy to check that

$$0 \leq s_\mu(t) - |t| \leq \frac{\mu}{2},$$

and hence $\lim_{\mu \downarrow 0} s_\mu(t) = |t|$. The smoothing function $\tilde{z}_\mu(f)$ of $z(f)$ can then be defined by replacing $\varphi(t)$ by $\varphi(s_\mu(t))$. For instance, when $\varphi(t) = |t|^p$ is employed, we set

$$\tilde{z}_\mu(f) = \|Hf - g\|^2 + \lambda \sum_{i=1}^{r} (s_\mu(d_i^T f))^p.$$

It is easy to see that the maximum difference between $s_\mu(t)$ and $|t|$ is at $t = 0$. Hence

$$0 \leq \tilde{z}_\mu(f) - z(f) \leq \lambda \sum_{i=1}^{r} (\frac{\mu}{2})^p \leq \lambda r (\frac{\mu}{2})^p,$$

which implies

$$\lim_{\mu \downarrow 0} z_\mu(f) = z(f).$$

The parameters of the SPG method are chosen to be

$$\sigma = 0.5; \ \sigma_1 = \sigma_2 = 10^3; \ \rho_2 = 0.25; \ \rho_3 = 10^3,$$

and $\hat{\rho} = 10^3$ or 10^5. We stop the SPG method if it reaches a maximum iteration $k_{\max}$ or $\mu < 10^{-5}$.

A. Test of simulated image

Let $f = (f_1, f_2)$ be a nonnegative image of two pixels, $D = (1, -1)$, $H = I$ and $\varphi(t) = |t|^p$. We generate 1000 samples of f independently, which are uniformly distributed in $[0,1]\times[0,1]$. Then we generate the observed image $g_i = f_i + \epsilon$ where the noise ϵ following the normal distribution with mean zero and standard deviation 0.5, i.e., $\epsilon \sim \text{Normal}(0, 0.5)$. We obtain the global optimal solution $f^* = (f_1^*, f_2^*)$ of

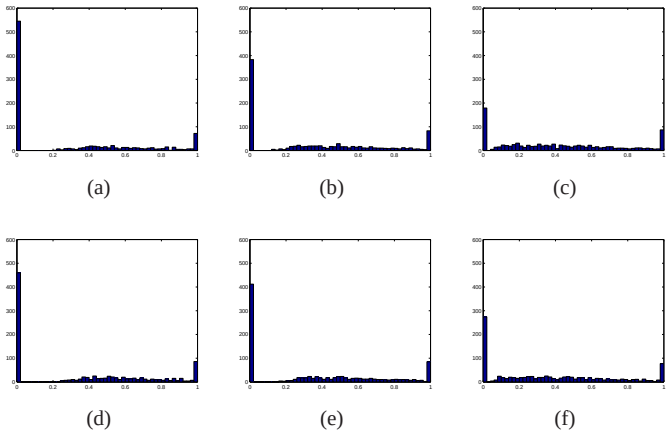
$$\begin{aligned} \min \quad & z(f_1, f_2) = (f_1 - g_1)^2 + (f_2 - g_2)^2 + \lambda |f_1 - f_2|^p \\ \text{s.t.} \quad & 0 \leq f_1, f_2 \leq 1. \end{aligned}$$

TABLE I: Parameters and Results for Fig. 1.

Subfigure	p	λ	θ	δ
(a)	0.5	0.2	0.0855	0.22
(b)	0.5	0.1	0.0539	0.14
(c)	0.5	0.02	0.0184	0.05
(d)	0.2	0.1	0.0684	0.25
(e)	0.4	0.1	0.0630	0.17
(f)	0.8	0.1	0.0179	0.05

The optimal solution is found by exhaustive search over $[0,1] \times [0,1]$ with the step size 0.01 in each direction. By Theorem 4, any local minimizer f^* satisfies

$$\text{either } f_1^* - f_2^* = 0, \text{ or } |f_1^* - f_2^*| \geq \theta = \left(\frac{\lambda p(1-p)}{2}\right)^{\frac{1}{2-p}}.$$



(a) (b) (c) (d) (e) (f)

Fig. 1: Histograms for absolute difference between f_1^* and f_2^*.

In Fig. 1, we show the histograms of $|f_1^* - f_2^*|$ for the 1000 generated samples using different p and λ. Their corresponding parameters p and λ, as well as the predicted threshold θ by Theorem 4, and the smallest nonzero absolute difference

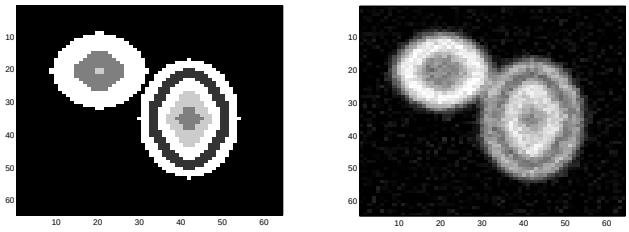
$$\delta = \min_{i=1,2,\ldots,1000} |f_1^*(\epsilon^i) - f_2^*(\epsilon^i)|$$

in the 6 cases are displayed in Table I.

Those nonzeros are larger than θ as predicted by Theorem 4. We observe that when λ decreases or p increases, the number of zero absolute differences decreases and the smallest nonzero entry tends nearer to zero at the same time. Since the effect of clear distinction of zero and nonzero differences is weakened as $\lambda \downarrow 0$, or $p \uparrow 1$, it might not be a good choice of very small λ or very big p for recovering the piecewise constant image with sharp varying regions.

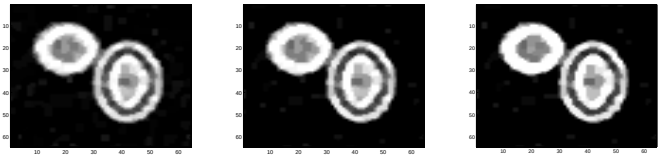
B. Test of Circles image

The Circles image is of size 64×64. The original Circles image and the observed image are shown in Fig. 2, which are used in [8]. The standard derivation of a Gaussian noise added to the blurred image is 0.05 in this experiment.



(a) original (b) observed (PSNR = 15.50dB)

Fig. 2: The Circles image of size 64×64.



(a) $f \in R^m$ (b) $f \geq 0$ (c) $0 \leq f \leq e$

Fig. 3: Image restoration results using $\varphi(t) = |t|^{0.5}$, and $\lambda = 0.003$. For reference, the PSNRs with / without projection onto $[0,1]$ are (a) 18.45dB / 18.16dB; and (b) 19.24dB / 19.13dB; and (c) 19.97dB.

We use $\varphi(t) = |t|^{0.5}$, the initial smoothing parameter $\mu_0 = 1$, and the maximum number of iteration $k_{\max} = 5000$ in SPG for this experiment. Figure 3 shows that restored images without constraints ($f \in R^m$), with one-sided constraints ($f \geq 0$) and with box constraints ($0 \leq f \leq e$). Here we apply the projection of the restored image pixel values to $[0,1]$. We display the restored image by selecting a suitable value of regularization parameter λ such that the corresponding PSNR is the highest. Visually, we see that the quality of the restored image with box constraint is better than the other two restored images. Also the PSNR of the restored image with box constraints is higher than those of the other two restored images. We also observe that the PSNRs of the restored images with the projection onto the box feasible region are slightly larger than those without the projection, see Figs. 3 (a) and (b). However, the PSNR improvement is more significant when we employ the box constraints in the proposed model.

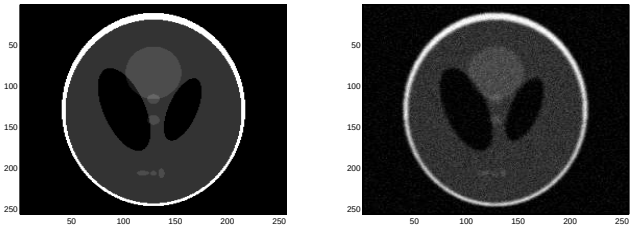
By Theorem 2, the threshold θ corresponding to $p = 0.5$ can be computed by

$$\theta = \left(\frac{\lambda}{512\|H^T H\|}\right)^{\frac{2}{3}}.$$

In our case, $\lambda = 0.003$ and $\|H\| = 1$, therefore the value of θ is equal to 3.25×10^{-4}. We check all the absolute differences $|d_i^T f|$ for $i = 1, 2, \ldots, 8064$, where f is the computed solution obtained by the SPG method using box constraints. We find that 6615 entries of $|Df|$, i.e., 82.03% of the absolute differences are close to 0, with the average to be 5.95×10^{-6}, as predicted by our theoretical results. The remaining nonzeros are, in fact, much larger than the threshold 3.25×10^{-4}, where the minimal entry is 0.0023 and the average is 0.2645.

C. Test of modified Shepp-Logan image

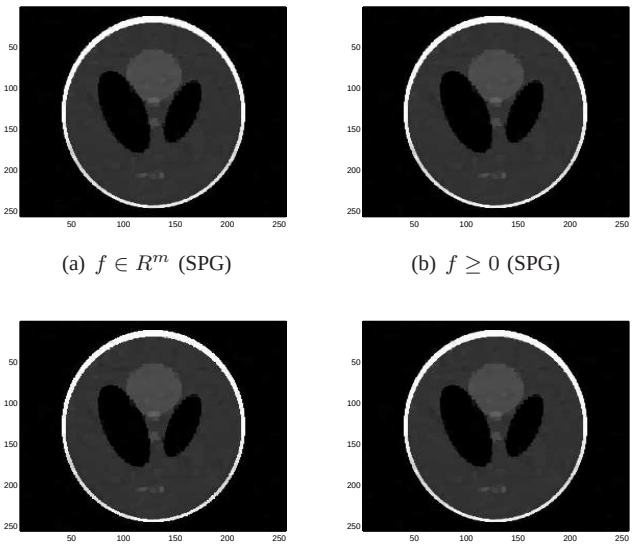
We use a modified Shepp-Logan image of size 256×256 shown in Fig. 4 to test the performance of the two potential nonsmooth nonconvex functions φ as well as the nonsmooth convex function $\varphi(t) = |t|$ which is used in TV regularization. The standard derivation of a Gaussian noise added to the blurred image is 0.05 in this experiment. We also compare the SPG method with the graduated nonconvexity (GNC) algorithm in [8], which is proposed for nonsmooth nonconvex minimization arising from image restoration.



(a) (b)

Fig. 4: The Shepp-Logan image of size 256×256.

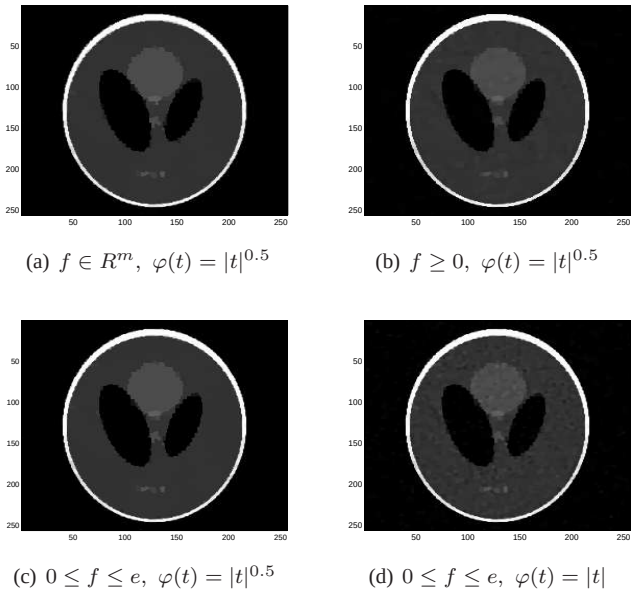
The initial smoothing parameter $\mu_0 = 10$, and the maximum number of iteration $k_{\max} = 5000$ for this experiment. The restored images are shown in Figs. 5 and 6. According to the figures, we find that the use of box constraints $0 \leq f \leq e$ can provide a better image restoration with a higher PSNR. Similarly as the Circles image, the projection of the solutions in the cases of without constraints and with only nonnegative constraints do help to improve the PSNR. However, the use of the box constraints provides the highest PSNR in the image restoration.



(a) $f \in R^m$ (SPG) (b) $f \geq 0$ (SPG)

(c) $f \geq 0$ (nonsmooth GNC) (d) $0 \leq f \leq e$ (SPG)

Fig. 5: Image restoration results using $\varphi(t) = \frac{0.5|t|}{1+0.5|t|}$, and $\lambda = 0.05$: (a) PSNR = 27.25dB; (27.21dB without projection); (b) PSNR = 27.64dB; (27.63dB without projection); (c) PSNR = 26.88dB; and (d) PSNR = 28.04dB.



(a) $f \in R^m,\ \varphi(t) = |t|^{0.5}$ (b) $f \geq 0,\ \varphi(t) = |t|^{0.5}$

(c) $0 \leq f \leq e,\ \varphi(t) = |t|^{0.5}$ (d) $0 \leq f \leq e,\ \varphi(t) = |t|$

Fig. 6: Image restoration results using SPG with $\lambda = 0.012$: (a) PSNR = 26.65dB (26.60dB without projection); (b) PSNR = 27.01dB (26.96dB without projection); (c) PSNR = 27.73dB; and (d) PSNR = 27.79dB;

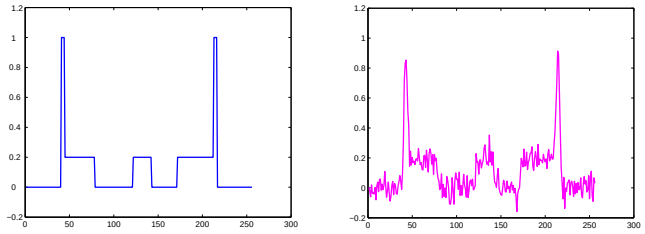
In order to see the edge-preserving property by the two different nonsmooth nonconvex potential functions and the nonsmooth convex potential function $\varphi(t) = |t|$, we display in Figs. 7 and 8 the 126th and 255th lines of the restored images in Fig. 5 (d) and Fig. 6 (c), (d). For a comparison, the original and blurred noisy lines are also displayed in Figs. 7 and 8 (a) and (b). We observe from the figures that both choices of nonsmooth nonconvex potential functions $\varphi(t)$ lead to the restored images with neat edges and correct amplitude. The restored image by using $\varphi(t) = \frac{0.5|t|}{1+0.5|t|}$, which is Lipschitzian at zero, has a few slightly blurred edges. While the restored image using $\varphi(t) = |t|^{0.5}$, which is non-Lipschitzian at zero, fits the original section quite well. It is easy to see that the restored image using the nonsmooth convex function $\varphi(t) = |t|$ has more blurred edges.

D. Test of MRI images

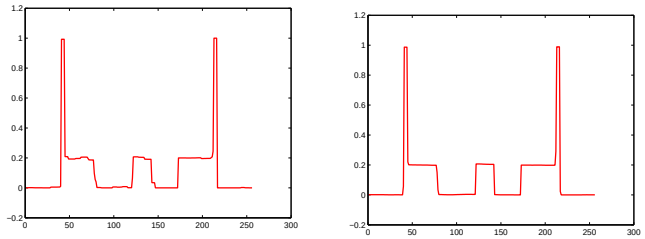
The experiment in this subsection is based on real data where they are 15 two-dimensional (2D) slices of 512×512 magnetic resonance imaging (MRI) scans for diagnosis of abdominal aortic aneurysm. In this experiment, we use 0.2 for the standard derivations of the Gaussian noise. The initial smoothing parameter of the SPG method is $\mu_0 = 10$, and the maximum number of iteration $k_{\max} = 1000$ for this experiment.

We show image restoration results for such fifteen MRI slices. The test was carried out on a Dell PC (3.00GHz, 2.00GB of RAM) with the use of Red Flag Linux Desktop 6.0 and Matlab R2009a (Version 7.8.0.347). We draw the highest PSNRs that can be obtained during the recovery process of the fifteen MRI slices by different methods in Fig. 9.

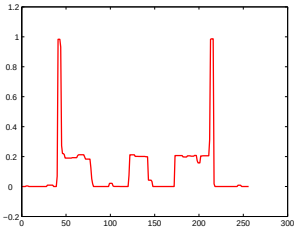
From the figure, we see that the use of box constraints in the proposed model provides better PSNRs than the other



(a) original (b) observed

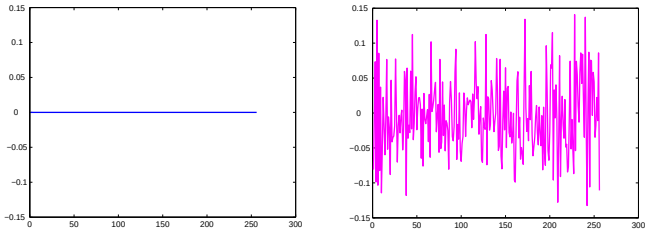


(c) $\varphi(t) = \frac{0.5|t|}{1+0.5|t|}$ (d) $\varphi(t) = |t|^{0.5}$

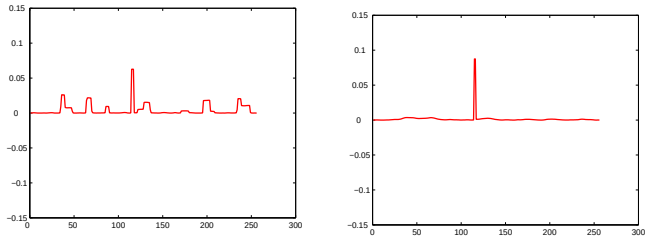


(e) $\varphi(t) = |t|$

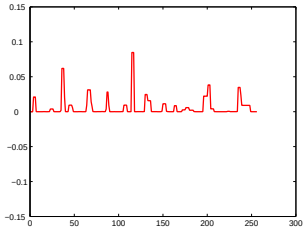
Fig. 7: The original, observed and restored 126th lines.



(a) original (b) observed



(c) $\varphi(t) = \frac{0.5|t|}{1+0.5|t|}$ (d) $\varphi(t) = |t|^{0.5}$



(e) $\varphi(t) = |t|$

Fig. 8: The original, observed and restored 255th lines.

TABLE II: Average computational results of the 15 slices with $\sigma = 0.2$.

Models	aver-PSNR	aver-cpu
$f \in R^m$	19.63 / 20.63 (without / with projection)	$1130sec.$
$f \geq 0$	20.59 / 20.82 (without / with projection)	$1020sec.$
$0 \leq f \leq e$	20.95	$886sec.$

methods. We also provide in Table II the information for the average PSNR (aver-PSNR), and the average computational time (aver-cputime) in seconds in order to obtain the restored MRI slices. We see in Table II that the proposed method is faster than the other methods in order to obtain about the same aver-PSNR of restored slices.

As an example, we show the original, the observed and the restored image of one MRI slice in Fig. 10. The restored image using $\varphi(t) = |t|^{0.5}$ by SPG method.

In subsections A-D, we perform numerical experiments on images that are (nearly) piecewise constant and obtain good recovery results using our box constrained nonsmooth nonconvex model. This validates the theoretical results given in this paper.

While many natural images are not piecewise constant images surrounded by edges. In the following subsections, we

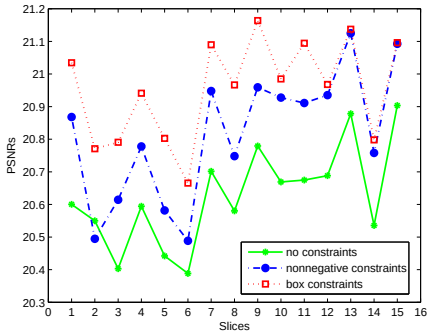
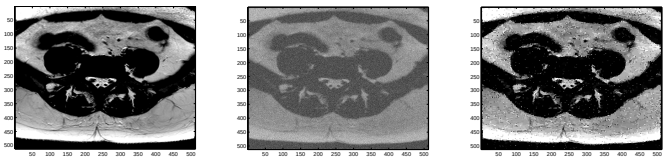


Fig. 9: PSNRs of the restored 15 MRI slices with $\sigma = 0.2$ for the Gaussian noise.

test our model on two common used images and an image with gradual change region that are not piecewise constant to see its performance.

The SPG method is employed to solve the minimization problem using the two nonsmooth nonconvex potential functions as well as the nonsmooth nonconvex function $\varphi(t) = |t|$. The initial smoothing parameter is set to be $\mu_0 = 10$ in the SPG method and the maximum number of iteration is set to be $k_{\max} = 2000$.



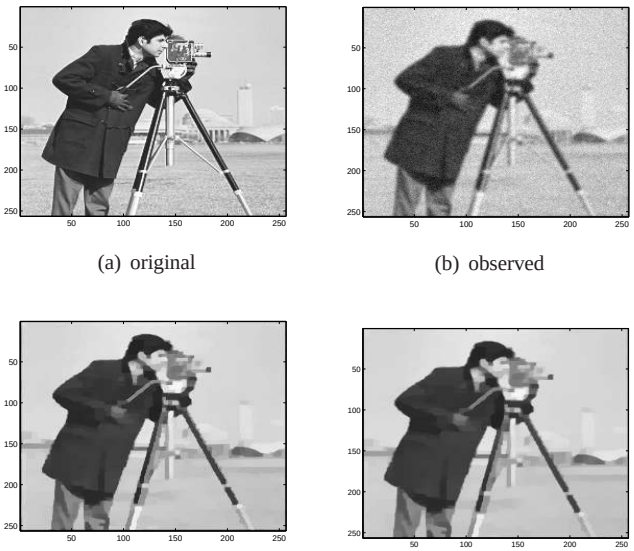
(a) Original (b) Observed (c) Restored

Fig. 10: Image restoration results using $\varphi(t) = |t|^{0.5}$ and $\lambda = 0.035$; The PSNRs are (b) 13.80dB and (c) 21.03dB.

E. Test of Cameraman image

The Cameraman image of size 256×256, which has more edges. In this experiment, we use 0.05 for the standard deviation of the Gaussian noise.

We employ the potential function $\varphi(t) = \frac{0.5|t|}{1+0.5|t|}$, as well as the potential function $\varphi(t) = |t|$ that used in TV regularization. We see from Fig. 11 (c)-(d) that the piecewise constant regions appear in the restoration results. The restoration using the nonsmooth nonconvex potential function $\varphi(t) = \frac{0.5|t|}{1+0.5|t|}$ provides higher PSNR than that using $\varphi(t) = |t|$.



(a) original (b) observed

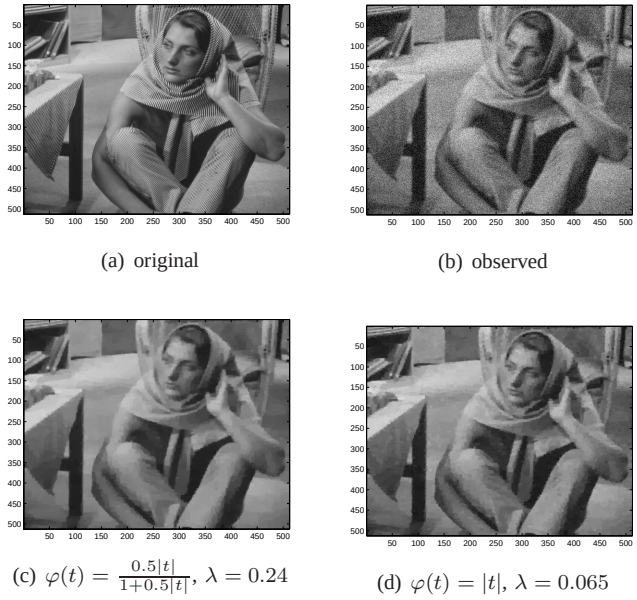
(c) $\varphi(t) = \frac{0.5|t|}{1+0.5|t|}$, $\lambda = 0.12$ (d) $\varphi(t) = |t|$, $\lambda = 0.05$

Fig. 11: The Cameraman image of size 256×256. Image restoration results are shown in (c) and (d) using SPG. The PSNRs are (b) PSNR=20.12; (c) PSNR= 21.48dB; and (d) PSNR=21.46dB.

F. Test of Barbara image

We test our proposed nonsmooth nonconvex regularization model on the Barbara image of size 512×512. This image has gradual changed pixel values, and is very different from piecewise constant image.

The observed image is constructed by adding a Gaussian noise with standard deviation 0.1 to the blurred image form the original image. We see from Fig. 12 (c)-(d) that the piecewise constant regions appear in the restoration results. We find that the restoration result using the nonsmooth nonconvex potential function $\varphi(t) = \frac{0.5|t|}{1+0.5|t|}$ provides higher PSNR than that using $\varphi(t) = |t|$.



(a) original (b) observed

(c) $\varphi(t) = \frac{0.5|t|}{1+0.5|t|}$, $\lambda = 0.24$ (d) $\varphi(t) = |t|$, $\lambda = 0.065$

Fig. 12: The Barbara image of size 512×512. Image restoration results are shown in (c) and (d) using SPG. The PSNRs are (b) PSNR= 18.58dB; (c) PSNR=23.40dB; and (d) PSNR=23.37.

G. Test of Books image

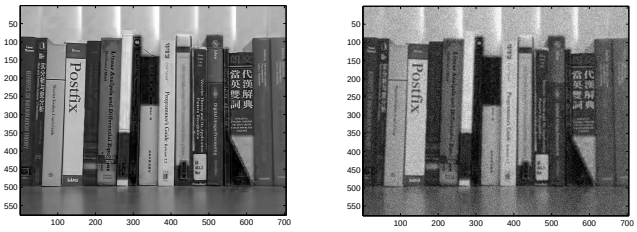
The original Books image of size 576×704 is obtained from website http://www.math.cuhk.edu.hk/ rchan/paper/csx, which looks like piecewise constant, but has some gradual shading due to illumination.

The standard deviation of a Gaussian noise added to the blurred image is 0.1 in this experiment. We try $p = 0.8$ and $p = 1$ in the regularization model and obtain the restoration results in Fig. 13 (c)-(d), respectively. Both the restored images improve the PSNR from the observed image a lot. And the restored image using $p = 0.8$ provides higher PSNR than that using $p = 1$.

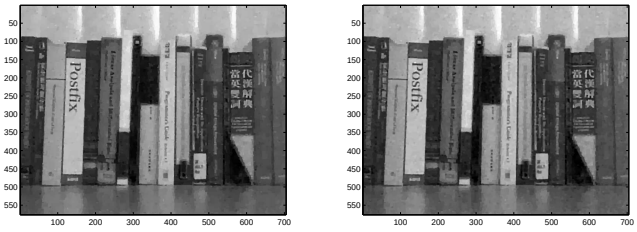
It is usually hard to find the global solutions for nonconvex model. There is no guarantee that the SPG method can find the global solutions for the proposed nonsmooth nonconvex regularization model. From numerical experiments in subsections E-G, by comparing with the restoration results using the convex TV regularization, we may say that the SPG method performs stable to avoid being trapped in bad local solutions for the nonsmooth nonconvex model.

V. Concluding Remarks

This paper studies a new box constrained minimization model with nonsmooth, concave regularization (4) for imaging restoration. We derive an easily computable constant θ for characterizing the sparsity of all local minimizers of (4). We show that every local minimizer f^* of (4) satisfies either $d_i^T f^* = 0$ or $|d_i^T f^*| \geq \min(\kappa, \theta)$. Moreover, we give an explicit form of θ for the box constrained image restoration



(a) original (b) observed



(c) $\varphi(t) = |t|^{0.8}$, $\lambda = 0.044$ (d) $\varphi(t) = |t|$, $\lambda = 0.05$

Fig. 13: The Books image of size 576×704. Image restoration results are shown in (c) and (d) using SPG. The PSNRs are (b) PSNR= 18.69dB; (c) PSNR=23.83dB; and (d) PSNR=23.80dB.

model with the non-Lipschitz nonconvex ℓ_p-norm ($0 < p < 1$) potential function in the regularization. Our numerical experiments validate the important characterization at local minimizers of (4) with ℓ_p-norm regularization. Moreover, the smoothing projected gradient method is shown very efficient to solve our proposed model. Our theoretical results, the box constrained minimization model with nonsmooth, non-Lipschitz regularization, and the SPG method contribute to the study of image processing.

APPENDIX

In order to show Theorems 1 and 2, we will use several index sets to prove these two theorems. For any $f \in \mathcal{F}$, define

- inactive and active sets of f

$$I = \{i \in \{1, \ldots, m\} \mid 0 < f_i < \kappa\}, \ \bar{I} = \{1, \ldots, m\} \setminus I \quad (12)$$

- index sets of zero and nonzero first-order differences

$$L = \{i \in \{1, \ldots, r\} \mid d_i^T f = 0\}, \ \bar{L} = \{1, \ldots, r\} \setminus L \quad (13)$$

- subset of L and subset of $\bar{L}$ in regard to the inactive set.

$$L_0 = \{i \in L \mid d_{iI} \neq 0\}, \ \hat{L}_0 = \{i \in \bar{L} \mid 0 < |d_i^T f| < \kappa\}. \quad (14)$$

The following simple example is used to explain these index sets.

Example 1: Let us consider an image of 3×3 pixels. The first-order difference matrix can be illustrated by the directed graph drawn in Fig. 14, where each vertex represents a pixel and each directed edge corresponds to a difference operator between two neighboring pixels.

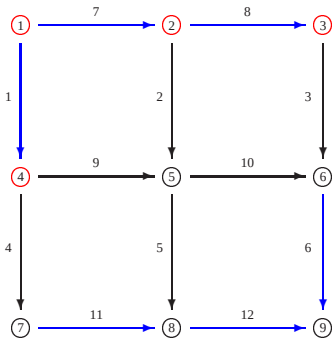


Fig. 14: illustration of the first-order difference matrix of 3×3 pixels

The first-order difference matrix $D \in R^{12 \times 9}$ corresponding to the above figure has the following form

$$D = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}.$$

Let $\kappa = 1$ and $f = (0.5, 0.5, 0.5, 0.5, 1, 0, 0, 0, 0)^T$. By direct computation, we get

$$Df = (0, -0.5, 0.5, 0.5, 1, 0, 0, 0, -0.5, 1, 0, 0)^T.$$

Then we obtain the index sets with respect to f as

$$I = \{1, 2, 3, 4\}, \qquad \bar{I} = \{5, 6, 7, 8, 9\},$$
$$L = \{1, 6, 7, 8, 11, 12\}, \qquad \bar{L} = \{2, 3, 4, 5, 9, 10\},$$
$$L_0 = \{1, 7, 8\}, \qquad \hat{L}_0 = \{2, 3, 4, 9\}.$$

If $I = \emptyset$ or $\hat{L}_0 = \emptyset$, then either $d_i^T f = 0$ or $|d_i^T f| = \kappa$ for $i = 1, \ldots, r$. Hence in the rest of this section, we assume $I \neq \emptyset$ and $\hat{L}_0 \neq \emptyset$,

Lemma 1: For a given vector $f \in \mathcal{F}$, let $I, \bar{I}, L, \bar{L}, L_0, \hat{L}_0$ be the index sets defined by (12)-(14). Then the following statements hold.
(i) either $d_{iI} = 0$ or $d_{i\bar{I}} = 0$ for any $i \in L$;
(ii) $d_{i\bar{I}} = 0$ for any $i \in L_0$;
(iii) $d_{iI} = 0$ for any $i \in L \setminus L_0$;
(iv) $d_{iI} \neq 0$ for any $i \in \hat{L}_0$.

Proof: Recall the notation introduced at the end of Section I, $d_{iI} \in R^{|I|}$, $d_{i\bar{I}} \in R^{|\bar{I}|}$ are the subvectors of d_i with $(d_{iI})_j, j \in I$ and $(d_{i\bar{I}})_j, j \in \bar{I}$, respectively. Note that each d_i contains only two nonzero entries, 1 and -1, and

$$d_i^T f = d_{iI}^T f_I + d_{i\bar{I}}^T f_{\bar{I}},$$

where $f_I = (f_i)_{i\in I} \in R^{|I|}$ and $f_{\bar{I}} = (f_i)_{i\in \bar{I}} \in R^{|\bar{I}|}$.

(i) If on the contrary for some $i \in L$, each of d_{iI} and $d_{i\bar{I}}$ contains one nonzero entry, then $0 < |d_{iI}^T f_I| < \kappa$, and $|d_{i\bar{I}}^T f_{\bar{I}}| \in \{0, \kappa\}$, which implies $d_i^T f \neq 0$. This contradicts to $i \in L$.

(ii) Since $d_{iI} \neq 0$ for any $i \in L_0 \subseteq L$, we immediately get from (i) that $d_{i\bar{I}} = 0$ for any $i \in L_0$.

(iii) This is directly from the definition of L_0.

(iv) Suppose on the contrary that $d_{iI} = 0$ for some $i \in \hat{L}_0$. Then $|d_i^T f| = |d_{i\bar{I}}^T f_{\bar{I}}| \in \{0, \kappa\}$, which contradicts to $i \in \hat{L}_0$. This completes the proof. ■

Proof of Theorem 1: Let $f^* \in \mathcal{F}^*$ be a local minimizer, and I, $\bar{I}$, L, $\bar{L}$, L_0, $\hat{L}_0$ be the index sets with respect to f^*. Recall that h_k is the kth column of H. Let $B = (h_k)_{k\in I}$ and $\bar{B} = (h_k)_{k\in \bar{I}}$ be the submatrices of H, whose entries lie in the columns of H indexed by I and $\bar{I}$, respectively. Denote

$$b = g - \bar{B} f_{\bar{I}}^*, \quad \text{and} \quad a_i = d_{i\bar{I}}^T f_{\bar{I}}^* \quad \text{for } i = 1, 2, \ldots, r.$$

It is easy to verify that

$$a_i = 0 \quad \text{for all} \quad i \in L, \tag{15}$$

since if $i \in L_0$, then $a_i = d_{iI}^T f_I^* = 0$ by (ii) of Lemma 1; and if $i \in L \setminus L_0$, then

$$a_i = d_{iI}^T f_I^* = d_{iI}^T f_I^* + d_{i\bar{I}}^T f_{\bar{I}}^* = d_i^T f^* = 0,$$

by employing (iii) of Lemma 1.

We then consider the following constrained optimization problem

$$\begin{array}{ll} \min\limits_{w\in R^{|I|}} & r(w) := \|Bw - b\|^2 + \lambda \sum\limits_{i\in\bar{L}} \varphi(d_{iI}^T w + a_i) \\ \text{s.t.} & d_{iI}^T w = 0, \quad \forall i \in L_0. \end{array} \tag{16}$$

Let us denote $w^* = f_I^*$ and $S_I = \{w \mid 0 \leq w \leq \kappa e_I\}$. Since $f^* \in \mathcal{F}^*$, there exists a neighborhood $\mathcal{N}(f^*) := \{f \mid \|f - f^*\| \leq \delta\}$ of f^* such that $\mathcal{N}(f^*) \cap \mathcal{F} \neq \emptyset$ and

$$r(w^*) = z(f^*) = \min\{z(f) \mid f \in \mathcal{N}(f^*) \cap \mathcal{F}\}. \tag{17}$$

Denote $\mathcal{N}(w^*) := \{w \mid \|w - w^*\| \leq \delta\}$ for a neighborhood of w^*. We now claim that w^* is a global minimizer of $r(w)$ in the region

$$\Omega_1 := \{w \mid d_{iI}^T w = 0 \quad \text{for all } i \in L_0, \quad w \in \mathcal{N}(w^*) \cap S_I\}.$$

Otherwise there exists $\hat{w} \in \Omega_1$ such that $r(\hat{w}) < r(w^*)$. Define $\hat{f} \in \mathcal{N}(f^*) \cap \mathcal{F}$ by letting $\hat{f}_I = \hat{w}$ and $\hat{f}_{\bar{I}} = f_{\bar{I}}^*$. We have

$$\begin{aligned} z(\hat{f}) &= r(\hat{w}) + \lambda \sum_{i\in L} \varphi(d_{iI}^T \hat{w} + a_i) \\ &= r(\hat{w}) + \lambda \sum_{i\in L} \varphi(d_{iI}^T \hat{w}) \\ &= r(\hat{w}) + \lambda \sum_{i\in L_0} \varphi(d_{iI}^T \hat{w}) + \lambda \sum_{i\in L\setminus L_0} \varphi(d_{iI}^T \hat{w}) \\ &= r(\hat{w}), \end{aligned}$$

where the second equality comes from (15), and the last equality can be obtained easily by employing $\hat{w} \in \Omega_1$, Lemma 1 (iii), and Assumption I (a). Thus we find $\hat{f} \in \mathcal{N}(f^*) \cap \mathcal{F}$, and

$$z(\hat{f}) = r(\hat{w}) < r(w^*) = z(f^*),$$

which contradicts (17).

Hence w^* is a local minimizer of the equality constrained minimization problem (16), where the objective function $r : R^{|I|} \to R$ is twice continuously differentiable at w^*. By the second-order necessary condition for the local minimizer w^*,

$$v^T \nabla^2 r(w^*) v = 2\|Bv\|^2 + \lambda \sum_{i\in\bar{L}} \varphi''(d_i^T f^*)(d_{iI}^T v)^2 \geq 0, \quad \text{for all } v \in \mathcal{A}(w^*) \tag{18}$$

where

$$\mathcal{A}(w^*) = \{v \mid d_{iI}^T v = 0, \ \forall i \in L_0\}.$$

For any $i \in \hat{L}_0$, $d_{iI} \neq 0$ according to (iv) of Lemma 1. Moreover, we can deduce that $d_{iI}^T w^* \neq 0$ since $0 < |d_i^T f^*| < \kappa$ for any $i \in \hat{L}_0$. Let $v^*(i) \in R^{|I|}$ be a solution to the quadratic programming

$$\begin{array}{ll} \min & \|v\|^2 \\ \text{s.t.} & v \in \mathcal{A}(w^*) \\ & d_{iI}^T v = 1, \end{array} \tag{19}$$

where $i \in \hat{L}_0$. The existence of $v^*(i)$ is guaranteed by the Frank-Wolfe theorem, by noting that $\|v\|^2 \geq 0$ and the feasible set is a polyhedron which is nonempty since $\tilde{v} = \frac{w^*}{d_{iI}^T w^*}$ belongs to it. Define

$$\mu(f^*) = \max_{i\in\hat{L}_0} \{\|v^*(i)\|\}.$$

Note that the minimizer of (19) is in fact determined by the index sets I, $\bar{L}$, L_0, $\hat{L}_0$ with respect to f^*. Taking all possible index sets I, $\bar{L}$, L_0, $\hat{L}_0$ with respect to $f^* \in \mathcal{F}^*$, which are finite, we can define

$$\mu = \max_{f^*\in\mathcal{F}^*} \mu(f^*).$$

Set $\lambda_0 = \frac{2\mu^2\|H^T H\|}{|\varphi''(0^+)|}$. Since $\lambda > \lambda_0$, we have $\frac{2\mu^2\|H^T H\|}{\lambda} < |\varphi''(0^+)|$. Then we can define the finite constant θ given by

$$\theta = \inf\Big\{t > 0 \mid \varphi''(t) = -\frac{2\mu^2\|H^T H\|}{\lambda}\Big\}.$$

We now prove that for λ_0 and θ defined above, statement (8) holds for the given local minimizer f^*. Suppose on the contrary that there is $j \in \hat{L}_0$ such that

$$0 < |d_j^T f^*| < \theta.$$

Consequently, we know that $\varphi''(d_j^T f^*) < \varphi''(\theta)$ from Assumption I (c), and

$$\begin{aligned} & v^*(j) \nabla^2 r(w^*) v^*(j) \\ &= 2\|Bv^*(j)\|^2 + \lambda \sum_{i\in\bar{L}} \varphi''(d_i^T f^*)(d_{iI}^T v^*(j))^2 \\ &\leq 2\mu^2\|H^T H\| + \lambda\varphi''(d_j^T f^*) \\ &< 2\mu^2\|H^T H\| + \lambda\varphi''(\theta) = 0, \end{aligned}$$

which contradicts (18). The fact that μ, and consequently λ_0 and θ, do not vary with any $f^* \in \mathcal{F}^*$ will yield that (8) holds for any local minimizer $f^* \in \mathcal{F}^*$. This completes the proof. ■

Proof of Theorem 2: The proof of this theorem is based on Theorem 1 and its proof. The difference between these two theorems is that instead of using a solution to (19), this theorem uses a feasible point of $\mathcal{A}(w^*)$ to give computable constants λ_0 and θ.

By Assumption I (c), the constant θ in (9) is well-defined and finite, since

$$|\varphi''(0^+)| = \frac{2\|H^T H\|m}{\lambda_0} > \frac{2\|H^T H\|m}{\lambda}.$$

We only need to show $|d_\ell^T f^*| \geq \theta$ for any $\ell \in \hat{L}_0$. We will fulfill this by analyzing two possible cases.

Case 1. $L_0 = \emptyset$. In this case $\mathcal{A}(w^*) = R^{|I|}$. For a fixed $\ell \in \hat{L}_0$, $d_{\ell I} \neq 0$ according to (iv) of Lemma 1. Assume $D_{\ell k} \neq 0$ for some $k \in I$. Let us define $\hat{v} \in R^{|I|}$ such that $\hat{v}_k = 1$, and $\hat{v}_i = 0$ for any $i \neq k$, $i \in I$. It is easy to check that $(d_{\ell I}^T v)^2 = 1$ and $\|\hat{v}\| = 1$. From (18) we have

$$0 \leq 2\|B\hat{v}\|^2 + \lambda\varphi''(d_\ell^T f^*)(d_{\ell I}^T v)^2 \leq 2\|H^T H\| + \lambda\varphi''(d_\ell^T f^*),$$

which, combined with Assumption 1 (c), indicates

$$|d_\ell^T f^*| \geq \inf\{t > 0 \mid \varphi''(t) = -\frac{2\|H^T H\|}{\lambda}\} \geq \theta.$$

Case 2. $L_0 \neq \emptyset$. Let us consider the homogeneous system

$$D_{L_0 I} v = 0, \tag{20}$$

where $D_{L_0 I} \in R^{|L_0| \times |I|}$ is the submatrix of D whose rows are d_{kI}^T, $k \in L_0$. It is easy to see that $\mathcal{A}(w^*)$ coincides to the null space of $D_{L_0 I}$, i.e., $\mathcal{A}(w^*) = \text{null}(D_{L_0 I})$. Denote ν for the rank of $D_{L_0 I}$. Since $0 \neq w^* \in \mathcal{A}(w^*)$, we have

$$\nu = \text{rank}(D_{L_0 I}) \leq \min\{|L_0|, |I| - 1\}.$$

According to (i) and (ii) of Lemma 1, each row d_{kI}^T, $k \in L_0$, has exactly two nonzero entries, 1 and -1. By performing the elementary row operations and rearranging the columns if necessary, we can get the equivalent system of (20) as follows

$$v_\beta - N v_\alpha = 0. \tag{21}$$

Here $v_\beta \in R^\nu$ and $v_\alpha \in R^{|I|-\nu}$ are the basic and nonbasic variables respectively. Entries of the matrix N are either 0 or 1, and each row of N has a single nonzero entry 1.

Let $v_\alpha(k)$ be the kth column of the identify matrix in $R^{(|I|-\nu)\times(|I|-\nu)}$, and $v_\beta(k) = N v_\alpha(k)$. Then we find $|I| - \nu$ solutions $v(k) \in R^{|I|}$ to the linear system (20), which form the basis of $\mathcal{A}(w^*)$. For a fixed $\ell \in \hat{L}_0$, we set

$$\hat{v}(\ell) = \text{argmax}\{(d_{\ell I}^T v(k))^2,\ k = 1, \ldots, |I| - \nu\}.$$

Note that $w^* \in$span$\{v(k), k = 1, \ldots, |I| - \nu\}$ and $d_{\ell I}^T w^* \neq 0$. We can claim that $d_{\ell I}^T v(k) \neq 0$ for some $v(k)$. Since $d_{\ell I}$ has only two nonzero elements, 1 and -1, and elements of $v(k)$ are either 0 or 1, we find that $(d_{\ell I}^T \hat{v}(\ell))^2 = 1$ and $\|\hat{v}(\ell)\|^2 \leq |I| \leq m$. From (18) we have

$$\begin{aligned} 0 &\leq 2\|B\hat{v}(\ell)\|^2 + \lambda\varphi''(d_\ell^T f^*)(d_{\ell I}^T \hat{v}(\ell))^2 \\ &\leq 2|I|\ \|H^T H\| + \lambda\varphi''(d_\ell^T f^*) \\ &\leq 2m\|H^T H\| + \lambda\varphi''(d_\ell^T f^*), \end{aligned}$$

which, combined with Assumption 1 (c), implies

$$|d_\ell^T f^*| \geq \inf\{t > 0 \mid \varphi''(t) = -\frac{2\|H^T H\|m}{\lambda}\} = \theta.$$

We complete the proof. ■

Proof of Theorem 3: First we consider

$$f_k^* \leq \min\{f_i^*, i \in \mathcal{C}_k\}. \tag{22}$$

If $f_k^* = \kappa$, then from (22) we have $f_i^* - f_k^* = 0$ for all $i \in \mathcal{C}_k$. Suppose $0 \leq f_k^* < \kappa$ and $f_i^* - f_k^* > 0$ for all $i \in \mathcal{C}_k$, which implies $|d_j^T f^*| > 0$ for all $j \in \mathcal{J}_k$. Let

$$\psi(t) = z(f^* + te_k) = \|H(f^* + te_k) - g\|^2 + \lambda\|D(f^* + te_k)\|_p^p, \tag{23}$$

where e_k is the kth column of the $m \times m$ identity matrix. We consider the following constrained minimization problem

$$\min_{t \geq 0} \psi(t). \tag{24}$$

Since f^* is a local minimizer of (5) satisfying $0 \leq f_k^* < \kappa$, we know that $t^* = 0$ is a local minimizer of (24). Moreover, we deduce that ψ is differentiable at $t^* = 0$ from the facts

$$\sum_{j=1}^r |d_j^T(f^* + te_k)|^p = \sum_{j \in \mathcal{J}_k} |d_j^T(f^* + te_k)|^p + \sum_{j \notin \mathcal{J}_k} |d_j^T f^*|^p$$

and

$$|d_j^T f^*| > 0, \quad \text{for all } j \in \mathcal{J}_k.$$

Hence the first order optimal condition of (24) holds at $t^* = 0$, that is,

$$2(Hf^* - g)^T h_k + \lambda p \sum_{j \in \mathcal{J}_k} |d_j^T f^*|^{p-1} \text{sign}(d_j^T f^*) D_{jk} \geq 0.$$

We have by (22) that

$$\text{sign}(d_j^T f^*) D_{jk} = -1, \quad \text{for all } j \in \mathcal{J}_k.$$

Therefore,

$$2(Hf^* - g)^T h_k \geq \lambda p \sum_{j \in \mathcal{J}_k} |d_j^T f^*|^{p-1} \geq \lambda p |d_j^T f^*|^{p-1},$$
$$\text{for all } j \in \mathcal{J}_k.$$

By the assumption $z(f^*) \leq z(f^0)$, we obtain

$$|(Hf^* - g)^T h_k|^2 \leq \|h_k\|^2 \|Hf^* - g\|^2 \leq$$
$$\|h_k\|^2(\|Hf^* - g\|^2 + \lambda \sum_{j=1}^r |d_j^T f^*|^p) \leq \|h_k\|^2 z(f^0).$$

This implies

$$|d_j^T f^*|^{1-p} \geq \frac{\lambda p}{2|(Hf^* - g)^T h_k|} \geq \frac{\lambda p}{2\|h_k\|\sqrt{z(f^0)}},$$
$$\text{for all } j \in \mathcal{J}_k.$$

From the definition of $\mathcal{C}_k$, for any $i \in \mathcal{C}_k$ there is $j \in \mathcal{J}_k$ such that $|f_i^* - f_k^*| = |d_j^T f^*|$. Hence we obtain

$$|f_i^* - f_k^*| \geq \left(\frac{\lambda p}{2\|h_k\| \sqrt{z(f^0)}} \right)^{\frac{1}{1-p}}, \quad \text{for all } i \in \mathcal{C}_k.$$

The case

$$f_k^* \geq \max\{f_i^*, i \in \mathcal{C}_k\} \tag{25}$$

can be proved similarly by employing the constrained minimization problem

$$\min_{t \leq 0} \psi(t), \tag{26}$$

and the fact that in this case

$$\text{sign}(d_j^T f^*) D_{jk} = 1, \quad \text{for all } j \in \mathcal{J}_k.$$

■

Proof of Theorem 4: Suppose

$$0 < f_k^* < \kappa, \quad \text{and} \quad |d_j^T f^*| > 0 \quad \text{for all } j \in \mathcal{J}_k.$$

Since f^* is a local minimizer of (5), we know that $t^* = 0$ is a local minimizer of the unconstrained minimization problem

$$\min \ \psi(t), \tag{27}$$

where $\psi(t)$ is given in (23). Moreover, we deduce that ψ is differentiable at $t^* = 0$ by noting

$$\sum_{j=1}^{r} |d_j^T (f^* + te_k)|^p = \sum_{j \in \mathcal{J}_k} |d_j^T (f^* + te_k)|^p + \sum_{j \notin \mathcal{J}_k} |d_j^T f^*|^p$$

and

$$|d_j^T f^*| > 0, \quad \text{for all } j \in \mathcal{J}_k.$$

The second order optimal condition of (27) yields

$$2h_k^T h_k + \lambda p(p-1) \sum_{j \in \mathcal{J}_k} |d_j^T f^*|^{p-2} \geq 0.$$

Since $0 < p < 1$, we find

$$2h_k^T h_k \geq \lambda p(1-p) \sum_{j \in \mathcal{J}_k} |d_j^T f^*|^{p-2} \geq \lambda p(1-p) |d_j^T f^*|^{p-2},$$
$$\text{for all } j \in \mathcal{J}_k.$$

This implies

$$|d_j^T f^*| \geq \left(\frac{\lambda p(1-p)}{2h_k^T h_k} \right)^{\frac{1}{2-p}} = \beta_k, \quad \text{for all } j \in \mathcal{J}_k. \tag{28}$$

From the definition of $\mathcal{C}_k$, for any $i \in \mathcal{C}_k$ there is $j \in \mathcal{J}_k$ such that $|f_i^* - f_k^*| = |d_j^T f^*|$. Hence we obtain $|f_i^* - f_k^*| \geq \beta_k$ for all $i \in \mathcal{C}_k$.

Now let us consider the case f_k^* reaches the boundary, i.e., $f_k^* = 0$ or $f_k^* = \kappa$. It is clear that

$$f_k^* = 0 \ \Rightarrow \ f_k^* \leq \min\{f_i^*, \ i \in \mathcal{C}_k\} \quad \text{and} \quad f_k^* = \kappa \ \Rightarrow \ f_k^* \geq \max\{f_i^*, \ i \in \mathcal{C}_k\}.$$

From Lemma 3, we obtain immediately that

$$|f_k^* - f_i^*| \geq \alpha_k, \quad \text{if} \quad f_k^* = 0, \ \text{or } f_k^* = \kappa.$$

We complete the proof. ■

ACKNOWLEDGMENT

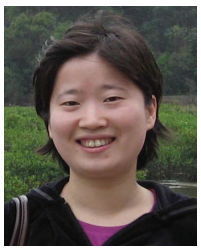
The authors would like to thank Dr. Juan Wei (Philips Research, Asia) for providing the MRI images. They are also very grateful to the two anonymous referees for valuable comments.

REFERENCES

[1] 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. 965–994, 1999.
[2] G. Golub, P. Hansen, and D. O'Leary, "Tikhonov regularization and total least squares," *SIAM J. Matrix Anal.*, vol. 21, pp. 185–194, 1999.
[3] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithm," *Physica*, vol. 60, pp. 259–268, 1992.
[4] N. Paragios, C. Chen, and O. Faugeras, *Handbook of Mathematical Models in Computer Vision*. Springer, 2006.
[5] X. Chen and W. Zhou, "Smoothing nonlinear conjugate gradient method for image restoration using nonsmooth nonconvex minimization," *SIAM J. Imaging Sci.*, pp. 765–790, 2010.
[6] M. Nikolova, "Minimizers of cost-functions involving non-smooth data-fidelity terms. application to the processing of outliers," *SIAM J. Numer. Anal.*, vol. 40, pp. 965–994, 2002.
[7] ——, "A variational approach to remove outliers and impulse noise," *J. Math. Imaging Vision*, vol. 20, pp. 99–120, 2004.
[8] M. Nikolova, M. K. Ng, S. Zhang, and W. Ching, "Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization," *SIAM J. Imaging Sci.*, vol. 1, pp. 2–25, 2008.
[9] M. Nikolova, "Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares," *Multisc. Model. Simul.*, vol. 4, pp. 960–991, 2005.
[10] M. Nikolova, M. K. Ng, and C. P. Tam, "Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction," *IEEE Tran. Image Processing*, vol. 19, pp. 3073–3088, 2010.
[11] R. Gonzalez and R. Woods, *Digital Image Processing*. Pearsol Int. Edition, 3rd ed., 2008.
[12] A. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[13] M. R. Banham and A. K. Katsaggelos, "Digital image restoration," *IEEE Signal Processing Magazine*, vol. 14, pp. 24–41, 1997.
[14] M. Hanke, J. Nagy, and C. Vogel, "Quasi-Newton approach to non-negative image restorations," *Linear Alg. Appl.*, vol. 316, pp. 223–236, 2000.
[15] J. Nagy and Z. Strakos, "Enforcing nonnegativity in image reconstruction algorithms," *Mathematical Modeling, Estimation, and Imaging, David C. Wilson, et.al., Eds.*, vol. 4121, pp. 182–190, 2000.
[16] M. Hong, T. Stathaki, and A. Katsaggelos, "Iterative regularized image restoration using local constraints," in *Proc. IEEE Workshop on Nonlinear Signal and Image Processing*, September 1997.
[17] D. Krishnan, L. Ping, and A. M. Yip, "A primal-dual active-set method for non-negativity constrained total variation deblurring problems," *IEEE Trans. Image Processing*, vol. 16, pp. 2766–2777, 2007.
[18] D. Geman, *Random fields and inverse problems in imaging*, ser. Lecture Notes in Mathematics. Berlin, Germany: Springer-Verlag, 1990.
[19] D. Geman and G. Reynolds, "Constrained restoration and recovery of discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 367–383, 1992.
[20] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Processing*, vol. 4, pp. 932–946, 1995.
[21] C. Zhang and X. Chen, "Smoothing projected gradient method and its application to stochastic linear complementarity problems," *SIAM J. Optim.*, vol. 20, pp. 627–649, 2009.
[22] X. Chen, F. Xu, and Y. Ye, "Lower bound theory of nonzero entries in solutions of ℓ_2-ℓ_p minimization," *SIAM J. Sci. Comput.*, vol. 32, pp. 2832–2852, 2010.

Xiaojun Chen is a Professor at the Department of Applied Mathematics, The Hong Kong Polytechnic University. Previously, she was a Professor at the Department of Mathematical Sciences, Hirosaki University, Japan. Her current research interests include nonsmooth, nonconvex optimization, stochastic variational inequalities and approximations on the sphere.

Michael K. Ng is a Professor in the Department of Mathematics at the Hong Kong Baptist University. He obtained his B.Sc. degree in 1990 and M.Phil. degree in 1992 at the University of Hong Kong, and Ph.D. degree in 1995 at Chinese University of Hong Kong. He was a Research Fellow of Computer Sciences Laboratory at Australian National University (1995-1997), and an Assistant/Associate Professor (1997-2005) of the University of Hong Kong before joining Hong Kong Baptist University. His research interests include bioinformatics, data mining, image processing, scientific computing and data mining, and he serves on the editorial boards of international journals, see http://www.math.hkbu.edu.hk/∼mng



Chao Zhang is an Associate Professor in the Department of Applied Mathematics at Beijing Jiaotong University. She received the B.S. degree in Mathematics from Normal School of Qingdao University, Qingdao, China in 2001, the M.S. degree from Beijng Jiaotong University, Beijing, China in 2004, and the Ph. D degree from Hirosaki University, Hirosaki, Japan in 2008. Her research interests include algorithms for nonsmooth nonconvex optimization and applications in data and image processing.