SIAM J. IMAGING SCIENCES
Vol. 1, No. 1, pp. 2–25





# Efficient Reconstruction of Piecewise Constant Images Using Nonsmooth Nonconvex Minimization*

Mila Nikolova†, Michael K. Ng‡, Shuqin Zhang§, and Wai-Ki Ching¶

**Abstract.** We consider the restoration of *piecewise constant images* where the number of the regions and their values are not fixed in advance, with a good difference of piecewise constant values between neighboring regions, from noisy data obtained at the output of a linear operator (e.g., a blurring kernel or a Radon transform). Thus we also address the generic problem of *unsupervised* segmentation in the context of linear inverse problems. The segmentation and the restoration tasks are solved jointly by minimizing an objective function (an energy) composed of a quadratic data-fidelity term and a nonsmooth nonconvex regularization term. The pertinence of such an energy is ensured by the analytical properties of its minimizers. However, its practical interest used to be limited by the difficulty of the computational stage which requires a nonsmooth nonconvex minimization. Indeed, the existing methods are unsatisfactory since they (implicitly or explicitly) involve a smooth approximation of the regularization term and often get stuck in shallow local minima. The goal of this paper is to design a method that efficiently handles the nonsmooth nonconvex minimization. More precisely, we propose a continuation method where one tracks the minimizers along a sequence of approximate nonsmooth energies $\{J_\varepsilon\}$, the first of which being strictly convex and the last one the original energy to minimize. Knowing the importance of the nonsmoothness of the regularization term for the segmentation task, each $J_\varepsilon$ is nonsmooth and is expressed as the sum of an $\ell_1$ regularization term and a smooth nonconvex function. Furthermore, the local minimization of each $J_\varepsilon$ is reformulated as the minimization of a smooth function subject to a set of linear constraints. The latter problem is solved by the modified primal-dual interior point method, which guarantees the descent direction at each step. Experimental results are presented and show the effectiveness and the efficiency of the proposed method. Comparison with simulated annealing methods further shows the advantage of our method.

**Key words.** image restoration, regularization, nonsmooth and nonconvex optimization, continuation, interior point method, unsupervised segmentation, inverse problems, deblurring, constrained optimization, graduated nonconvexity (GNC)

**AMS subject classifications.** 65F22, 65F10, 65K05, 65K10

**DOI.** 10.1137/070692285

## 1. Introduction.

Digital image restoration and reconstruction plays an important part in various applied areas such as medical and astronomical imaging, film restoration, image

*Received by the editors May 17, 2007; accepted for publication (in revised form) October 31, 2007; published electronically March 20, 2008. This research was supported by the PROCORE - France/Hong Kong Joint Research Scheme sponsored by the Research Grant Council of Hong Kong and the Consulate General of France in Hong Kong F-HK 30/04T.

http://www.siam.org/journals/siims/1-1/69228.html

†Centre de Mathématiques et de Leurs Applications, ENS de Cachan, 61 av. President Wilson, 94235 Cachan Cedex, France (nikolova@cmla.ens-cachan.fr).

‡Corresponding author. Centre 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 was supported in part by Hong Kong Research Grants Council grants 7035/04P and 7035/05P and HKBU FRGs.

§School of Mathematical Sciences, Fudan University, Shanghai, 200433, China (zhangs@fudan.edu.cn).

¶Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong (wkc@maths.hku.hk).

and video coding, and many others [35, 13, 46, 2, 32]. We focus on the most common image degradation model where the observed data $\mathbf{g} \in \mathbb{R}^q$ are related to the underlying image $\mathbf{f} \in \mathbb{R}^p$ as

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

where $\mathbf{n}$ represents the noise 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, and a Fourier transform in diffraction tomography. Usually the information provided by the forward model (1.1) alone is not sufficient to find an acceptable solution $\hat{\mathbf{f}}$. Prior information on the underlying image can then help the restoration of $\hat{\mathbf{f}}$. A flexible means of combining the forward model with priors is regularization, which is formally equivalent to maximum a posteriori (MAP) estimation [7, 18, 22]. The solution $\hat{\mathbf{f}}$ is defined as the (global) minimizer of a cost function $J(\mathbf{f})$ of the form

$$J(\mathbf{f}) = \Theta(H\mathbf{f} - \mathbf{g}) + \beta\Phi(\mathbf{f}), \quad \mathbf{f} \geq 0, \tag{1.2}$$

where $\Theta$ forces closeness to data, $\Phi$ embodies the priors, and $\beta > 0$ is a parameter that controls the trade-off between these two terms. The constraint $\mathbf{f} \geq 0$ is optional, but in our context it reflects the fact that the pixels are nonnegative. The most usual choice for $\Theta$ is

$$\Theta(\mathbf{v}) = \|\mathbf{v}\|_2^2. \tag{1.3}$$

In a statistical setting, this $\Theta$ corresponds to assuming that $\mathbf{n}$ is white Gaussian noise. Our computational method is designed for $\Theta$ as given in (1.3), or for any smooth and convex function $\Theta$. 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 [39, 14, 27]. Our method is straightforward to extend to this situation, and we will indicate how this can be done.

The role of $\Phi$ in (1.2) is to push the solution $\hat{\mathbf{f}}$ to exhibit some a priori expected features, such as the presence of edges, smooth regions, and textures. Since [7] a useful class of regularization functions is of the form

$$\Phi(\mathbf{f}) = \sum_{i=1}^{r} \varphi(\mathbf{d}_i^T \mathbf{f}), \tag{1.4}$$

where $\varphi$ is called a potential function (PF) and $\mathbf{d}_i^T : \mathbb{R}^p \to \mathbb{R}$, for $i = 1, \ldots, r$, are linear operators. In order to fix the frame of the presentation, we refer to the $\mathbf{d}_i^T\mathbf{f}$'s as differences. Usually $\{\mathbf{d}_i^T\}$ yield the first-order differences between each pixel and its 4 or 8 adjacent neighbors. Various PFs $\varphi$ have been used in the literature; a complete review can be found, for instance, in [9]. An important requirement is that $\varphi$ allows the recovery of large differences $|\mathbf{d}_i^T\hat{\mathbf{f}}|$ at the locations of edges in the original image and smooth differences elsewhere. It is well known that this requirement cannot be met by $\varphi(t) = t^2$, which was originally used in [50]. Since the pioneering work of S. Geman and D. Geman [22], different nonconvex functions $\varphi$ have been considered either in a statistical or in a variational framework [1, 7, 23, 20, 21, 31, 34, 42]. In

order to avoid the numerical intricacies arising with nonconvex regularization, an important effort was done later to derive edge-preserving convex PFs [12, 15, 25, 29, 44, 48]. Nevertheless, nonconvex regularization offers better possibilities for restoring images with neat edges. This question was systematically analyzed in [38].

We focus on the restoration of *piecewise constant images* $\hat{\mathbf{f}}$ where the number of the regions and their values are not fixed in advance, with a good contrast between neighboring regions. Such an $\hat{\mathbf{f}}$ thus provides both an *unsupervised* segmentation and a piecewise constant approximation of the underlying $\mathbf{f}$. This is important in various situations to aid in the interpretation of restored images, for the automatic recognition of objects, and finding profiles in computed tomography and nondestructive evaluation. This objective—to combine segmentation and restoration—is challenging especially in the context of inverse problems (1.1) involving a general nondiagonal matrix $H$. For instance, the usual segmentation techniques based on curve representation cannot be applied because of $H$. Under the same reason, hierarchical Markov random fields cannot be used either. For a general $H$, two main approaches can be evoked. Using a variational framework, image restoration and segmentation are tackled by minimizing an objective function involving a possibly continuous-valued line process; let us cite [20, 21, 15, 26, 49, 3] among others. Usually there is no guarantee that the resultant restoration yields an unambiguous segmentation since the obtained contours are not necessarily closed. The other approach is to construct a two-step process where image restoration and image segmentation are realized in an alternating way, with each dependent on the result of the other [33]. The process has to start with an initial guess for the segmentation whose importance is critical for the success of the overall method. Clearly all these difficulties are shrugged off if the restored image is piecewise constant. The *unsupervised* obtention of a piecewise constant image as a solution to a linear (possibly ill-conditioned) inverse problem still remains an open problem in the low-level vision community.

**1.1. Choice of energy: Nonsmooth nonconvex regularization.** We solve the problem formulated above by using nonsmooth nonconvex functions $\varphi$. More precisely, we suppose that

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

By (b), $\varphi$ is nonsmooth at zero and increasing on $(0,+\infty)$. By (c), it is nonconvex. These basic requirements are satisfied by the main nonconvex nonsmooth functions $\varphi$ used in practice; see, e.g., (f3), (f4), and (f5) in Table 1. Our interest in this class of functions $\varphi$ is motivated by the following theoretical result (see [38, Theorem 3.3]).

Theorem 1.1. *There is a constant $\mu > 0$ entirely determined by $\{\mathbf{d}_i^T : 1 \leq i \leq r\}$ such that if*

$$\beta > \frac{\mu \, \|H^T H\|_2}{|\varphi''(0^+)|}, \tag{1.5}$$

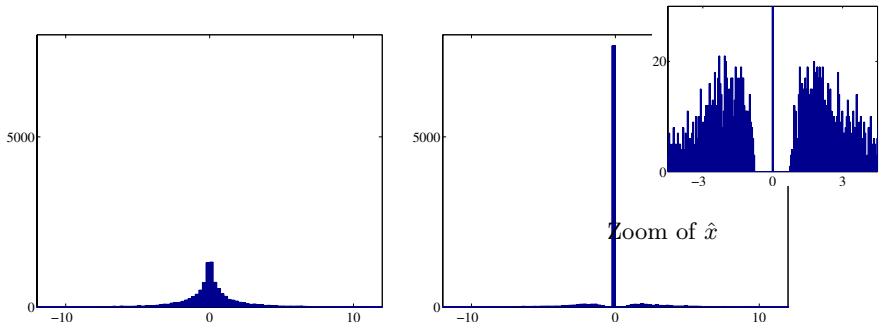
*then there exists $\theta > 0$ such that for every $\mathbf{g} \in \mathbb{R}^q$, every (local) minimizer $\hat{\mathbf{f}}$ of $J$ satisfies*

$$\textit{either} \quad |\mathbf{d}_i^T \hat{\mathbf{f}}| = 0 \quad \textit{or} \quad |\mathbf{d}_i^T \hat{\mathbf{f}}| \geq \theta \quad \forall i \in \{1, \ldots, r\}. \tag{1.6}$$

**Table 1**

*Nonsmooth at zero PFs $\varphi$ where $\alpha > 0$ is a parameter. Some references are* [7, 44, 20, 21, 45, 30].

| | |
|---|---|
| Convex PFs | |
| (f1) | $\varphi(t) = \lvert t\rvert$ |
| Nonconvex PFs | |
| (f2) | $\varphi(t) = \lvert t\rvert^{\alpha},\ 0 < \alpha < 1$ |
| (f3) | $\varphi(t) = \dfrac{\alpha\lvert t\rvert}{1+\alpha\lvert t\rvert}$ |
| (f4) | $\varphi(t) = \log(\alpha\lvert t\rvert + 1)$ |
| (f5) | $\varphi(0)=0,\ \varphi(t)=1$ if $t\neq 0$ |




(a) All underlying $\mathbf{f}$ (b) The relevant minimizers $\hat{\mathbf{f}}$

**Figure 1.** *Histograms for* 10 000 *independent trials:* $\mathbf{f}$ *and* $\hat{\mathbf{f}}$ *for* $\varphi(t) = \sqrt{\lvert t\rvert}$.

Let us recall that $\|H^T H\|_2$ equals the largest eigenvalue of $H^T H$. If $\varphi''(0^+) = \infty$, as is the case with (f3), the requirement in (1.5) is simplified to $\beta > 0$. The constant $\mu$ is defined in [38] and is generally difficult to calculate for images.

*Example* 1. In order to illustrate the theorem, we consider a simple example where the underlying $\mathbf{f}$ is a scalar, $\mathbf{d}_1 = 1$, $H = 1$, and $\varphi$ is of the form (f3) for $\alpha = 0.5$, i.e., $\varphi(t) = \sqrt{\lvert t\rvert}$. We realized 10 000 independent trials where an original $\mathbf{f} \in \mathbb{R}$ is sampled from $p(\mathbf{f}) \propto \exp(-\lambda\varphi(\mathbf{f}))$ 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 $\mathbf{f}$ is shown in Figure 1(a). After this, the solution $\hat{\mathbf{f}}$ is calculated by minimizing $J(\mathbf{f}) = (\mathbf{f} - \mathbf{g})^2 + \varphi(\sqrt{\lvert\mathbf{f}\rvert})$ for $\beta = 2\sigma^2\lambda$. For every $\mathbf{g} \neq 0$ the function $J$ has two local minimizers, $\hat{\mathbf{f}}_1 = 0$ and $\hat{\mathbf{f}}_2$ satisfying $\lvert\hat{\mathbf{f}}_2\rvert > \theta$ for $\theta \approx 0.47$, and the global minimizer $\hat{\mathbf{f}}$ is found by comparing $J(\hat{\mathbf{f}}_1)$ and $J(\hat{\mathbf{f}}_2)$. The empirical histogram of all solutions $\hat{\mathbf{f}}$, shown in Figure 1(b), illustrates the theorem well: we have $\hat{\mathbf{f}} = 0$ in 77% of the trials, while the smallest nonzero $\lvert\hat{\mathbf{f}}\rvert$ is $0.77 > \theta$.

If $\{\mathbf{d}_i\}$ yield the first-order differences between the adjacent neighbors in an image, then $\hat{\mathbf{f}}$ is composed of constant regions surrounded by closed contours and neat edges higher than $\theta$. The solution $\hat{\mathbf{f}}$ is hence a segmented image. Let us mention that energies $J$ with PFs $\varphi$ satisfying (a), (b), and (c) have already been used in [20, 21], but never with the goal to obtain piecewise constant, segmented images. At that time, the property of Theorem 1 was not known, and thus the minimization schemes were unable to produce these features. Indeed, it is shown in [40] that replacing $\varphi$ with a smooth approximation gives rise to a solution which is

nowhere constant. Stochastic minimization with a finite number of iterations cannot produce these features either. The minimization of $J$ is tricky since $J$ may have many local minima that are usually located at points where $J$ is nonsmooth. The existing methods for solving this nonsmooth nonconvex optimization problem are reviewed in section 2.

**1.2. Content of the paper.** In the following we provide efficient tools to minimize energies of the form (1.2)–(1.4), where $\varphi$ is nonconvex and nonsmooth according to assumptions (a), (b), and (c) of the previous subsection. In section 3 we design a specialized minimization method that uses the idea of continuation. It is based on local minimization along a family of approximations $\{J_\varepsilon\}_{\varepsilon\in[0,1]}$ of $J$ that are nonsmooth in order to favor the recovery of constant patches in each $\hat{\mathbf{f}}_\varepsilon$ and that tackle the local minima problem by starting with a convex $J_0$. Each $J_\varepsilon$ is decomposed into a nonsmooth total variational-type regularization term and a smooth nonconvex term. In section 4, the minimization of every $J_\varepsilon$ is equivalently expressed as the minimization of a smooth nonconvex function subject to a set of linear constraints. The latter problems are solved by the modified primal-dual interior point method, which guarantees the descent direction at each step (section 5). Experimental results, presented in section 6, show the effectiveness and the efficiency of the proposed method. Comparison with simulated annealing methods further show the advantage of our method (section 7).

## 2. Existing minimization methods.

The minimization of $J$ involves three major difficulties that drastically restrict the methods that can be envisaged. Because of the nonconvexity of $\varphi$, $J$ may exhibit a large number of local minima which are not global. Second, $J$ 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.

**2.1. Simulated annealing (SA) for global minimization.** In this subsection, we denote $p_T(\mathbf{f}|\mathbf{g}) = \exp[-J(\mathbf{f})/T]/Z_T$, where $T > 0$ is temperature and $Z_T$ is the partition constant. Stochastic algorithms are based on the observation that when $T \to 0$, $p_T$ converges to a Dirac distribution centered on the maximizers of $\mathbf{f} \to p_T(\mathbf{f}|\mathbf{g})$, which are also the global minimizers of $J$. Since [28], asymptotically convergent global minimization of nonconvex functions $J$ has been conducted using annealing schemes where images $\mathbf{f}^k$ are sampled from $p_{T_k}(\mathbf{f}|\mathbf{g})$ for a slowly decreasing sequence of temperatures $T_k$. For discrete-valued images, this was sometimes done using the Metropolis dynamics [43]. For real-valued images, $\mathbf{f}^k$ is obtained using the Gibbs sampler [22], where each pixel $\mathbf{f}_i^k$ is a realization of $p_{T_k}(\mathbf{f}_i|\mathbf{f}_j, j \neq i, \mathbf{g})$. The update of each $\mathbf{f}_i^k$ thus involves all other pixels $\mathbf{f}_j^k$ such that $\mathbf{d}_i^T\mathbf{d}_j \neq 0$ or $\mathbf{h}_i^T\mathbf{h}_j \neq 0$, where $\mathbf{h}_j$ stands for the $j$th column of $H$. If $H$ is the identity, this set is given by the neighbors determined by $\{\mathbf{d}_i^T\}$ (e.g., the 4 or 8 nearest neighbors). However, if $H$ is a blurring operator or any nonsparse operator such as a Radon or a Fourier transform, updating $\mathbf{f}_i^k$ involves a large number of pixels $\mathbf{f}_j^k$, or even all $\mathbf{f}_j^k$, $j \neq i$, and the computation is cumbersome.

In order to lighten this computational burden, two different forms of half-quadratic reformulation of $J$ were proposed in [20] and later in [21]. They rely on the minimization of a new objective $\mathcal{J}$ which is quadratic with respect to $\mathbf{f}$ and separable with respect to an auxiliary array $\mathbf{b} \in \mathbb{R}^r$. SA is then applied to $p_T(\mathbf{f}, \mathbf{b}|\mathbf{g}) = \exp[-\mathcal{J}(\mathbf{f}, \mathbf{b})/T]/Z_T$ in order to find the

minimizer $(\hat{\mathbf{f}}, \hat{\mathbf{b}})$ of $\mathcal{J}$. Sampling $\mathbf{f}$ from $p_T(\mathbf{f}|\mathbf{b}, \mathbf{g})$ is then fully parallel and can be efficiently done using the FFT if $H$ is block-circulant. Sampling $\mathbf{b}$ from $p_T(\mathbf{b}|\mathbf{f}, \mathbf{g})$ is tractable since each entry $\mathbf{b}_i^k$ is realized independently. Solving our original problem (1.1) naturally requires that $\min_{\mathbf{b}} \mathcal{J}(\mathbf{f}, \mathbf{b}) = J(\mathbf{f})$ for every $\mathbf{f}$. For the additive form [21], there is no equivalence unless $\varphi$ is smooth; otherwise a smooth approximation of $\varphi$ is introduced implicitly and null differences $\mathbf{d}_i^T \hat{\mathbf{f}}$ cannot be found. For the multiplicative form [21], one can see that null differences $\mathbf{d}_i^T \hat{\mathbf{f}}$ correspond to $\hat{\mathbf{b}}_i = \infty$, which is hard to approximate numerically.

**2.2. Deterministic continuation.** If a plausible initial guess for the sought image is available (e.g., the maximum likelihood estimate for simple images and low noise corruption), local minimization in its vicinity can be an easy means of providing local regularity. It should be emphasized that this local minimization technique must be adapted to handle the nonsmoothness of $J$. To the best of our knowledge, the iterated conditional mode [6, 7], which is to maximize deterministically $p_{T_k}(\mathbf{f}_i|\mathbf{f}_j, j \neq i, \mathbf{g})$, for all $i$ and for a unique fixed temperature, $T_k = 1$, has only been used along with smooth energies $J$. A family of methods called ARTUR [15, 17] are methods that locally minimize with respect to $(\mathbf{f}, \mathbf{b})$ in the half-quadratic form $\mathcal{J}$ mentioned in section 2.1. According to the comments given there, they are not properly adapted to nonsmooth energies $J$. In all cases, local minimization techniques are easily trapped in shallow local minima that provide unsatisfactory solutions.

According to [52], continuation ideas (namely, to track local minima along a family of approximated energies $J_\varepsilon$, $\varepsilon \in [0, 1]$) are a good deterministic alternative to deal with nonconvex energies $J$. The graduated nonconvexity (GNC) algorithm was proposed by Blake and Zisserman [10] in the context of computer vision where $H$ is the identity and $\Phi$ is the Mumford–Shah regularization whose discrete version corresponds to $\varphi$ as given in (f4). More information about GNC is given in section 3. Even though there is no guarantee for global convergence, extensive experiments have shown that for a finite number of iterations GNC leads to minimizers $\hat{\mathbf{f}}$ having a lower (hence better) energy than SA [11]. Further extensions of GNC were proposed in [4, 5, 36, 37]. Another continuation approach is mean-field annealing [8, 47], which applies when $H$ is the identity and $\Phi$ is smooth, and depends on the initialization. However, none of these methods properly addresses the nonsmoothness of $J$: since all approximations $J_\varepsilon$ are smooth, their minimizers $\hat{\mathbf{f}}^\varepsilon$ cannot contain null differences; hence $\mathbf{d}_i^T \hat{\mathbf{f}}^\varepsilon \neq 0$.

**3. Nonsmooth GNC.** For $\varepsilon \in [0, 1]$ we consider a continuous family of approximated objectives $J_\varepsilon$ such that $\lim_{\varepsilon \to 1} J_\varepsilon = J$. The idea behind GNC continuation, as pioneered in [10], is to replace the global minimization of $J$ by a family of local minimizations along $J_\varepsilon$, when $\varepsilon$ is increasing from 0 to 1, by starting with a strictly convex $J_0$. Consider a sequence

$$\varepsilon_0 = 0 < \varepsilon_1 < \cdots < \varepsilon_n = 1$$

(or possibly $\varepsilon_n \overset{<}{\approx} 1$ or $\lim_{n \to \infty} \varepsilon_n = 1$). The GNC minimization starts by calculating the unique minimizer $\mathbf{f}^0$ of $J_0$. Afterwards, for each $k \geq 1$, an intermediate solution $\hat{\mathbf{f}}^k$ is calculated by minimizing $J_{\varepsilon_k}$ locally, starting from the previously obtained $\mathbf{f}^{k-1}$:

$$\hat{\mathbf{f}}^k = \arg \min_{\mathbf{f} \in V(\hat{\mathbf{f}}^{k-1})} J_{\varepsilon_k}(\mathbf{f}) \quad \text{for} \quad k = 1, \ldots, n.$$

In this expression, $V(\hat{\mathbf{f}}^{k-1})$ stands for the attraction valley of $J_{\varepsilon_k}$ near $\hat{\mathbf{f}}^{k-1}$. The natural way to construct $\{J_\varepsilon\}$ is to replace $\varphi$ in (1.2) by approximated potential functions $\varphi_\varepsilon$, $\varepsilon \in [0,1]$, such that $\varphi_0$ is convex and $\varphi_1 = \varphi$, and for every $\varepsilon \in [0,1]$, $\varphi_\varepsilon$ is symmetric and increasing on $[0,\infty]$. Thus we consider

$$J_\varepsilon(\mathbf{f}) = \Theta(H\mathbf{f} - \mathbf{g}) + \beta \sum_{i=1}^{r} \varphi_\varepsilon(\mathbf{d}_i^T \mathbf{f}). \tag{3.1}$$

When $\Theta$ is strictly convex, the initial $J_0$ is strictly convex. Such is the case for instance when $\Theta$ is of the form (1.3) with rank$H = p$. We note that when $\Theta$ is only nonstrictly convex (e.g. rank $H < p$), $J_0$ is only nonstrictly convex since $\mathbf{d}_i^T \mathbf{f} = 0$ for all constant images $\mathbf{f}$ (i.e., $\mathbf{f}_i = \mathbf{f}_j$ for all $i, j$). In the latter case, we can consider

$$J_\varepsilon(\mathbf{f}) = \Theta(H\mathbf{f} - \mathbf{g}) + \mu_\varepsilon \|\mathbf{f}\|^2 + \beta \sum_{i=1}^{r} \varphi_\varepsilon(\mathbf{d}_i^T \mathbf{f}),$$

where $\mu_\varepsilon$ decreases from $\mu_0 > 0$ to $\mu_\varepsilon = 0$ for all $\varepsilon \geq \varepsilon_1$ with $0 < \varepsilon_1 < 1$. Clearly, $J_0$ is then strictly convex. In order to simplify the presentation, we will consider $J_\varepsilon$ as given in (3.1).

The choice of the approximating family $\varphi_\varepsilon$ is crucial for the success of the GNC minimization. One of the novelties of our approach is to require that the intermediate solutions $\hat{\mathbf{f}}^k$ exhibit the same type of features as the global minimizer $\hat{\mathbf{f}}$ of $J$. In particular, we require that they are piecewise constant. Such a goal cannot be reached if $\varphi_\varepsilon$ are smooth functions [40]. Instead, we choose $\varphi_\varepsilon$ smooth on $(0,\infty)$ and such that

$$\varphi'_\varepsilon(0^+) > 0 \quad \forall \varepsilon \in [0,1]. \tag{3.2}$$

For instance, the function (f3) in Table 1 can be approximated by

$$\varphi_\varepsilon(t) = \frac{\alpha|t|}{1 + \varepsilon\alpha|t|}.$$

For (f8), we can take $\varphi_\varepsilon(t) = |t|^{\varepsilon\alpha+1-\varepsilon}$. A systematical way to construct $\varphi_\varepsilon$ when $\varphi'(0^+)$ is finite is simply

$$\varphi_\varepsilon(t) = \varepsilon\varphi(t) + \rho(1-\varepsilon)|t|, \tag{3.3}$$

where $\rho$ is any positive constant. In all these cases, every $\varphi_\varepsilon$ for $\varepsilon \in (0,1]$ satisfies the conditions (a), (b), and (c) in section 1.1 which ensures that all intermediate solutions $\hat{\mathbf{f}}^k$, $k \geq 1$, exhibit the property (1.6) in Theorem 1.1, whereas $\hat{\mathbf{f}}^0$ is likely to have constant pieces thanks to (3.2); see [40] for the theoretical explanation. We can argue that the precise shape of a $\varphi_\varepsilon$ satisfying (a), (b), and (c) has no significant influence on the ultimate solution as long as the differences $\mathbf{d}_i\mathbf{f}^k$ are either null or pushed away from zero, according to Theorem 1.1. This qualitative argument was confirmed by the numerical experiments.

There is some degree of freedom in the choice of the initial convex $\varphi_\varepsilon$—for instance, $\varphi_0$ in (3.3) is convex for any $\rho > 0$. When we start with $\rho = \varphi'(0^+)$, the regularization is strong, so the early solutions must have few details. As long as $\varepsilon$ increases, regularization decreases

and details start to appear. A better strategy is to start with a very small $\rho$ so that the early solutions are underregularized. When $\varepsilon$ increases, the regularization becomes stronger and spurious details are swept out. Our experiments show that this strategy leads to a deeper decrease of the energy although the difference with the case $\rho = \varphi'(0^+)$ is not sensitive.

**4. Equivalent formulation as smooth GNC under linear constraints.** At each step of our GNC method we have to properly solve a nonsmooth (local) minimization problem. Recently, several methods have specially been designed to minimize convex objectives $J$ with a convex but nonsmooth regularization term, such as the TV norm [16, 24, 19]. We will extend the method of [19] in order to handle nonconvex regularization under the additional condition that

(d) $\varphi$ is such that $\varphi'(0^+) > 0$ is finite.

To this end, we split each $J_\varepsilon$ into a (nonsmooth) total variational-type regularization term and a smooth, nonconvex term. More precisely, for every $\varepsilon \in [0, 1]$, we decompose

$$\varphi_\varepsilon(t) = \psi_\varepsilon(t) + \varphi'_\varepsilon(0^+)|t|. \tag{4.1}$$

Since $\psi_\varepsilon(t) = \varphi_\varepsilon(t) - \varphi'_\varepsilon(0^+)|t|$ satisfies

$$\psi'_\varepsilon(0^+) = 0 = \psi'_\varepsilon(0^-),$$

$\psi_\varepsilon$ is smooth at 0 and hence on $\mathbb{R}$. Notice also that $\psi_\varepsilon$ is symmetric and nonconvex if $\varepsilon > 0$. For instance, when $\phi$ is a function given by (f3), we have

$$\varphi_\varepsilon(t) = \frac{\alpha|t|}{1 + \varepsilon\alpha|t|} = \psi_\varepsilon(t) + \alpha|t|, \tag{4.2}$$

$$\psi_\varepsilon(t) = -\frac{\varepsilon\alpha^2 t^2}{1 + \varepsilon\alpha|t|} \tag{4.3}$$

with $\alpha = \varphi'_\varepsilon(0^+)$. For every $\varepsilon \in [0, 1]$, let us define

$$\alpha_\varepsilon = \varphi'_\varepsilon(0^+). \tag{4.4}$$

For $\varphi_\varepsilon$ as given in (3.3), $\alpha_\varepsilon = \varepsilon\varphi'(0^+) + \rho(1 - \varepsilon)$. Using (4.1), $J_\varepsilon$ can be put into the form

$$J_\varepsilon(\mathbf{f}) = \Theta(H\mathbf{f} - \mathbf{g}) + \beta\widetilde{\Psi}_\varepsilon(\mathbf{f}) + \beta\alpha_\varepsilon\|D\mathbf{f}\|_1, \tag{4.5}$$

where $D$ is the $r \times p$ matrix whose $i$th row is $\mathbf{d}_i^T$ for $i = 1, \ldots, r$ and

$$\widetilde{\Psi}_\varepsilon(\mathbf{f}) = \sum_{i=1}^{r} \psi_\varepsilon(\mathbf{d}_i^T\mathbf{f})$$

is a smooth function. Notice that whenever $\mathbf{d}_i\mathbf{f}$ are the first-order differences between the 4 adjacent neighbors, $\|D\mathbf{f}\|_1$ is the $\ell_1$ approximation of the discrete gradient of the image $\mathbf{f}$.

For any $\varepsilon \in [0, 1]$, we next focus on the local minimization of $J_\varepsilon(\mathbf{f})$ as given in (4.5). Let us introduce the following notation:

$$\mathbf{v} = D\mathbf{f} = \mathbf{v}^+ - \mathbf{v}^-, \quad \text{where} \begin{cases} \mathbf{v}^+ = \max(\mathbf{v}, 0), \\ \mathbf{v}^- = \max(-\mathbf{v}, 0), \end{cases} \tag{4.6}$$

as well as

$$\mathbf{y} = \begin{pmatrix} \mathbf{f} \\ \mathbf{v}^+ \\ \mathbf{v}^- \end{pmatrix}, \qquad \mathbf{c} = \begin{pmatrix} -2H^T\mathbf{g} \\ \beta\alpha_\varepsilon \mathbf{1} \\ \beta\alpha_\varepsilon \mathbf{1} \end{pmatrix},$$

$$M = \begin{pmatrix} 2H^TH & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad B = [D, -I, I],$$

$$\Psi_\varepsilon(\mathbf{y}) = \beta \sum_{1\le i\le r} \big(\psi_\varepsilon(\mathbf{v}_i^+) + \psi_\varepsilon(\mathbf{v}_i^-)\big) = \beta \sum_{i\ge p+1} \psi_\varepsilon(\mathbf{y}_i). \tag{4.7}$$

Observe that $\Psi_\varepsilon(\mathbf{y}) = \beta\widetilde{\Psi}_\varepsilon(\mathbf{f})$. The minimization of $J_\varepsilon$ in (4.5) is equivalent to finding $\mathbf{y} \in \mathbb{R}^{p+2r}$ that solves the problem

$$\begin{aligned} \min \quad & \frac{1}{2}\mathbf{y}^T M\mathbf{y} + \mathbf{c}^T\mathbf{y} + \Psi_\varepsilon(\mathbf{y}) \\ \text{subject to} \quad & B\mathbf{y} = 0, \quad \mathbf{y} \ge 0. \end{aligned} \tag{4.8}$$

Now the nonconvex nonsmooth local minimization problem is reformulated as the minimization of a nonconvex differentiable objective function under a set of linear constraints, which is much more tractable. Notice that any additional linear constraint on $\mathbf{f}$ is easy to incorporate in the above formulation.

*Remark* 1 (*$\ell_1$ data fidelity*). The case when in (1.2) we have $\Theta(\mathbf{f}) = \|H\mathbf{f} - \mathbf{g}\|_1$ can be considered in a very similar way. Instead of (4.5), the approximated $J_\varepsilon$ now reads

$$J_\varepsilon(\mathbf{f}) = \|H\mathbf{f} - \mathbf{g}\|_1 + \beta\widetilde{\Psi}_\varepsilon(\mathbf{f}) + \beta\alpha_\varepsilon\|D\mathbf{f}\|_1. \tag{4.9}$$

In addition to (4.6), we define

$$\boldsymbol{u} = H\mathbf{f} - \mathbf{g} \quad \text{and} \quad \begin{cases} \boldsymbol{u}^+ = \max(\boldsymbol{u}, 0), \\ \boldsymbol{u}^- = \max(-\boldsymbol{u}, 0), \end{cases}$$

$$\mathbf{y} = \begin{pmatrix} \mathbf{f} \\ \boldsymbol{u}^+ \\ \boldsymbol{u}^- \\ \mathbf{v}^+ \\ \mathbf{v}^- \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} \mathbf{0} \\ \mathbf{1} \\ \mathbf{1} \\ \beta\alpha_\varepsilon\mathbf{1} \\ \beta\alpha_\varepsilon\mathbf{1} \end{pmatrix}, \quad \boldsymbol{h} = \begin{pmatrix} \mathbf{g} \\ \mathbf{0} \end{pmatrix}, \quad B = \begin{bmatrix} H & -I & I & \mathbf{0} & \mathbf{0} \\ D & \mathbf{0} & \mathbf{0} & -I & I \end{bmatrix},$$

$$\Psi_\varepsilon(\mathbf{y}) = \beta \sum_{i\ge p+2q+1} \psi_\varepsilon(\mathbf{y}_i).$$

Finding a minimizer $\hat{\mathbf{f}}$ of $J_\varepsilon$ is equivalent to finding a solution $\hat{\mathbf{y}}$ of the following problem:

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{y} + \Psi_\varepsilon(\mathbf{y}) \\ \text{subject to} \quad & B\mathbf{y} = \boldsymbol{h}, \quad \mathbf{y} \ge \mathbf{0}. \end{aligned}$$

The minimization of $J_\varepsilon$ in (4.9) is hence equivalent to minimizing a smooth nonconvex function under a set of linear constraints. This optimization problem can be solved using the approach developed in section 5.

**5. Interior point method for local constrained nonconvex minimization.** The necessary conditions for a solution to (4.8) are formulated with the aid of its Lagrangian function. The latter reads

$$\mathcal{L}(\mathbf{y},\lambda,s)=\frac{1}{2}\mathbf{y}^T M\mathbf{y}+\mathbf{c}^T\mathbf{y}+\Psi_\varepsilon(\mathbf{y})-\lambda^T B\mathbf{y}-\mathbf{s}^T\mathbf{y}.$$

Let $Y=\ \mathrm{diag}(\mathbf{y})$ and $S=\ \mathrm{diag}(\mathbf{s})$. The optimality condition then reads

$$F(\mathbf{y},\boldsymbol{\lambda},\mathbf{s})=\begin{bmatrix} M\mathbf{y}+\mathbf{c}+\nabla\Psi_\varepsilon(\mathbf{y})-B^T\boldsymbol{\lambda}-\mathbf{s}\\ B\mathbf{y}\\ YS\mathbf{1}\end{bmatrix}=0,$$

$$\mathbf{y}\geq 0,\quad s\geq 0, \tag{5.1}$$

where

$$\nabla\Psi_\varepsilon(\mathbf{y})=\begin{bmatrix}\mathbf{0}\\ \nabla_{\mathbf{v}^+}\Psi_\varepsilon(\mathbf{v}^+)\\ \nabla_{\mathbf{v}^-}\Psi_\varepsilon(\mathbf{v}^-)\end{bmatrix}$$

with components calculated according to (4.7),

$$\nabla_{\mathbf{v}^+}\Psi_\varepsilon(\mathbf{y})=\beta\begin{bmatrix}\psi'_\varepsilon(v_1^+)\\ \psi'_\varepsilon(v_2^+)\\ \vdots\\ \psi'_\varepsilon(v_r^+)\end{bmatrix},\quad \nabla_{\mathbf{v}^-}\Psi_\varepsilon(\mathbf{y})=\beta\begin{bmatrix}\psi'_\varepsilon(v_1^-)\\ \psi'_\varepsilon(v_2^-)\\ \vdots\\ \psi'_\varepsilon(v_r^-)\end{bmatrix}.$$

One may be induced to solve (4.8), and correspondingly (5.1), using standard tools such as the projected gradient method. Finding at each iteration a descent direction that satisfies the active constraints involves projection operators that are singular, so the numerical task is impractical. Instead, we focus on a primal-dual interior point method where the constraints are strictly satisfied. A description of these methods can be found in [53, 41]. In the primal-dual interior point methods, the search direction is modified with respect to the Newton method in such a way that it focuses on points on a *central* path $(\mathbf{y}_\tau,\boldsymbol{\lambda}_\tau,\mathbf{s}_\tau)$ of the form

$$F(\mathbf{y}_\tau,\boldsymbol{\lambda}_\tau,\mathbf{s}_\tau)=\begin{bmatrix}0\\ 0\\ \tau\mathbf{1}\end{bmatrix},\quad \mathbf{y}_\tau>0,\quad s_\tau>0.$$

The parameter $\tau$ is often written as $\sigma\mu$, where $\sigma\in[0,1]$ is a centering parameter, and $\mu$ is the duality measure defined by

$$\mu=\frac{1}{p+2r}\sum_{i=1}^{p+2r}y_i s_i=\frac{\mathbf{y}^T\mathbf{s}}{p+2r}. \tag{5.2}$$

The Newton search direction $(\Delta\mathbf{y},\Delta\boldsymbol{\lambda},\Delta\mathbf{s})$ is computed by solving the system

$$\begin{bmatrix} M+\nabla^2\Psi_\varepsilon(\mathbf{y}) & -B^T & -I\\ B & 0 & 0\\ S & 0 & Y\end{bmatrix}\begin{bmatrix}\Delta\mathbf{y}\\ \Delta\boldsymbol{\lambda}\\ \Delta\mathbf{s}\end{bmatrix}=\begin{bmatrix}-\mathbf{r}_c\\ -\mathbf{r}_b\\ -\mathbf{r}_a\end{bmatrix},$$

where

$$\mathbf{r}_c = M\mathbf{y} + c + \nabla\Psi_\varepsilon(\mathbf{y}) - B^T\boldsymbol{\lambda} - \mathbf{s},$$

$$\mathbf{r}_b = B\mathbf{y}, \quad \mathbf{r}_a = YS\mathbf{1} - \sigma\mu\mathbf{1}.$$

By eliminating $\Delta s$ from the above equation, we obtain

$$\begin{bmatrix} M + \nabla^2\Psi_\varepsilon + Y^{-1}S & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{y} \\ -\Delta\lambda \end{bmatrix} = \begin{bmatrix} -\hat{\mathbf{r}}_c \\ -\mathbf{r}_b \end{bmatrix}, \tag{5.3}$$

where

$$\hat{\mathbf{r}}_c = \mathbf{r}_c + Y^{-1}\mathbf{r}_a.$$

Since $\nabla^2\Psi_\varepsilon(\mathbf{y}_i)$ is negative semidefinite, we would like to guarantee that $\Delta\mathbf{y}$ and $\Delta\lambda$ can be determined in the Newton step toward points on the central path; therefore the term $\nabla^2\Psi_\varepsilon(\mathbf{y}_i)$ is removed in (5.3). Now the equation in (5.3) becomes

$$\begin{bmatrix} M + Y^{-1}S & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{y} \\ -\Delta\lambda \end{bmatrix} = \begin{bmatrix} -\hat{\mathbf{r}}_c \\ -\mathbf{r}_b \end{bmatrix}.$$

Since $M+Y^{-1}S$ is always positive definite, the solutions $\Delta\lambda$ and $\Delta\mathbf{y}$ can always be determined as follows:

$$B(M + Y^{-1}S^{-1})B^T\Delta\lambda = -\mathbf{r}_b + B(M + Y^{-1}S)^{-1}\hat{\mathbf{r}}_c$$

and

$$\Delta\mathbf{y} = (M + Y^{-1}S)^{-1}(-\hat{\mathbf{r}}_c + B^T\Delta\lambda).$$

Furthermore, as $B$ is a full-rank matrix, $\Delta\lambda$ can be computed.

Taking advantage of the special structures of the matrices $M$, $B$, and $\nabla\Psi_\varepsilon(\mathbf{y})$, the above system can be further reduced. Let $L_1$, $L_2$, and $L_3$ be the diagonal matrices of size $p \times p$, $r \times r$, and $r \times r$, respectively, such that

$$S^{-1/2}Y^{1/2} = \begin{bmatrix} L_1 & 0 & 0 \\ 0 & L_2 & 0 \\ 0 & 0 & L_3 \end{bmatrix}.$$

Then the system can be put into the following form:

$$\begin{bmatrix} 2H^TH + L_1^{-2} & 0 & 0 & D^T \\ 0 & L_2^{-2} & 0 & -I \\ 0 & 0 & L_3^{-2} & I \\ D & -I & I & 0 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{f} \\ \Delta\mathbf{v}^+ \\ \Delta\mathbf{v}^- \\ -\Delta\lambda \end{bmatrix} = \begin{bmatrix} -\hat{\mathbf{r}}_{c1} \\ -\hat{\mathbf{r}}_{c2} \\ -\hat{\mathbf{r}}_{c3} \\ -\mathbf{r}_b \end{bmatrix}.$$

Here $\hat{\mathbf{r}}_{ci}(i = 1, 2, 3)$ are subvectors of $\hat{\mathbf{r}}_c$ of size $p$, $r$, and $r$, respectively. By eliminating $\Delta\mathbf{v}^+$ and $\Delta\mathbf{v}^-$, we get

$$\begin{bmatrix} 2H^TH + L_1^{-2} & D^T \\ D & -L_2^2 - L_3^2 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{f} \\ -\Delta\boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\hat{\mathbf{r}}_{c1} \\ -\hat{\mathbf{r}}_b \end{bmatrix},$$

where

$$\hat{\mathbf{r}}_b = \mathbf{r}_b + L_2^2\mathbf{r}_{c2} - L_3^2\mathbf{r}_{c3}.$$

Eliminating $\Delta\lambda$, we get

$$[2H^TH + L_1^{-2} + D^T(L_2^2 + L_3^2)^{-1}D]\Delta\mathbf{f} = -\tilde{\mathbf{r}}_{c1}, \tag{5.4}$$

where

$$\tilde{\mathbf{r}}_{c1} = \hat{\mathbf{r}}_{c1} + D^T(L_2^2 + L_3^2)^{-1}\mathbf{r}_b.$$

The other unknowns can be calculated by

$$\begin{aligned} \Delta\lambda &= (L_2^2 + L_3^2)^{-1}(-\hat{\mathbf{r}}_b - D\Delta\mathbf{f}), \\ \Delta\mathbf{v}^+ &= L_2^2(-\hat{\mathbf{r}}_{c2} + \Delta\boldsymbol{\lambda}), \\ \Delta\mathbf{v}^- &= L_3^2(-\hat{\mathbf{r}}_{c3} - \Delta\boldsymbol{\lambda}), \\ \Delta s &= M\Delta\mathbf{y} - B^T\Delta\lambda + \mathbf{r}_c. \end{aligned} \tag{5.5}$$

**5.1. Initialization.** Interior point methods must start with a feasible initialization. In our case, the strictly feasible set reads

$$\begin{aligned} \mathcal{F}_\varepsilon^0 = \{(\mathbf{f},\boldsymbol{\lambda},\mathbf{s}) | M\mathbf{y} + c + \nabla\Psi_\varepsilon(\mathbf{y}) - B^T\boldsymbol{\lambda} - \mathbf{s} = 0, \\ B\mathbf{y} = 0, \mathbf{f} > 0, \mathbf{s} > 0\}. \end{aligned}$$

Next we explain how to find an element belonging to $\mathcal{F}_\varepsilon^0$. Let $\mathbf{f}_{es}$ be any estimate of the solution, and let $\delta > 0$. We choose $\mathbf{f}^{(0)}$ to be

$$\mathbf{f}^{(0)}[i] = \max\{\mathbf{f}_{es}[i], \delta\}. \tag{5.6}$$

We let $\mathbf{v}^{(0)+}$ be the sum of the nonnegative part of $D\mathbf{f}^{(0)}$ and $\delta$, and $\mathbf{v}^{(0)-}$ be the sum of the nonpositive part of $D\mathbf{f}^{(0)}$ and $\delta$. Then the vector

$$\mathbf{y}^0 = [\mathbf{f}^{(0)^T}, \mathbf{v}^{+(0)^T}, \mathbf{v}^{-(0)^T}]^T$$

satisfies $B\mathbf{y}^0 = \mathbf{0}$. To find $\mathbf{s}^{(0)}$ and $\boldsymbol{\lambda}^{(0)}$, there are many different ways only if they can satisfy $M\mathbf{y} + \mathbf{c} + \nabla\Psi_\varepsilon(\mathbf{y}) - B^T\boldsymbol{\lambda} - \mathbf{s} = \mathbf{0}$. One simple method is to set $\boldsymbol{\lambda} = 0$ and to iterate $\mathbf{f}^0$ by adding $\delta$ in (5.6) until $s = M\mathbf{y}^0 + \mathbf{c} + \nabla\Psi_\varepsilon(\mathbf{y}^0) - B^T$ is strictly positive.

**5.2. Algorithm for nonconvex programming problems.** To solve the whole problem, as described in section 3, we need to compute a sequence of minimization problems $\min J_\varepsilon$ from $\varepsilon = 0$ to $\varepsilon = 1$. The initial point in solving $\min J_{\varepsilon_k}$ is the solution of the $\min J_{\varepsilon_{k-1}}$.

Let $\varepsilon_0 = 0 < \varepsilon_1 < \cdots < \varepsilon_n = 1$. Start from $\varepsilon(0)$, $j_0 = \frac{1}{\Delta\varepsilon}$. Initialize $(\mathbf{y}^{(0)}, \boldsymbol{\lambda}^{(0)}, \mathbf{s}^{(0)})$.
Calculate $\mu$ according to (5.2).
For $k = 1 \to n$, $\varepsilon = \varepsilon_k$
    j=0
    While $\mu > tol$ do
        Choose $\sigma \in [0, 1]$;
        Solve (5.4) for $\Delta\mathbf{f}$;
        Use (5.5) to find $\Delta\mathbf{v}^+, \Delta\mathbf{v}^-, \Delta\boldsymbol{\lambda}, \Delta\mathbf{s}$;

Choose a step length $\nu$;
$\mathbf{y}^{(j+1)} = \mathbf{y}^{(j)} + \nu\Delta\mathbf{y}$;
$\boldsymbol{\lambda}^{(j+1)} = \boldsymbol{\lambda}^{(j)} + \nu\Delta\boldsymbol{\lambda}$;
$\mathbf{s}^{(j+1)} = \mathbf{s}^{(j)} + \nu\Delta\mathbf{s}$;
Update $\mu$;
$j = j + 1$;
End While
$\mathbf{y}^{(0)} = \mathbf{y}^{(k)}$;
$\boldsymbol{\lambda}^{(0)} = \boldsymbol{\lambda}^{(k)}$;
$\mathbf{s}^{(0)} = \mathbf{s}^{(k)}$;
$\varepsilon(j+1) = \varepsilon(j) + \Delta\varepsilon$;
End For

We remark that in our experiments, a linear increase for $\varepsilon_k$ is considered. We also tested other possible increases for $\varepsilon_k$, but their differences are not significant. How to determine an optimal increase for $\varepsilon_k$ is a question for future research work.

Finally, let us discuss the computational cost of the proposed continuation method. In the continuation method, we must solve (4.5) for several values of $\varepsilon$ by the interior point method. In each interior point iteration, we must solve a linear system in (5.4). In [19], we have shown how to use the preconditioned conjugate gradient technique with an effective preconditioner to solve such a system efficiently. The computational cost of the preconditioning conjugate gradient technique is $O(p \log p)$, where $p$ is the size of the computed solution where the image is in a vector form. It follows that the overall computational cost of the proposed method is $O(cp \log p)$ operations, where $c$ depends on the number of values of $\varepsilon$ to be tested and the total number of interior point iterations. According to our numerical results, the number of values of $\varepsilon$ is 10 and the total number of interior point iterations is about 200.

**6. Numerical experiments.** In this section, we present the experimental results to show the effectiveness of our method. The PF is applied:

$$\varphi(t) = \frac{\alpha|t|}{1+\alpha|t|}, \quad \varphi_\varepsilon(t) = \frac{\alpha|t|}{1+\varepsilon\alpha|t|}, \ 0 \le \varepsilon \le 1, \tag{6.1}$$

$$\varphi(t) = \log(1+\alpha|t|), \quad \varphi_\varepsilon(t) = \log(1+\varepsilon\alpha|t|), \ 0 \le \varepsilon \le 1. \tag{6.2}$$

Three images are tested, which are all gray level images with intensity values ranging from 0 to 1. The first two are based on two artificial images: a circles image (Figure 2) and a diamonds image (Figure 6). The first image is of size $64 \times 64$, and the second is of size $50 \times 50$. For each of these two images, we generate two observed images. One is that all the pixels are contaminated by Gaussian noise with the standard deviation of 0.1 without blurring, and the other is that the pixels are contaminated by Gaussian noise with the standard deviation of 0.05 with blurring. The blurring function is chosen to be a two-dimensional Gaussian,

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

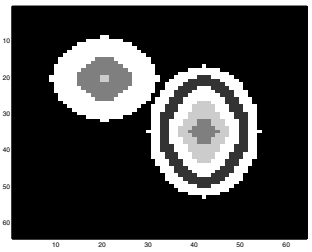


**Figure 2.** *The original circles image.*

truncated such that the function has a support of $7 \times 7$. For the circles image, we choose $D$ to be a matrix of a first-order difference operator:

$$D_1 = \begin{pmatrix} L_1 \otimes I \\ I \otimes L_1 \end{pmatrix} \quad \text{with} \quad L_1 = \begin{pmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \end{pmatrix}.$$

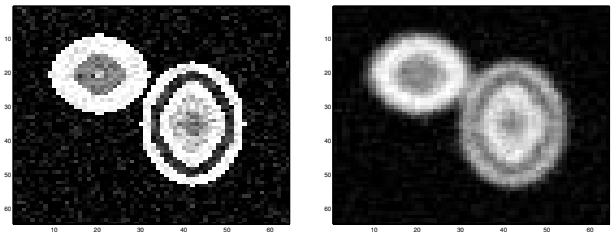
For the diamonds image, we choose $D$ to be a matrix of a second-order difference operator:

$$D_2 = \begin{pmatrix} L_2 \otimes I \\ I \otimes L_2 \end{pmatrix} \quad \text{with} \quad L_2 = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix}$$

with the Neumann boundary condition. The third image is the modified Shepp–Logan image (Figure 10), which is used to test the effectiveness of restoration algorithms. For this image, we perform the restoration when it is under the Radon transform rather than the convolution.

To demonstrate the effectiveness of our GNC method, we also applied the interior point method to the problem with $\varepsilon = 1$; i.e., we directly solve the involved optimization problem. In the following description, "$\varepsilon = 1$" means an interior point method is directly applied to the optimization problem, while "$\varepsilon = 0 \to 1$" means the GNC technique is resorted.

**6.1. Test of circles image.** Figure 3 shows the observed images without blurring and with blurring described in the above setting. Four initial guesses are employed: the observed image (OB), the least square solution (LS), a flat image (FL) (all the pixel values are 0.5), and a random image (RD). Tables 2 and 3 show the objective function values and the peak signal noise ratio (PSNR) values by using different initial guesses. Here the PF in (6.1) and the optimal regularization parameters $\alpha$ and $\beta$ are used. In the computational process, we tested several values of $\alpha$ and $\beta$ such that the resulting PSNR of the restored image is the largest. It is clear from the tables that the proposed method ($\varepsilon = 0 \to 1$) can provide similar image restoration results for different initial guesses and give smaller objective function values and larger PSNR values of the restored images than those of the noncontinuation method ($\varepsilon = 1$) when initial guesses are flat images and random images. As an example, we further show in



(a) (b)

**Figure 3.** *Observed circles images:* (a) *all pixels are contaminated by Gaussian noise with deviation* 0.1 *without blurring;* (b) *Gaussian blurring and all pixels are contaminated by Gaussian noise with deviation* 0.05.

**Table 2**

*Objective values and PSNRs for the observed circles image without blurring when* $\alpha = 0.5$, $\beta = 0.3$. *The objective function value for the original image is* 84.48.

| | | FL | RD | OB | LS |
|---|---|---|---|---|---|
| $\varepsilon = 1$ | Objective | 82.02 | 81.89 | 81.62 | 81.63 |
| $\varepsilon = 0 \to 1$ | Objective | 81.63 | 81.63 | 81.63 | 81.63 |
| $\varepsilon = 1$ | PSNR | 30.66 | 30.85 | 31.31 | 31.28 |
| $\varepsilon = 0 \to 1$ | PSNR | 31.28 | 31.28 | 31.28 | 31.28 |

**Table 3**

*Objective values and PSNRs for the observed circles image with blurring when* $\alpha = 0.5$, $\beta = 0.02$. *The objective value for the true image is* 13.56.

| | | FL | RD | OB | LS |
|---|---|---|---|---|---|
| $\varepsilon = 1$ | Objective | 14.65 | 21.02 | 13.51 | 13.14 |
| $\varepsilon = 0 \to 1$ | Objective | 12.84 | 12.84 | 12.84 | 12.84 |
| $\varepsilon = 1$ | PSNR | 18.92 | 17.35 | 19.42 | 19.64 |
| $\varepsilon = 0 \to 1$ | PSNR | 19.03 | 19.03 | 19.03 | 19.03 |

Figures 4 and 5 that the proposed method can provide better quality of restored images than those of the non-continuation method when initial guesses are flat images and random images. Also we remark that we tried the PF in (6.2) and found that the numerical results obtained are similar to those in Tables 2 and 3.

**6.2. Test of diamonds image (Figure 6).** Figure 7 is the observed blurred image. In Figures 8 and 9, we showed the restored images by using the noncontinuation method and the proposed method, respectively. Table 4 shows their objective function values and their PSNR values by using different initial guess that we used in the previous subsection. We see that the proposed continuation method provides slightly better image restoration results in the table than those obtained by the noncontinuation method when different initial guesses are used. Here the PF in (6.1) and the optimal regularization parameters $\alpha$ and $\beta$ are used. When the PF in (6.2) is used, we find that the numerical results obtained are similar to those in Table 4.

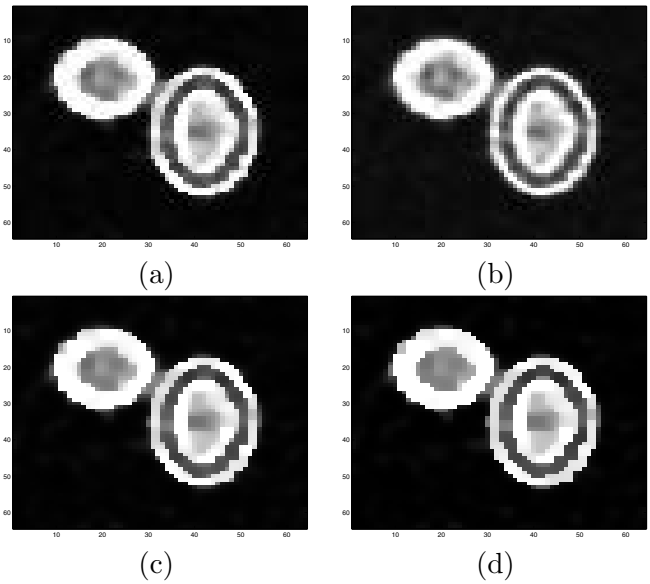


**Figure 4.** *Restored blurred circles images for PF* (6.1) *when* $\alpha = 0.5$, $\beta = 0.02$, $\varepsilon = 1$: (a) *the initial guess is the flat image;* (b) *the initial guess is the random image;* (c) *the initial guess is the observed image;* (d) *the initial guess is the least square solution.*

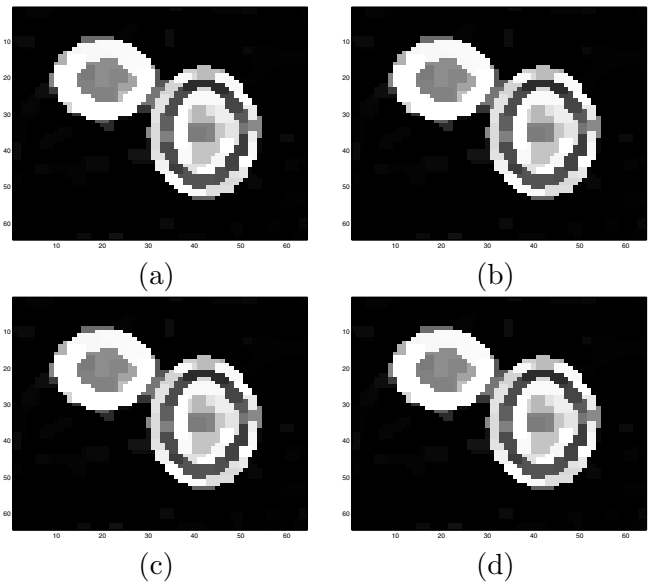


**Figure 5.** *Restored blurred circles images for PF* (6.1) *when* $\varepsilon = 0 \to 1$, $\alpha = 0.5$, $\beta = 0.02$: (a) *the initial guess is the flat image;* (b) *the initial guess is the random image;* (c) *the initial guess is the observed image;* (d) *the initial guess is the least square solution.*

**6.3. Radon transform of modified Shepp–Logan image.** Radon transform is able to transform two-dimensional images with lines into a domain of possible line parameters, where each line in the image will give a peak positioned at the corresponding line parameters. This had led to many line detection applications within image processing, computer vision, and

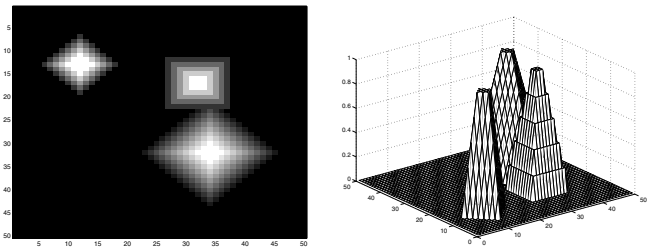


**Figure 6.** *The original diamonds image.*

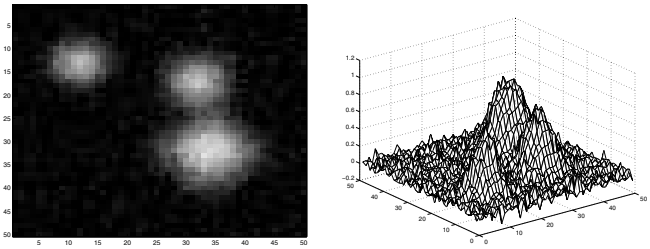


**Figure 7.** *Observed diamonds images: Gaussian blurring and all pixels are contaminated by Gaussian noise with deviation* 0.05.

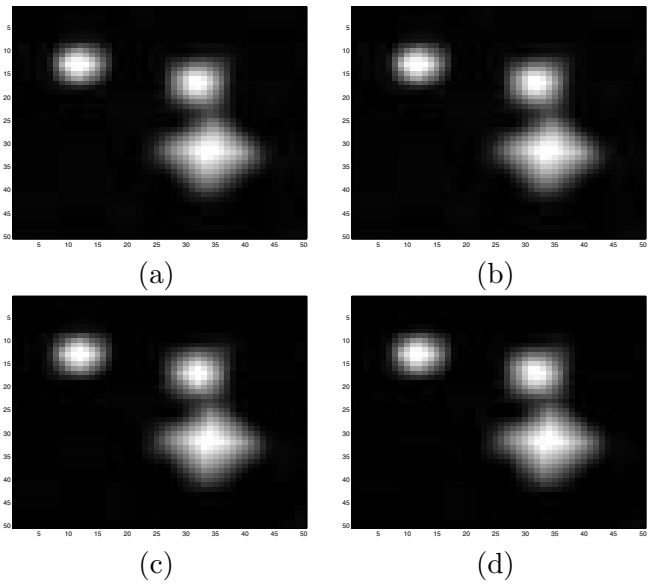



**Figure 8.** *Restored blurred diamonds images for PF* (6.1) *when* $\alpha = 0.5$, $\beta = 0.05$, $\varepsilon = 1$: (a) *the initial guess is the flat image;* (b) *the initial guess is the random image;* (c) *the initial guess is the observed image;* (d) *the initial guess is the least square solution.*

seismics [51]. The inverse of the transform can reconstruct the images. The Radon transform of a two-dimensional function is found by stacking or integrating the function along some slanted lines. A discrete approximation to the Radon transform that transforms digital images is very useful. As mentioned in section 1, the matrix $H$ in the image degradation model is the discrete Radon transform matrix in this application.

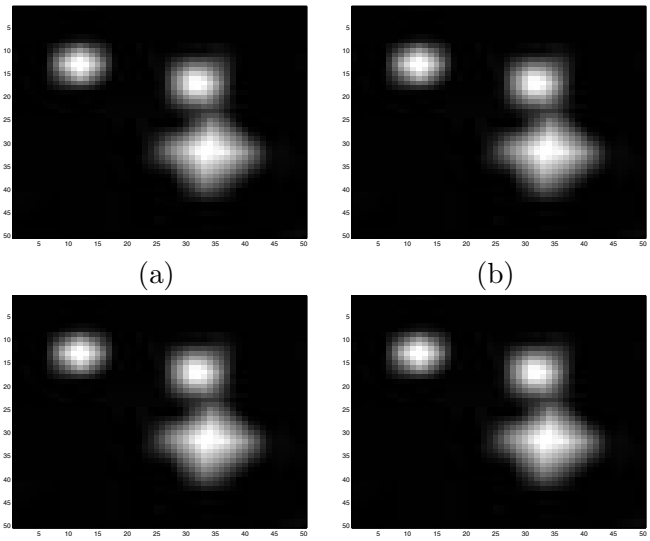



**Figure 9.** *Restored blurred diamonds images for PF (6.1) when $\varepsilon = 0 \to 1$, $\alpha = 0.5$, $\beta = 0.05$:* (a) *the initial guess is the flat image;* (b) *the initial guess is the random image;* (c) *the initial guess is the observed image;* (d) *the initial guess is the least square solution.*

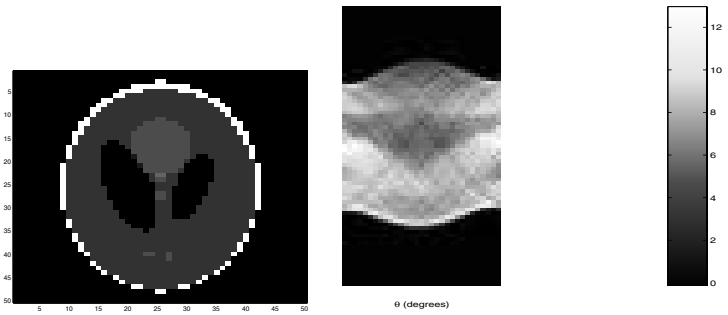
**Table 4**

*Objective values and PSNRs for the observed diamond image with blurring when $\alpha = 0.5$, $\beta = 0.05$. The objective value for the true image is* 7.38.

| | | FL | RD | OB | LS |
|---|---|---|---|---|---|
| $\varepsilon = 1$ | Objective | 7.08 | 7.17 | 6.93 | 6.90 |
| $\varepsilon = 0 \to 1$ | Objective | 6.88 | 6.88 | 6.88 | 6.88 |
| $\varepsilon = 1$ | PSNR | 28.01 | 28.07 | 28.25 | 28.55 |
| $\varepsilon = 0 \to 1$ | PSNR | 28.59 | 28.59 | 28.59 | 28.59 |

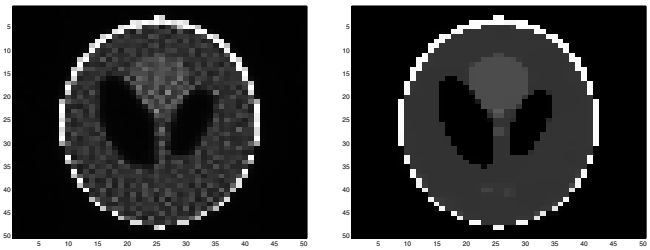
In this subsection, the image reconstruction from images transformed by the Radon transform using our proposed method is presented. The modified Shepp–Logan image which is used in image restoration is applied to illustrate the efficiency of our algorithm. Here we set the image to be of size $50 \times 50$. Figure 10(a) is the original modified Shepp–Logan image. The image is transformed along the angles from 0 to 180 degrees. We choose the transformed line along one angle out of 6 angles. That is, we choose the transformed line along angles 0, 6, $12, \ldots, 180$. There are in total 31 angles in the calculation. For an image of size $50 \times 50$, the Radon transform of it along one angle is of size $75 \times 1$. We represent the image as a column vector of size 2500. The corresponding matrix $H$ has $75 \times 31$ rows and 2500 columns. Figure 10(b) is the Radon transform of the image. The noise-based normal distribution with mean zero and deviation 0.05 is added to the transformed image.

The regularization parameter $\beta$ is chosen from 0 to 1. We find that when $\beta$ is 0.8, the largest PSNR value of the restored image is obtained. Figure 11 shows the restored image when the initial guess is the flat image. Figure 11(a) is the restored image when the interior



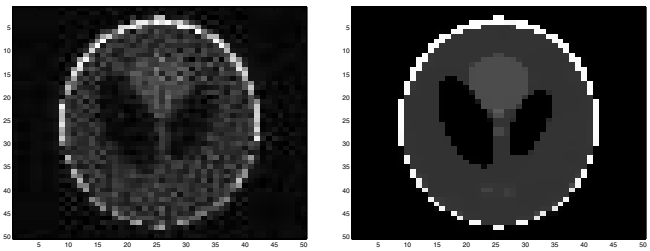

(a) (b)

**Figure 10.** (a) *The original modified Shepp–Logan image with size* $50 \times 50$*;* (b) *the obtained image after Radon transform along the angles from* 0 *to* 180 *with the increasing of* 6 *degrees. The noise follows the normal distribution with mean zero and deviation* 0.05*.*



(a) (b)

**Figure 11.** *Restored Shepp–Logan images for PF* (6.1) *when the initial guess is a flat image of constant value* 25: (a) *the restored image when* $\varepsilon = 1$*,* $PSNR = 25.82$*;* (b) *the restored image when* $\varepsilon = 0 \to 1$*,* $PSNR = 41.96$*.*



(a) (b)

**Figure 12.** *Restored Shepp–Logan images for PF* (6.1) *when the initial guess is a random image*: (a) *the restored image when* $\varepsilon = 1$*,* $PSNR = 20.23$*;* (b) *the restored image when* $\varepsilon = 0 \to 1$*,* $PSNR = 41.96$*.*

point method is directly applied to solve the formulated problem and Figure 11(b) is that when $\varepsilon$ is taken from 0 to 1. It is clear that Figure 11(b) is very close to the original image, while Figure 11(a) is far from the original. Figure 12 shows the restored image when the initial guess is a random image. We also see that when the continuation technique is applied, the restored image is quite good.

We make a comparison between the nonconvex regularization and the convex regularization. We choose one convex, nonsmooth PF: $\psi(t) = |t|$. We choose the best results for both

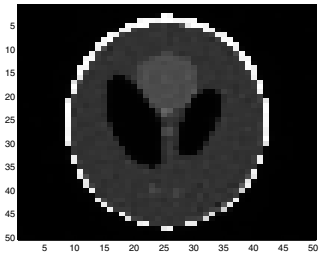
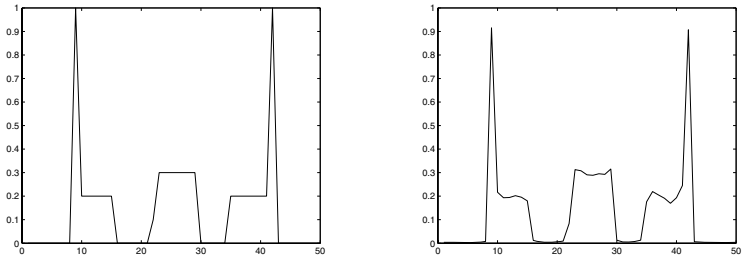
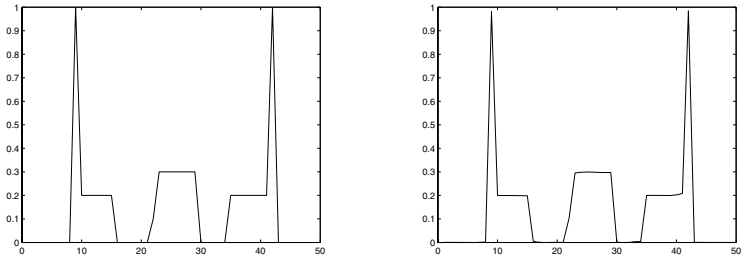


**Figure 13.** *The restored Shepp–Logan image with PF* $\psi(t) = |t|$. *The initial guess is a flat image,* $PSNR = 34.71$.



(a) (b)

**Figure 14.** (a) *The* 21*st line in the original Shepp–Logan image;* (b) *the restored* 21*st line with convex regularization.*

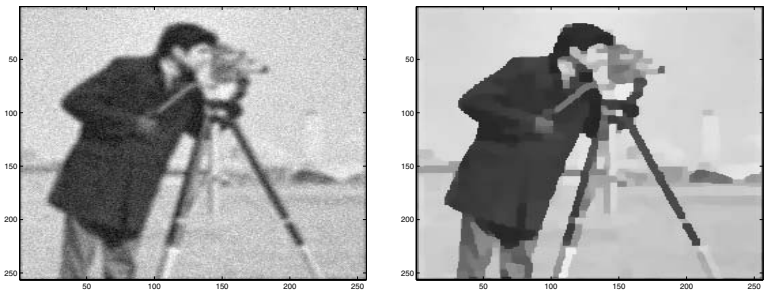


(a) (b)

**Figure 15.** (a) *The restored* 21*st line with PF in* (6.1)*;* (b) *the restored* 21*st line with PF in* (6.2).

when $\beta$ is taken from 0 to 1. For the convex regularization, the best PSNR value is obtained when $\beta = 0.1$. Figure 13 shows the restored image with convex regularization, and this PSNR value is 34.71, which is less than those with nonconvex regularization (the PSNR values are 41.96 in Figures 11(b) and 12(b)). In particular, we choose the 21st line of both restored images to see the restoration effect. Figure 14(a) is the image corresponding to the 21st line in the original image. Figure 14(b) is the restored result when the convex PF is applied. Figures 15(a) and 15(b) show the restored results when the nonconvex PFs in (6.1) and (6.2) are applied, respectively. The restored images are very close to the original image when the nonconvex PF is used, where sharp edges appear and the flat parts restore very well.

**Figure 16.** *The original cameraman image.*



(a) (b)

**Figure 17.** (a) *Observed image;* (b) *the restored image.*

**6.4. Test of cameraman image.** In this subsection, we present the restoration of a blurred and noisy $256 \times 256$ "cameraman" image that contains more edges. The original image is shown in Figure 16. The image is blurred by a Gaussian function $h(i,j) = e^{-(i/3)^2-(j/3)^2}$ truncated such that the function has a support of $7 \times 7$, and then the pixels are contaminated by Gaussian noise with the standard deviation of 0.05. The blurred and noisy image is shown in Figure 17(a). The restored image using the proposed method using a nonconvex PF in (6.1) is shown in Figure 17(b) with $\alpha = 0.5$ and $\beta = 0.12$. We see that the piecewise constant image appears in the restoration result.

**6.5. Comparison with simulated annealing method.** In this subsection, we make a comparison between our method and the SA method when the blur is not present. When the blur is present, the SA method does not work in practice.

We choose a temperature sequence $\{T_k, k = 1, \ldots\}$ and initialize the configuration $\mathbf{f}(0)$ of the image. For each $k = 1, 2, \ldots$, the pixels are updated one by one. Each pixel will follow the distribution

$$\Pi_T(\mathbf{f}_{i,j} | \mathbf{f}_{k,l}, (i,j) \neq (k,l)) = \exp(-J(\mathbf{f})/T),$$

where $T$ is the temperature. To generate this distribution, the acceptance-rejection method is applied. Let $c$ be a constant which satisfies

$$\frac{\Pi_T(\mathbf{f}_{i,j} | \mathbf{f}_{k,l}, (i,j) \neq (k,l))}{g(y)} \leq c \quad \forall y,$$

**Table 5**

*Results using SA when $\alpha = 0.5$, $\beta = 0.3$ for the observed circles image without blurring. The objective function value and PSNR obtained by the proposed continuation method are* 81.62 *and* 31.28.

| Initial image | Initial $T$ | Number of iterations $k$ | Relation between $k$ and $T$ | Objective function value | PSNR |
|---|---|---|---|---|---|
| Random image | 1 | 40,000 | $T = \frac{1}{\log(k+2)}$ | 206.52 | 17.69 |
| Observed image | 0.08 | 30,000 | $T = \frac{0.08}{\log(k+2)}$ | 93.77 | 27.68 |
| Computed image by the interior point method with $\varepsilon = 1$ | $\frac{0.005}{\log(20000)}$ | 15,000 | $T = \frac{0.005}{\log(k+20000)}$ | 82.75 | 31.17 |

where $g(y) = 1$, which is the probability distribution function of the uniform distribution on [0,1]. The algorithm is as follows:

Step 1: Generate $Y$ has the distribution g(y);

Step 2: Generate a random number $U$;

Step 3: If $U \leq \frac{\Pi_T(\mathbf{f}_{i,j}|\mathbf{f}_{k,l},(i,j)\neq(k.l))}{cg(Y)}$, set $\mathbf{f}_{i,j} = U$. Otherwise, return to step 1.

We choose three initial images and a proper $T$ such that the initial image is a distribution at temperature $T$. When $T$ decreases, the objective value should decrease with some fluctuations. We did experiments for the circles image without blurring. When there is no blurring, we test three initial guesses. The first initial image is the random image, the second is the observed image, and the third is the image we get from the interior point method with $\varepsilon = 1$ (i.e., no continuation).

Table 5 shows the settings and the results of SA tests. In the SA tests, the objective function value decreases with some fluctuations when $T$ decreases. When the decrease rate of $T$ is slower, the decrease rate of the objective function values for the first two initial images is also slower. For the third image, since the objective function value is already close to the minimal value, we need to choose a small $T$ correspondingly. In such a case, when $T$ is further decreasing, the objective function value cannot decrease any more. The main reason is that $T$ cannot be smaller than $\frac{0.005}{\log(20000)}$, as the corresponding probability distribution function is very close to zero. For the computational time, 10,000 iterations in the SA method cost more than 12 hours. It took a few days in order to get the resulting objective function value in the table. However, the proposed continuation method takes only a few minutes to provide a very good solution comparable to that determined by the SA method. The objective function value and the PSNR value obtained by the proposed continuation method are 81.62 and 31.28, which are much better than those obtained by the SA method (82.75 and 31.17) in Table 5.

## 7. Concluding remarks.

The restoration of images with nonconvex and nonsmooth regularization can give rise to restored images with neat edges and homogeneous regions elsewhere. This benefits many practical applications of image restoration. Solving such an optimization problem is a challenging task since it involves many minima due to the nonconvex property. In this paper, we proposed an efficient algorithm for solving the optimization problem. The problem is formulated as a series of nonsmooth approximations which are solved by a primal-dual

interior point method. Experiment results showed that the method is quite effective and that the proposed algorithm can determine restored images within a reasonably short period of time.

## REFERENCES


[1] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing*, Springer-Verlag, Berlin, 2002.

[2] M. R. Banham and A. K. Katsaggelos, *Digital image restoration*, IEEE Signal Processing Magazine, 14 (2) (1997), pp. 24–41.

[3] L. Bar, N. Sochen, and N. Kiryati, *Semi-blind image restoration via Mumford-Shah regularization*, IEEE Trans. Image Process., 15 (2006), pp. 483–493.

[4] L. Bedini, I. Gerace, and A. Tonazzini, *A deterministic algorithm for reconstructing images with interacting discontinuities*, CVGIP: Graph. Models Image Process., 56 (1994), pp. 109–123.

[5] L. Bedini, I. Gerace, and A. Tonazzini, *A GNC algorithm for constrained image reconstruction with continuous-value line process*, Pattern Recogn. Lett., 15 (1994), pp. 907–918.

[6] J. E. Besag, *On the statistical analysis of dirty pictures,* J. Roy. Statist. Soc. Ser. B, 48 (1986), pp. 259–302.

[7] J. E. Besag, *Digital image processing: Towards Bayesian image analysis*, J. Appl. Statist., 16 (1989), pp. 395–407.

[8] G. Bilbro, W. Snyder, and R. Mann, *Mean field approximation minimizes relative entropy*, J. Opt. Soc. Amer. A, 8 (1991), pp. 290–294.

[9] M. Black and A. Rangarajan, *On the unification of line processes, outlier rejection, and robust statistics with applications to early vision*, Internat. J. Comput. Vision, 19 (1996), pp. 57–91.

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

[11] A. Blake, *Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction*, IEEE Trans. Pattern Anal. Machine Intell., 11 (1989), pp. 2–12.

[12] C. Bouman and K. Sauer, *A generalized Gaussian image model for edge-preserving MAP estimation*, IEEE Trans. Image Process., 2 (1993), pp. 296–310.

[13] 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. Med. Imag., 12 (1993), pp. 610–621.

[14] T. F. Chan and S. Esedoglu, *Aspects of total variation regularized $L^1$ function approximation*, SIAM J. Appl. Math., 65 (2005), pp. 1817–1837.

[15] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, *Deterministic edge-preserving regularization in computed imaging*, IEEE Trans. Image Process., 6 (1997), pp. 298–311.

[16] P. Combettes and J. Luo, *An adaptive level set method for nondifferentiable constrained image recovery*, IEEE Trans. Image Process., 11 (2002), pp. 1295–1304.

[17] A. Delaney and Y. Bresler, *Globally convergent edge-preserving regularized reconstruction: An application to limited-angle tomography*, IEEE Trans. Image Process., 7 (1998), pp. 204–221.

[18] G. Demoment, *Image reconstruction and restoration: Overview of common estimation structure and problems*, IEEE Trans. Acoust. Speech Signal Process., 37 (1989), pp. 2024–2036.

[19] H. Fu, M. K. Ng, M. Nikolova, and J. L. Barlow, *Efficient minimization methods of mixed $\ell 2$-$\ell 1$ and $\ell 1$-$\ell 1$ norms for image restoration*, SIAM J. Sci. Comput., 27 (2006), pp. 1881–1902.

[20] D. Geman and G. Reynolds, *Constrained restoration and recovery of discontinuities*, IEEE Trans. Pattern Anal. Mach. Intell., 14 (1992), pp. 367–383.

[21] D. Geman and C. Yang, *Nonlinear image recovery with half-quadratic regularization*, IEEE Trans. Image Process., 4 (1995), pp. 932–946.

[22] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Trans. Pattern Anal. Mach. Intell., 6 (1984), pp. 721–741.

[23] S. Geman and D. E. McClure, *Statistical methods for tomographic image reconstruction*, in Proceedings of the 46th Session of the ISI, Bulletin of the ISI, 52 (1987), pp. 22–26.

[24] D. Goldfarb and W. Yin, *Second-order cone programming methods for total variation-based image restoration*, SIAM J. Sci. Comput., 27 (2005), pp. 622–645.

[25] P. J. Green, *Bayesian reconstructions from emission tomography data using a modified EM algorithm*, IEEE Trans. Med. Imag., 9 (1990), pp. 84–93.
[26] G. Hewer, C. Kenney, and B. Manjunath, *Variational image segmentation using boundary functions*, IEEE Trans. Image Process., 7 (1998), pp. 1269–1281.
[27] T. Kärkkäinen, K. Kunisch, and K. Majava, *Denoising of smooth images using $L^1$-fitting*, Computing, 74 (2005), pp. 353–376.
[28] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, *Optimization by simulated annealing*, Science, 220 (1983), pp. 671–680.
[29] K. Lange, *Convergence of EM image reconstruction algorithms with Gibbs priors*, IEEE Trans. Med. Imag., 9 (1990), pp. 439–446.
[30] Y. Leclerc, *Constructing simple stable descriptions for image partitioning*, Int. J. Comput. Vis., 3 (1989), pp. 73–102.
[31] S. Li, *Markov Random Field Modeling in Computer Vision*, Springer-Verlag, New York, 1995.
[32] Z. K. Liu and J. Y. Xiao, *Restoration of blurred TV pictures caused by uniform linear motion*, CVGIP: Graph. Models Image Process., 44 (1988), pp. 30–34.
[33] M. Mignotte, *A segmentation-based regularization term for image deconvolution*, IEEE Trans. Image Process., 15 (2006), pp. 1973–1984.
[34] D. Mumford and J. Shah, *Boundary detection by minimizing functionals*, in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 1985, pp. 22–26.
[35] M. R. Nagel, *Introduction to evaluation of motion-degraded images*, in Proceedings of the NASA Electronics Research Center Seminar, Cambridge, MA, 1968.
[36] M. Nikolova, J. Idier, and A. Mohammad-Djafari, *Inversion of large-support ill-posed linear operators using a piecewise Gaussian MRF*, IEEE Trans. Image Process., 8 (1998), pp. 571–585.
[37] M. Nikolova, *Markovian reconstruction using a GNC approach*, IEEE Trans. Image Process., 8 (1999), pp. 1204–1220.
[38] M. Nikolova, *Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers*, SIAM J. Numer. Anal., 40 (2002), pp. 965–994.
[39] M. Nikolova, *A variational approach to remove outliers and impulse noise*, J. Math. Imaging Vision, 20 (2004), pp. 99–120.
[40] M. Nikolova, *Weakly constrained minimization application to the estimation of images and signals involving constant regions*, J. Math. Imaging Vision, 21 (2004), pp. 155–175.
[41] J. Nocedal and S. Wright, *Numerical Optimization*, Springer-Verlag, New York, 1999.
[42] P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion*, IEEE Trans. Pattern Anal. Mach. Intell., 12 (1990), pp. 629–639.
[43] M. Robini, T. Rastello, and I. Magnin, *Simulated annealing, acceleration techniques, and image restoration*, IEEE Trans. Image Process., 8 (1999), pp. 1374–1387.
[44] L. Rudin, S. Osher, and C. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
[45] S. Saquib, C. Bouman, and K. Sauer, *ML parameter estimation for Markov random fields, with applications to Bayesian tomography*, IEEE Trans. Image Process., 7 (1998), pp. 1029–1044.
[46] C. H. Slump, *Real-time image restoration in diagnostic X-ray imaging, the effects on quantum noise*, in Proceedings of the 11th IAPR International Conference on Pattern Recognition, Vol. II, Conference B: Pattern Recognition Methodology and Systems, 1992, pp. 693–696.
[47] W. Snyder, Y. Han, G. Bilbro, R. Whitaker, and S. Pizer, *Image relaxation: Restoration and feature extraction*, IEEE Trans. Pattern Anal. Machine Intell., 17 (1995), pp. 620–624.
[48] R. Stevenson and E. Delp, *Fitting curves with discontinuities*, in Proceedings of the First International Conference on Robust Computer Vision, Seattle, WA, 1990, pp. 127–136.
[49] S. Teboul, L. Blanc-Féraud, G. Aubert, and M. Barlaud, *Variational approach for edge-preserving regularization using coupled PDEs*, IEEE Trans. Image Process., 7 (1998), pp. 387–397.
[50] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*, Winston, Washington, D.C., 1977.
[51] P. Toft, *The Radon Transform - Theory and Implementation*, Ph.D. Thesis, Department of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark, 1996.
[52] E. Wasserstrom, *Numerical solutions by the continuation method*, SIAM Rev., 15 (1973), pp. 89–119.
[53] M. Wright, *Interior methods for constrained optimization*, in Acta Numerica, 1992, Acta Numer., Cambridge University Press, Cambridge, UK, 1992, pp. 341–407.