

# A Smoothing Descent Method for Nonconvex $TV^q$ -Models

Michael Hintermüller<sup>1</sup>(✉) and Tao Wu<sup>2</sup>

<sup>1</sup> Department of Mathematics, Humboldt-University of Berlin, Unter den Linden 6,  
10099 Berlin, Germany

hint@math.hu-berlin.de

<sup>2</sup> Institute for Mathematics and Scientific Computing, University of Graz,  
Heinrichstrasse 36, 8010 Graz, Austria

tao.wu@edu.uni-graz.at

**Abstract.** A novel class of variational models with nonconvex  $\ell_q$ -norm-type regularizations ( $0 < q < 1$ ) is considered, which typically outperforms popular models with convex regularizations in restoring sparse images. Due to the fact that the objective function is nonconvex and non-Lipschitz, such models are very challenging from an analytical as well as numerical point of view. In this work a smoothing descent method with provable convergence properties is proposed for computing stationary points of the underlying variational problem. Numerical experiments are reported to illustrate the effectiveness of the new method.

## 1 Introduction

In signal or image recovery from sparse data, it has been observed that models based on nonconvex regularization typically outperform convex ones. In particular, variational models relying on  $\ell_q$ -norm-type regularization with  $q \in (0, 1)$  are of interest [3, 9, 16]. Due to the particular choice of  $q$ , the associated functional  $\|v\|_q^q := \sum_i |v_i|^q$  turns out to be nonconvex, nondifferentiable and not even locally Lipschitz continuous, and it is not a norm in the usual sense. These properties imply several challenges from an analytical as well as numerical point of view in the treatment of such problems. In fact, analytically generalized derivative concepts are challenged [7] and, thus, the first-order optimality description becomes an issue. Concerning the numerical solution of general nonsmooth minimization problems, we mention that while convex problems are rather well understood [13, 14, 19] nonconvex ones are still challenging [1, 14]. For nonconvex and non-locally-Lipschitz problems, the literature on methods based on generalized derivatives is rather scarce. However, taking the particular format of the underlying nonsmoothness into account and possibly applying tailored (vanishing) smoothing concepts, in [4, 6, 9, 12, 15, 16] solvers of the associated minimization problems in sparse signal or image recovery were developed recently.

---

This research was supported by the Austrian Science Fund (FWF) through START project Y305 “Interfaces and Free Boundaries” and through SFB project F3204 “Mathematical Optimization and Applications in Biomedical Sciences”.

In view of these references, we note that [6, 15] requires conditions for the successful analysis which ultimately rule out the  $\ell_q$ -norm-type regularization, [9] needs a sparsity assumption, and [4] provides a method based on Bergman iterations and specific shrinkage-procedures, but does not include a convergence analysis. In [12] a regularized nonsmooth Newton technique is proposed which relies on some kind of local smoothing.

Motivated by smoothing methods (see for example [6]), in this work, for solving the problem (1) below, we present a descent algorithm combining a Huber-type smoothing (which we call *Huberization* in the sequel) with elements of the nonsmooth Newton solver from [12]. In fact, the smoothing method provides a mechanism which allows us to drive the Huber-type smoothing of the  $\ell_q$ -norm to zero, thus, genuinely approaching, along a subsequence, a stationary point of the  $\ell_q$ -norm-type problem. From the gradient of the *Huberized* objective, a suitable descent direction for that objective is computed by the so-called *R*-regularized nonsmooth Newton method [12]. By this procedure, a variable-metric-type scaling is applied to the steepest descent direction, thus, improving the convergence behavior of the overall method. Moreover, convergence of the algorithmic scheme towards a stationary point of (1) is established.

The rest of the paper is organized as follows. In Sect. 2, we introduce the nonconvex  $\text{TV}^q$ -model in finite dimension and establish the existence of a solution and the necessary optimality condition. A smoothing descent method and its convergence analysis are presented in Sect. 3. Finally, the proposed method is implemented for various image processing tasks and the results are reported in Sect. 4.

## 2 Nonconvex $\text{TV}^q$ -Model

The nonconvex  $\text{TV}^q$ -model considered in this paper is formulated as follows:

$$\min_{u \in \mathbb{R}^{|\Omega|}} f(u) := \sum_{(i,j) \in \Omega} \left( \frac{\mu}{2} |(\nabla u)_{ij}|^2 + \frac{\alpha}{q} |(\nabla u)_{ij}|^q + \frac{1}{2} |(Ku - z)_{ij}|^2 \right). \quad (1)$$

Here,  $\Omega$  is a two-dimensional discrete image domain with  $|\Omega|$  being the number of pixels in  $\Omega$ ,  $u \in \mathbb{R}^{|\Omega|}$  represents the digital image which is to be reconstructed from observed data  $z \in \mathbb{R}^{|\Omega|}$ ,  $\nabla \in \mathbb{R}^{2|\Omega| \times |\Omega|}$  is the gradient operator,  $|(\nabla u)_{ij}|$  denotes the Euclidean norm of  $(\nabla u)_{ij}$  in  $\mathbb{R}^2$ ,  $K \in \mathbb{R}^{|\Omega| \times |\Omega|}$  is a continuous linear operator (which, for instance, might describe blurring), and  $\alpha > 0$ ,  $0 < q < 1$ ,  $0 < \mu \ll \alpha$  are user-chosen regularization parameters. In particular, if  $q = 1$  and  $\mu = 0$ , then (1) corresponds to the conventional (convex) total variation ( $\text{TV}$ -) model according to Rudin, Osher and Fatemi [18].

Throughout this paper, we shall assume that

$$\text{Ker} \nabla \cap \text{Ker} K = \{0\}. \quad (2)$$

Under the condition (2), the existence of a solution of (1) is guaranteed by the following theorem. Its proof can be found in [12].

**Theorem 1 (Existence of Solution).** *There exists a global minimizer for the variational problem (1).*

Note that due to the presence of the power- $q$  term the objective  $f$  in (1) is not even locally Lipschitz continuous. This causes difficulties when using the Clarke generalized gradient of  $f$  [7] for describing stationarity of a point  $u$ . However, distinguishing smooth and nonsmooth regions of  $f$  through the norm of the gradient of  $u$ , stationary points of (1) can still be characterized by the following Euler-Lagrange-type system:

$$\begin{cases} -\mu\Delta u + K^\top(Ku - z) + \alpha\nabla^\top(|\nabla u|^{q-2}\nabla u) = 0, & \text{if } (\nabla u)_{ij} \neq 0; \\ \nabla u = 0, & \text{otherwise.} \end{cases} \quad (3)$$

The disjunctive nature of the above system, which is due to the nonsmoothness of  $f$ , causes severe difficulties in the design of solution algorithms. In the following, we propose a smoothing descent method, which generates a sequence that has an accumulation point satisfying (3).

### 3 Smoothing Descent Method

The smoothing descent method proceeds iteratively as follows. In each iteration, the TV<sup>q</sup>-objective is smoothed locally around the nondifferentiability by a Huber function which is controlled through the parameter  $\gamma > 0$ . The Huber function reads

$$\varphi_\gamma(s) := \begin{cases} \frac{1}{q}s^q - (\frac{1}{q} - \frac{1}{2})\gamma^q, & \text{if } s > \gamma, \\ \frac{1}{2}\gamma^{q-2}s^2, & \text{if } 0 \leq s \leq \gamma, \end{cases}$$

and the associated Huberized version of (1) is given by

$$\min_{u \in \mathbb{R}^{|\Omega|}} f_\gamma(u) := \sum_{(i,j) \in \Omega} \left( \frac{\mu}{2} |(\nabla u)_{ij}|^2 + \alpha \varphi_\gamma(|(\nabla u)_{ij}|) + \frac{1}{2} |(Ku - z)_{ij}|^2 \right). \quad (4)$$

The corresponding Huberized Euler-Lagrange equation is

$$\nabla f_\gamma(u) = -\mu\Delta u + K^\top(Ku - z) + \alpha\nabla^\top(\max(|\nabla u|, \gamma)^{q-2}\nabla u) = 0. \quad (5)$$

Here the max-operation is understood in a componentwise sense.

Clearly, the Huberized objective  $f_\gamma$  is continuously differentiable and bounded from below. Therefore, (4) can be solved by well-known standard solvers for smooth minimization problems; see, e.g., [17]. In this work, however, we utilize a tailored approach by employing the so-called *R-regularized Newton scheme* which was very recently proposed in [12]. In order to globalize the Newton iteration, a backtracking Armijo line search is used which is particularly tuned to the structure of the problem of interest.

For completeness, we provide a self-contained introduction of the  $R$ -regularized Newton method in the appendix. Essentially, this method is a generalization of the lagged-diffusivity fixed-point iteration [2, 5, 20], which alleviates diffusion nonlinearity by using information from the previous iterate. Moreover, with the aid of the infeasible Newton technique [11], a sufficient condition can be derived for obtaining a descent direction in each iteration; see Theorem 5 in the appendix. The descent property is important for the Armijo-based line search globalization; see [17] and the references therein for a general account of Armijo's line search rule.

When (4) is solved with sufficient accuracy, for instance with respect to the residual norm  $\|f_{\gamma^k}(u^{k+1})\|$ , then the Huber parameter is reduced and the current solution serves as the initial guess for solving the next Huberized problem. The resulting overall algorithmic scheme is specified next.

---

**Algorithm 1.** Smoothing descent method.

---

Choose  $0 < \tau < 1$ ,  $0 < \theta < 1$ ,  $\rho > 0$ ,  $\bar{\gamma} > 0$ . Then iterate as follows:

1. Compute a descent direction  $\delta u^k$  for  $f_{\gamma^k}$  at  $u^k$ , i.e.  $\nabla f_{\gamma^k}(u^k)^\top \delta u^k < 0$ , for the Huberized problem (4) by the  $R$ -regularized Newton method..
2. Perform an Armijo line search, i.e. determine the size  $a^k > 0$  such that

$$f_{\gamma^k}(u^k + a^k \delta u^k) \leq f_{\gamma^k}(u^k) + \tau a^k \nabla f_{\gamma^k}(u^k)^\top \delta u^k, \tag{6}$$

and set  $u^{k+1} := u^k + a^k \delta u^k$ .

3. If  $\|\nabla f_{\gamma^k}(u^{k+1})\| \geq \rho \gamma^k$ , then set  $\gamma^{k+1} := \gamma^k$ ; otherwise, set  $\gamma^{k+1} := \theta \gamma^k$ .
  4. If  $\gamma^k \geq \bar{\gamma}$ , then set  $k := k + 1$  and return to step 1; otherwise stop.
- 

In our experiments, we shall always fix the parameters  $\tau = 0.1$ ,  $\theta = 0.8$ ,  $\rho = 0.1$ .

Next we present the convergence analysis for the smoothing descent method. For this purpose, we take  $\bar{\gamma} = 0$ .

**Lemma 2.** *The sequence generated by Algorithm 3 satisfies*

$$\lim_{k \rightarrow \infty} \gamma^k = 0, \quad \text{and} \quad \liminf_{k \rightarrow \infty} \|\nabla f_{\gamma^k}(u^{k+1})\| = 0.$$

*Proof.* Define the index set

$$\mathcal{K} := \{k : \gamma^{k+1} = \theta \gamma^k\}. \tag{7}$$

If  $\mathcal{K}$  is finite, then there exists some  $\bar{k}$  such that for all  $k > \bar{k}$  we have  $\gamma^k = \gamma^{\bar{k}}$  and  $\|\nabla f_{\gamma^k}(u^{k+1})\| \geq \rho \gamma^{\bar{k}}$ . This contradicts the fact that the first-order method with Armijo line search for solving the smooth problem  $\min_u f_{\gamma^{\bar{k}}}(u)$  generates a sequence  $(u^k)$  such that  $\liminf_{k \rightarrow \infty} \|\nabla f_{\gamma^{\bar{k}}}(u^k)\| = 0$ , cf. [10]. Thus,  $\mathcal{K}$  is infinite and  $\lim_{k \rightarrow \infty} \gamma^k = 0$ . Moreover, let  $\mathcal{K} = (k^l)_{l=1}^\infty$  with  $k^1 < k^2 < \dots$ , then we have  $\|\nabla f_{\gamma^{k^l}}(u^{k^l+1})\| \leq \rho \gamma^{k^l} \rightarrow 0$  as  $l \rightarrow \infty$ . Hence,  $\liminf_{k \rightarrow \infty} \|\nabla f_{\gamma^k}(u^{k+1})\| = 0$ .

**Theorem 3.** *Assume that the sequence  $(u^k)$  generated by Algorithm 3 is bounded. Then there exists an accumulation point  $u^*$  of  $(u^k)$  such that  $u^* \in \mathbb{R}^{|\Omega|}$  satisfies the Euler-Lagrange Eq. (3).*

*Proof.* In view of the result in Lemma 2, there exists a subsequence of  $(u^k)$ , say  $(u^{k'})$ , such that  $\lim_{k' \rightarrow \infty} \|\nabla f_{\gamma^{k'-1}}(u^{k'})\| = 0$ . Since  $(u^{k'})$  is bounded, by compactness there exists a subsequence of  $(u^{k'})$ , say  $(u^{k''})$ , such that  $(u^{k''})$  converges to some  $u^*$  as  $k'' \rightarrow \infty$ . We show that  $u^*$  is a solution to (3). On the set  $\{(i, j) \in \Omega : (\nabla u^*)_{ij} = 0\}$ , the conclusion follows automatically. On the set  $\{(i, j) \in \Omega : (\nabla u^*)_{ij} \neq 0\}$ , we have that  $\max(|\nabla u^{k''}|, \gamma^{k''-1}) \rightarrow |\nabla u^*| > 0$  as  $k'' \rightarrow \infty$ . Therefore, it follows from

$$|\nabla f(u^*)| \leq |\nabla f_{\gamma^{k''-1}}(u^{k''}) - \nabla f(u^*)| + |\nabla f_{\gamma^{k''-1}}(u^{k''})| \rightarrow 0, \quad (8)$$

that  $u^*$  satisfies (3).

## 4 Numerical Experiments

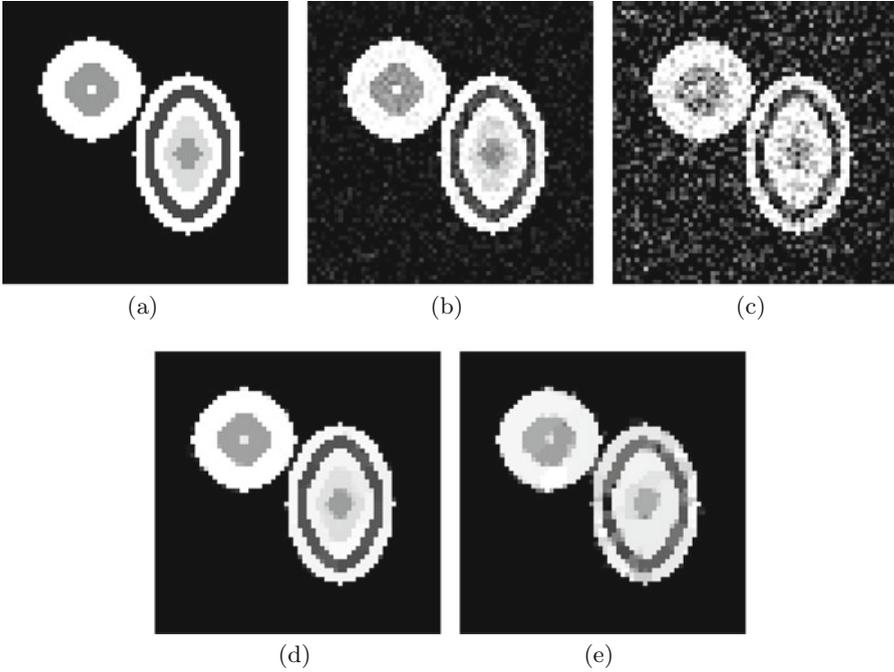
In this section, we report on numerical results obtained by our algorithm for various tasks in TV<sup>q</sup>-model based image restoration. The experiments are performed under MATLAB R2009b on a 2.66 GHz Intel Core Laptop with 4 GB RAM.

### 4.1 Denoising

We first test our algorithm on denoising the “Two Circles” image; see Fig. 1. This image, shown in plot (a), is degraded by zero-mean white Gaussian noise of 7% and 14% standard deviation respectively; see plots (b) and (c). The parameters  $q = 0.75$ ,  $\mu = 0$  are used in this example. The restored images of two different noisy images are given in plots (d) and (e). In the following, we use the data set in (b) to investigate the numerical behavior of Algorithm 3 in details.

**Robustness to Initialization.** Note that our algorithm is intended to find a stationary point of the TV<sup>q</sup>-model (which is often a local minimizer in our numerical experiments). It is worthwhile to check the quality of such local solutions. In Fig. 2, we implement the algorithm starting from three different choices of initializations; see the first row. The corresponding restored images are shown in the second row, which are visually indistinguishable. The energy values of the restorations in (e), (f), (g), (h) are equal to 29.4488, 29.4499, 29.4497, 29.4594, respectively, which also indicates that the restorations have small differences in quality. We remark that choosing a relatively large initial Huber parameter  $\gamma^0$  in general strengthens the robustness of the algorithm against poor initializations.

**Choice of Stopping Criteria.** As the stopping criteria (step 4) of Algorithm 3 depends on  $\bar{\gamma}$ , here we suggest an empirical way based on the *histogram* for choosing a proper  $\bar{\gamma}$  in an a posteriori way. As we know, the TV<sup>q</sup>-model tends



**Fig. 1.** Denoising: (a) “Two Circles” image. (b) Corrupted with 7% Gaussian noise. (c) Corrupted with 14% Gaussian noise. (d) Restoration of (b) with  $\alpha = 0.05$ . (e) Restoration of (c) with  $\alpha = 0.12$ .

to promote a solution with very sparse histogram. This is numerically confirmed by the histogram plots in Fig. 3. Therefore, it is reasonable to terminate the algorithm once there is no longer significant change in the sparsity pattern of the histogram. In our particular example, this suggests that  $\bar{\gamma} = 10^{-4}$  is a proper choice. The same choice of  $\bar{\gamma}$  will be used in all the following experiments.

**Miscellaneous Numerical Behavior.** We further demonstrate the numerical behavior of the algorithm in Fig. 4. All data points in the plots are taken from those iterations with  $k \in \mathcal{K}$ ; see (7) for the definition of  $\mathcal{K}$ . As  $\gamma^k$  decreases, our algorithm is well behaved in terms of objective value, PSNR, and residual norm. Qualitatively a similar numerical behavior is observed in the experiments that follow.

## 4.2 Deblurring

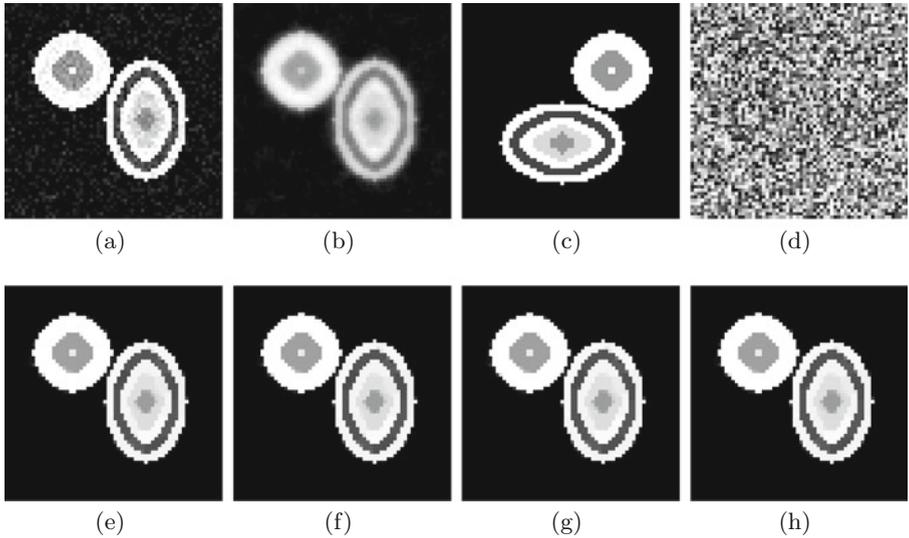
In Fig. 5 we test our algorithm in the context of deblurring the 256-by-256 phantom image depicted in plot (a). The original image is blurred by a two-dimensional truncated Gaussian kernel yielding

$$(Ku)_{ij} = \sum_{|i'|\leq 3, |j'|\leq 3} \exp\left(-\frac{|i'|^2 + |j'|^2}{2|\sigma_K|^2}\right) u_{i-i', j-j'}.$$

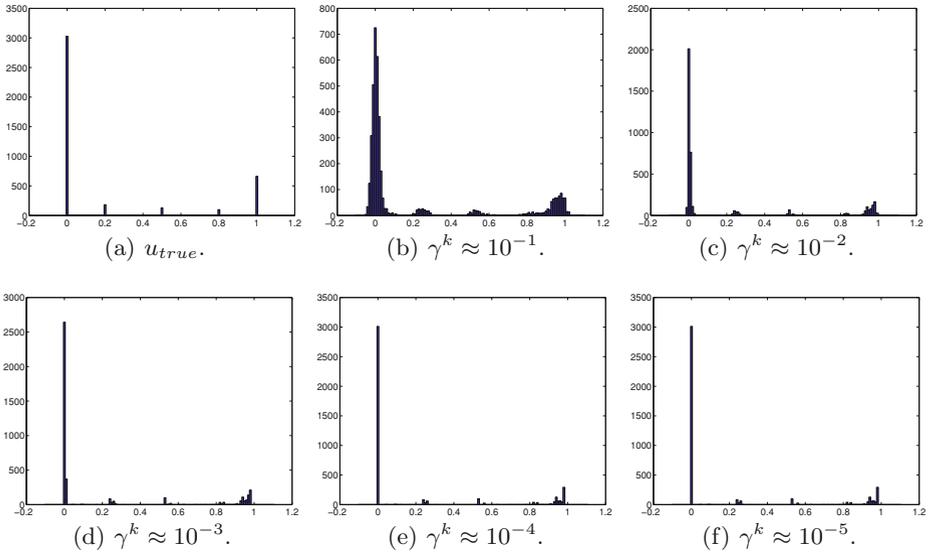
Then white Gaussian noise of zero mean and 0.05 standard deviation is added to the blurred image; see (b). We apply our algorithm with  $q = 0.75$ ,  $\alpha = 0.01$ ,  $\mu = 10^{-6}$ , and  $u^0 = z$ . In plot (c) the restored image is shown. Its PSNR-value is 27.6672.

### 4.3 Compressed Sensing

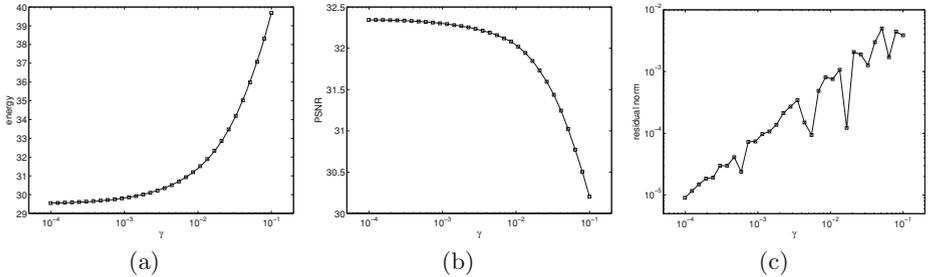
We also apply our algorithm to a  $k$ -space compressed sensing problem; see Fig. 6. The observed data  $z$  is constructed as follows:  $z = SFu_{true}$ , where  $F$  is the 2D discrete Fourier transform and  $S$  is a 20%  $k$ -space random sampling matrix. We reconstruct the image by solving our  $TV^q$ -model with  $q = 0.75$ ,  $\alpha = 0.001$ ,  $\mu = 10^{-6}$ , and  $u^0 = 0$ . The corresponding restored image is shown in (e). This result is compared with the reconstruction obtained from the inverse Fourier transform in (c) and the reconstruction obtained from the TV-model in (d) with  $q = 1$ ,  $\alpha = 0.02$ ,  $\mu = 10^{-6}$ . In our implementation of the TV-model,  $\alpha$  is chosen in order to obtain an image with optimal PSNR. The (convex) TV-model, here as well as in Sect. 4.4, is solved by a primal-dual Newton method [11] with Huber parameter  $\bar{\gamma}$ . We remark that many other algorithms in the literature may also work efficiently in the same context. In the left part of Table 1, we provide the comparison of the three candidate methods with respect to PSNR and CPU time. It is observed



**Fig. 2.** Initialization test: (a) Observed image as initial guess. (b) Tikhonov regularized image as initial guess. (c) Rotated image as initial guess (d) Random initial guess. (e), (f), (g), and (h) are the restorations from (a), (b), (c), and (d) respectively.



**Fig. 3.** Histogram of  $u^{k+1}$  ( $k \in \mathcal{K}$ ).

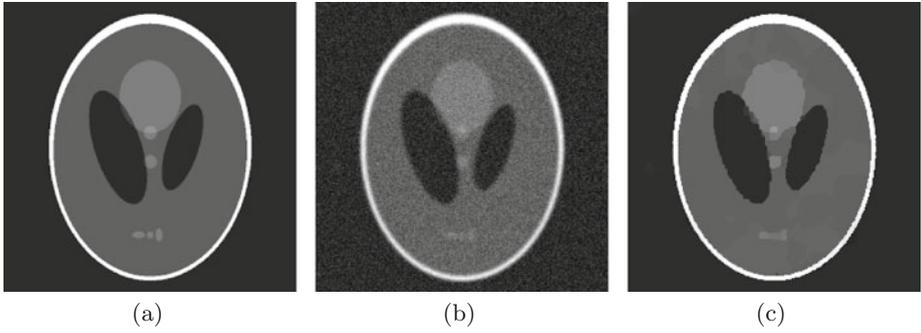


**Fig. 4.** Numerical behavior ( $k \in \mathcal{K}$ ): (a)  $\text{TV}^q$ -energy  $f(u^{k+1})$ . (b)  $\text{PSNR}(u^{k+1})$ . (c) Residual norm  $\|\nabla f_{\gamma^k}(u^{k+1})\|$ .

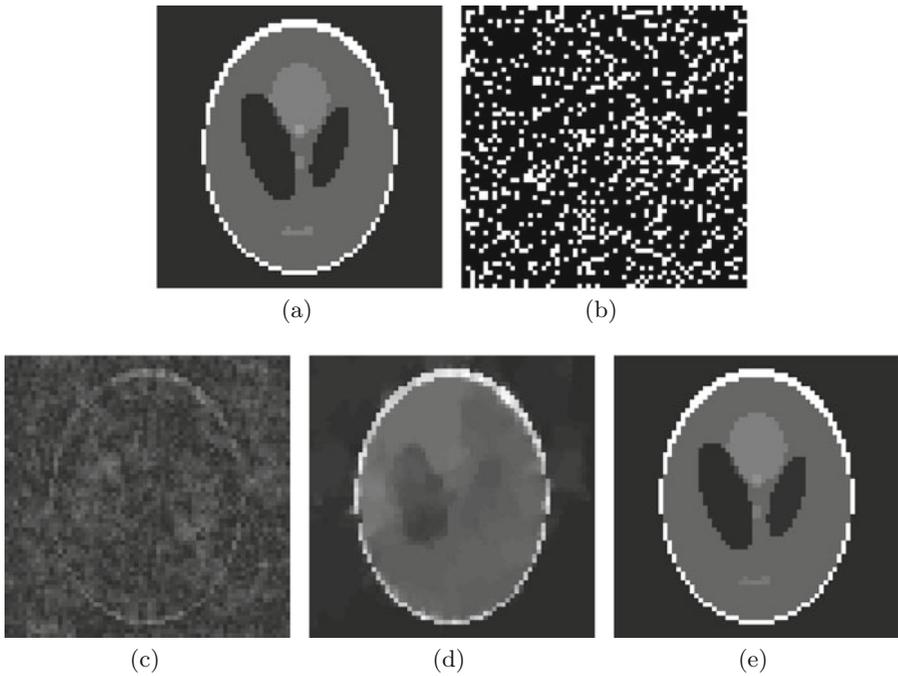
that the inverse Fourier transform is computationally cheap but only yields a poor result. The TV method takes about 6 seconds but still cannot recover the image to fine details. Our  $\text{TV}^q$  method takes about double the CPU time of TV and provides an almost perfect reconstruction.

#### 4.4 Integral-Geometry Tomography

In Fig. 7, we apply our algorithm to integral-geometry tomography. The given data  $z$  in (b), also known as the *sinogram*, is constructed by taking the 2D Radon transform of the underlying image every 15 degrees (out of 180 degrees). The matrix  $K$  in this example is a discrete Radon transform of size 1235-by-4096.

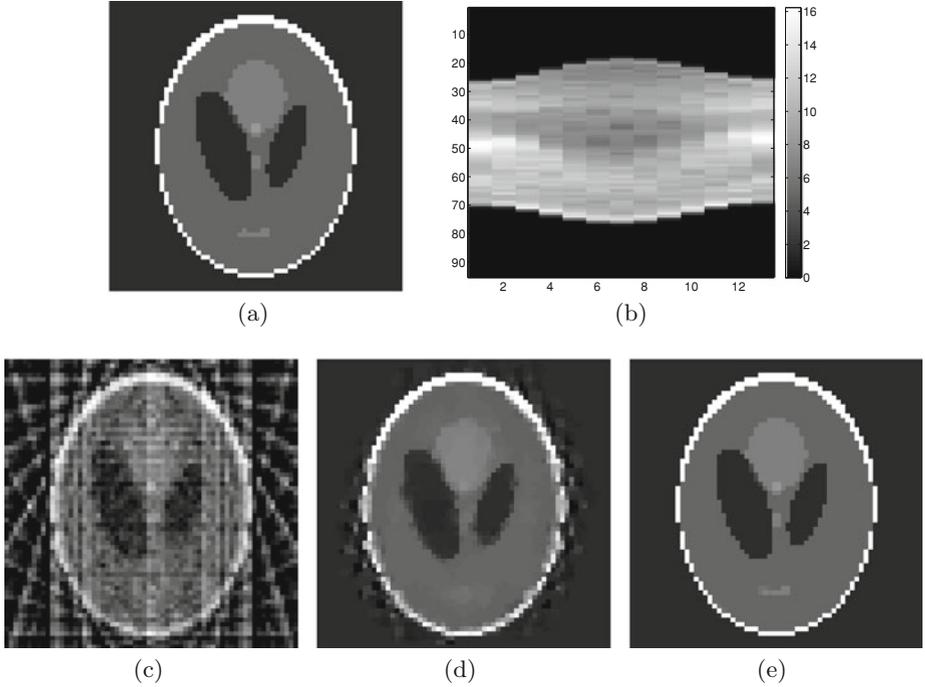


**Fig. 5.** Deblurring: (a) 256-by-256 Phantom. (b) Noisy blurred image; PSNR=21.7276. (c) Restored image; PSNR=27.6672.



**Fig. 6.** Compressed sensing: (a) 64-by-64 Phantom. (b) 20%  $k$ -space random sampling. (c) Direct reconstruction by FFT. (d) Reconstruction by TV-model. (e) Reconstruction by  $TV^q$ -model.

We utilize our  $TV^q$ -model with  $q = 0.75$ ,  $\alpha = 0.001$ ,  $\mu = 10^{-6}$ , and  $u^0 = 0$ . The restoration is shown in plot (e). This result clearly is superior to the one shown in plot (c), which is obtained by filtered backprojection, and the one shown in plot (d), which is obtained from the TV-model with  $q = 1$ ,  $\alpha = 0.02$ ,



**Fig. 7.** Integral-geometry tomography: (a) 64-by-64 Phantom. (b) Sinogram. (c) Reconstruction by filtered backprojection. (d) Reconstruction by TV-model. (e) Reconstruction by  $TV^q$ -model.

$\mu = 10^{-6}$ . In our implementation of the TV-model,  $\alpha$  is chosen in order to obtain an image with optimal PSNR. In the right part of Table 1, we again compare the three candidate methods with respect to PSNR and CPU time. Similar to the compression sensing example, the  $TV^q$  method costs more CPU time than the other two methods (but still less than double the CPU time of the TV method) but yields an almost perfect reconstruction.

**4.5 Reconstruction of Multi-coil MRI Raw Data**

We now extend the methodology to magnetic resonance imaging (MRI), by considering the following model:

$$\min_{u \in \mathbb{R}^{|\Omega|}} \frac{1}{2} \sum_{l=1}^L \|K_l u - z_l\|^2 + \sum_{(i,j) \in \Omega} \frac{\alpha_g}{q} |(\nabla u)_{ij}|^q + \sum_{(i',j') \in \Xi} \frac{\alpha_w}{q} |(Wu)_{i'j'}|^q, \quad (9)$$

with  $K_l := PF(\sigma_l u)$ . Here  $\alpha_g$  and  $\alpha_w$  are two positive parameters,  $L$  is the number of coils of the MRI machine,  $(z_l)$  denote the raw data collected by each coil,  $(\sigma_l)$  are the given (or precomputed) coil sensitivities,  $F$  is the two-dimensional

**Table 1.** Comparison with respect to PSNR and CPU time (in seconds).

	Compressed sensing			Integral-geometry tomography		
	IFT	TV	$TV^q$	FBP	TV	$TV^q$
PSNR	13.68	25.74	44.48	PSNR	15.14	23.18
CPU	$\ll 1$	6.68	12.72	CPU	$\ll 1$	3.82

discrete Fourier transform, and  $P$  represents some given radial projection operator in the k-space. Moreover,  $W : \mathbb{R}^{|\Omega|} \rightarrow \mathbb{R}^{|\Xi|}$  is a user-chosen transform, typically a 2D discrete wavelet transform, and  $\Xi$  denotes the wavelet transformed domain.

Note that in addition to the  $TV^q$  regularization, we include the  $\ell^q$ -norm of wavelet coefficients in the regularization in order to allow the reconstruction to be richer than patchwise constant images. Nevertheless, our algorithm presented in this paper can be extended without agonizing pain to the problem (9).

Indeed, as a straightforward extension of Theorem 1, the solution of (9) exists provided that  $\text{Ker} \nabla \cap \text{Ker} W \cap (\cap_{l=1}^L \text{Ker} K_l) = \{0\}$ . The Euler-Lagrange equation for (9) appears as

$$\begin{cases} \alpha_g \nabla^\top p_g + \alpha_w W^\top p_w + \sum_{l=1}^L K_l^\top (K_l u - z_l) = 0, \\ (p_g)_{ij} = |(\nabla u)_{ij}|^{q-2} (\nabla u)_{ij}, & \text{if } (i, j) \in \Omega \wedge (\nabla u)_{ij} \neq 0, \\ (p_w)_{i'j'} = |(Wu)_{(i',j')}|^{q-2} (Wu)_{i'j'}, & \text{if } (i', j') \in \Xi \wedge (Wu)_{i'j'} \neq 0. \end{cases}$$

The associated Huberized problem can be analogously formulated as

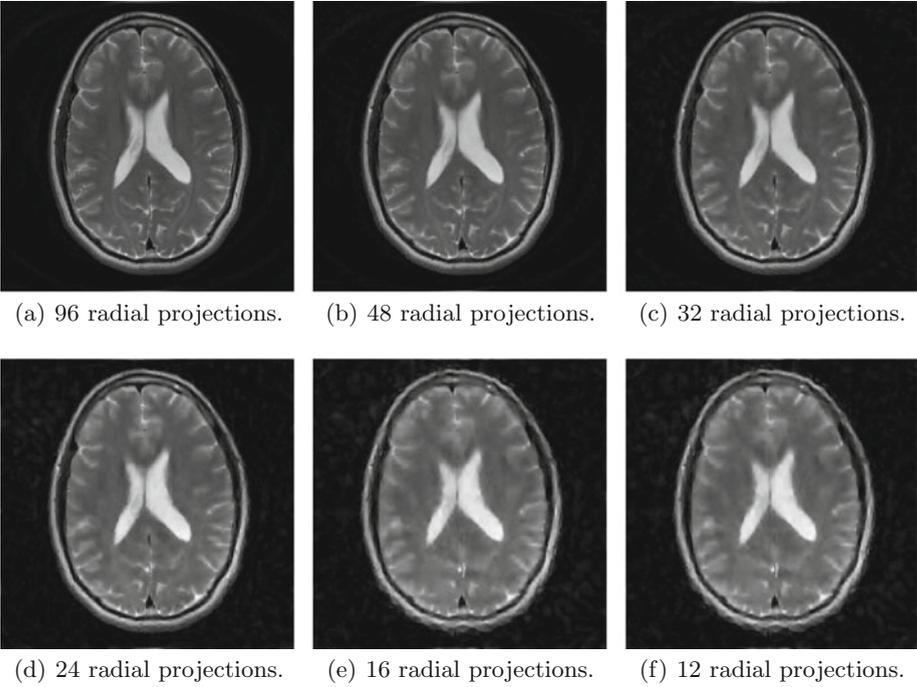
$$\min_{u \in \mathbb{R}^{|\Omega|}} \frac{1}{2} \sum_{l=1}^L \|K_l u - z_l\|^2 + \alpha_g \sum_{(i,j) \in \Omega} \varphi_\gamma(|(\nabla u)_{ij}|) + \alpha_w \sum_{(i',j') \in \Xi} \varphi_\gamma(|(Wu)_{i'j'}|),$$

and the corresponding Huberized Euler-Lagrange equation is given by

$$\alpha_g \nabla^\top (\max(|\nabla u|, \gamma)^{q-2} \nabla u) + \alpha_w W^\top (\max(|Wu|, \gamma)^{q-2} Wu) + \sum_{l=1}^L K_l^\top (K_l u - z_l) = 0.$$

We shall not go into further details but remark that the  $R$ -regularized method in the appendix can be used to solve the above smooth problem by treating the gradient term and the wavelet term independently. Thus, Algorithm 3 can be implemented.

In this experiment, the MRI data are collected from four coils, i.e.  $L = 4$ . We choose  $q = 0.75$ ,  $\alpha_g = 10^{-5}$ ,  $\alpha_w = 2 \times 10^{-5}$ , and  $W$  to be the 4th-order Daubechies wavelet transform [8]. The reconstructed images using various numbers of radial projections are shown in Fig. 8. Depending on the resolution (or detail) desired by practitioners, our method helps to reduce the necessary number of k-space samples and therefore to speed up the overall MRI data acquisition.



**Fig. 8.** Reconstruction of four-coil MRI raw data.

### Appendix: $R$ -Regularized Newton Method

Here we provide a brief and self-contained description of the  $R$ -regularized Newton method. The interested readers can find more details in the recent work [12].

A regularized-Newton-type structure generically arises in the classical lagged-diffusivity fixed-point iteration [20]. Let  $u^k$  be our current iterate in solving the Huberized problem (4). By introducing a lagged-diffusivity weight

$$w^k := \max(|\nabla u^k|, \gamma)^{q+r-2},$$

and a dual variable

$$p := w^k \max(|\nabla u|, \gamma)^{-r} \nabla u,$$

with  $0 \leq r \leq 2 - q$ , the reweighted Euler-Lagrange Eq. (5) appears as

$$\begin{cases} -\mu \Delta u + K^\top (Ku - z) + \alpha \nabla^\top p = 0, \\ (w^k)^{-1} \max(|\nabla u|, \gamma)^r p = \nabla u. \end{cases} \quad (10)$$

Given a current approximation  $(u^k, p^k)$ , we apply a generalized linearization to (10) and obtain the generalized Newton system

$$\begin{bmatrix} -\mu \Delta + K^\top K & \alpha \nabla^\top, \\ -\tilde{C}^k(r) \nabla & D((m^k)^{2-q} e) \end{bmatrix} \begin{bmatrix} \delta u^k \\ \delta p^k \end{bmatrix} = \begin{bmatrix} \mu \Delta u^k - K^\top (Ku^k - z) - \alpha \nabla^\top p^k \\ \nabla u^k - (m^k)^{2-q} p^k \end{bmatrix},$$

where

$$\begin{aligned} m^k &:= \max(|\nabla u^k|, \gamma), \\ (\chi_{\mathcal{A}^k})_{ij} &:= \begin{cases} 1, & \text{if } |(\nabla u^k)_{ij}| > \gamma, \\ 0, & \text{otherwise,} \end{cases} \\ \tilde{C}^k(r) &:= I - rD(\chi_{\mathcal{A}^k}(m^k)^{-q}p^k) \begin{bmatrix} D(\nabla_x u^k) & D(\nabla_y u^k) \\ D(\nabla_x u^k) & D(\nabla_y u^k) \end{bmatrix}. \end{aligned}$$

Here  $D(v)$  denotes a diagonal matrix with the vector  $v$  as diagonal entries. For  $v \in \mathbb{R}^{|\Omega|}$  and  $p = (p^1, p^2) \in \mathbb{R}^{2|\Omega|}$ , the notation  $vp$  is understood as a vector in  $\mathbb{R}^{2|\Omega|}$  such that  $(vp)_{ij} = (v_{ij}p_{ij}^1, v_{ij}p_{ij}^2)$  for all  $(i, j) \in \Omega$ .

After eliminating  $\delta p^k$ , we are left with the linear system

$$\tilde{H}^k(r)\delta u^k = -\nabla f_\gamma(u^k), \quad (11)$$

where

$$\tilde{H}^k(r) := -\mu\Delta + K^\top K + \alpha\nabla^\top D((m^k)^{q-2}e)\tilde{C}^k(r)\nabla.$$

The Newton system (11) can be rewritten as

$$(H^k + \beta R^k)\delta u^k = -\nabla f_\gamma(u^k), \quad (12)$$

with  $\beta = 2 - q - r$ , where  $H^k := \tilde{H}^k(2 - q)$  is the Hessian from the non-reweighted primal-dual Newton method [11, 20], and

$$R^k := \alpha\nabla^\top D(\chi_{\mathcal{A}^k}(m^k)^{-2}p^k) \begin{bmatrix} D(\nabla_x u^k) & D(\nabla_y u^k) \\ D(\nabla_x u^k) & D(\nabla_y u^k) \end{bmatrix} \nabla.$$

serves as a regularizer on the Hessian  $H^k$ . This coins the name *R-regularization* in connection with Newton's method.

Next we aim to establish a sufficient condition on the *R-regularization* weight  $\beta$  in order to guarantee that  $\tilde{H}^k(r)$  is positive definite and, therefore,  $\delta u^k$  is a descent direction for  $f_\gamma$  at  $u^k$ . For this purpose we invoke an *infeasible Newton technique* [11, 12].

The infeasible Newton technique involves two modifications in constructing the system matrix  $\tilde{H}^k(r)$ . First, we replace  $p^k$  by  $p_+^k$ , where

$$p_+^k := \frac{\chi_{\mathcal{A}^k}(m^k)^{q-1}p^k}{\max((m^k)^{q-1}, |p^k|)} + (1 - \chi_{\mathcal{A}^k})p^k.$$

Note that the modified  $p_+^k$  satisfies its feasibility condition, i.e.

$$|(p_+^k)_{ij}| \leq |(\nabla u^k)_{ij}|^{q-1}, \quad \text{whenever } |(\nabla u^k)_{ij}| > \gamma. \quad (13)$$

Secondly, we replace  $\tilde{C}^k(r)$  by its symmetrization denoted by  $\tilde{C}_+^k(r)$ , i.e.

$$\begin{aligned} \tilde{C}_+^k(r) &:= \frac{\tilde{C}^k(r) + \tilde{C}^k(r)^\top}{2} = I - rD(\chi_{\mathcal{A}^k}(m^k)^{-q}) \\ &\cdot \begin{bmatrix} D((p_+^k)_x \nabla_x u^k) & D(\frac{1}{2}((p_+^k)_x \nabla_y u^k + (p_+^k)_y \nabla_x u^k)) \\ D(\frac{1}{2}((p_+^k)_x \nabla_y u^k + (p_+^k)_y \nabla_x u^k)) & D((p_+^k)_y \nabla_y u^k) \end{bmatrix}. \end{aligned}$$

Accordingly, the system matrix  $H^k$  in (12) is replaced by  $H_+^k$  with

$$H_+^k := -\mu\Delta + K^\top K + \alpha\nabla^\top D((m^k)^{q-2}e)\tilde{C}_+^k(2-q)\nabla,$$

and the regularizer  $R^k$  is replaced by  $R_+^k$  with

$$R_+^k := \alpha\nabla^\top D(\chi_{\mathcal{A}^k}(m^k)^{-2}) \cdot \begin{bmatrix} D((p_+^k)_x \nabla_x u^k) & D(\frac{1}{2}((p_+^k)_x \nabla_y u^k + (p_+^k)_y \nabla_x u^k)) \\ D(\frac{1}{2}((p_+^k)_x \nabla_y u^k + (p_+^k)_y \nabla_x u^k)) & D((p_+^k)_y \nabla_y u^k) \end{bmatrix} \nabla.$$

**Lemma 4.** *Let  $0 \leq r \leq 1$  (or equivalently  $1 - q \leq \beta \leq 2 - q$ ) and the feasibility condition (13) hold true. Then the matrix  $\tilde{C}_+^k(r)$  is positive semidefinite.*

The proof of Lemma 4 is given in [12]. Thus, the positive definiteness of the  $R$ -regularized Hessian  $H_+^k + \beta R_+^k$  follows immediately from its structure and Lemma 4.

**Theorem 5.** *Suppose the assumptions of Lemma 4 are satisfied. Then the  $R$ -regularized Hessian  $H_+^k + \beta R_+^k$  is positive definite.*

We remark that our choice of  $\mu$  is related to Theorem 6. Under the condition (2), for any positive  $\mu$  the matrix  $-\mu\Delta + K^\top K$  is positive definite, and therefore Theorem 6 follows. However, whenever  $K$  is injective the same conclusion holds with  $\mu = 0$ . This is why we are allowed to choose  $\mu = 0$  in the denoising example in Sect. 4.1.

Given the result in Theorem 5, the descent direction  $\delta u^k$  in the  $R$ -regularized Newton method implemented in step 1 of Algorithm 3 can be now obtained by solving the linear system

$$(H_+^k + \beta R_+^k)\delta u^k = -\nabla f_\gamma(u^k),$$

with  $1 - q \leq \beta \leq 2 - q$ . Given  $\delta u^k$ , one can compute  $\delta p^k$  as

$$\delta p^k = (m^k)^{q-2}(\nabla u^k + \tilde{C}_+^k(r)\nabla\delta u^k) - p^k,$$

and then update  $u^{k+1} := u^k + a^k\delta u^k$  and  $p^{k+1} := p^k + a^k\delta p^k$  with some step size  $a^k$  determined by the Armijo line search as in step 2 of Algorithm 3.

In all experiments in Sect. 4, we consistently choose  $\beta = 1 - q$ . This choice has the interesting interpretation that we actually relax the  $\text{TV}^q$ -model to a weighted  $\text{TV}$ -model (with weight updating in the outer iterations); see [12].

**Acknowledgement.** The authors would like to thank Dr. Florian Knoll for contributing his data and codes to our experiments on MRI.

## References

1. Burke, J.V., Lewis, A.S., Overton, M.L.: A robust gradient sampling algorithm for nonsmooth, nonconvex optimization. *SIAM J. Optim.* **15**, 751–779 (2005)

2. Chan, T.F., Mulet, P.: On the convergence of the lagged diffusivity fixed point method in total variation image restoration. *SIAM J. Numer. Anal.* **36**, 354–367 (1999)
3. Chartrand, R.: Exact reconstruction of sparse signals via nonconvex minimization. *IEEE Signal Process. Lett.* **14**, 707–710 (2007)
4. Chartrand, R.: Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data. In: *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 262–265 (2009)
5. Chartrand, R., Yin, W.: Iteratively reweighted algorithms for compressive sensing. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3869–3872 (2008)
6. Chen, X., Zhou, W.: Smoothing nonlinear conjugate gradient method for image restoration using nonsmooth nonconvex minimization. *SIAM J. Imaging Sci.* **3**, 765–790 (2010)
7. Clarke, F.H.: *Optimization and Nonsmooth Analysis*. John Wiley & Sons, New York (1983)
8. Daubechies, I.: *Ten Lectures on Wavelets*. SIAM, Philadelphia (1992)
9. Daubechies, I., DeVore, R., Fornasier, M., Güntürk, C.: Iteratively reweighted least squares minimization for sparse recovery. *Comm. Pure Appl. Math.* **63**, 1–38 (2010)
10. Grippo, L., Lampariello, F., Lucidi, S.: A nonmonotone line search technique for Newton’s method. *SIAM J. Numer. Anal.* **23**, 707–716 (1986)
11. Hintermüller, M., Stadler, G.: An infeasible primal-dual algorithm for total bounded variation-based inf-convolution-type image restoration. *SIAM J. Sci. Comput.* **28**, 1–23 (2006)
12. Hintermüller, M., Wu, T.: Nonconvex  $TV^q$ -models in image restoration: analysis and a trust-region regularization based superlinearly convergent solver. *SIAM J. Imaging Sciences* (to appear)
13. Kiwiel, K.C.: Convergence of the gradient sampling algorithm for nonsmooth nonconvex optimization. *SIAM J. Optim.* **18**, 379–388 (2007)
14. Mäkelä, M., Neittaanmäki, P.: *Nonsmooth Optimization: Analysis and Algorithms with Applications to Optimal Control*. World Scientific, River Edge (1992)
15. Nikolova, M., Ng, M.K., Tam, C.-P.: Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction. *IEEE Trans. Image Process.* **19**, 3073–3088 (2010)
16. Nikolova, M., Ng, M.K., Zhang, S., Ching, W.-K.: Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization. *SIAM J. Imaging Sci.* **1**, 2–25 (2008)
17. Nocedal, J., Wright, S.: *Numerical Optimization*, 2nd edn. Springer, New York (2006)
18. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* **60**, 259–268 (1992)
19. Schramm, H., Zowe, J.: A version of the bundle idea for minimizing a nonsmooth function: Conceptual idea, convergence analysis, numerical results. *SIAM J. Optim.* **2**, 121–152 (1992)
20. Vogel, C.R., Oman, M.E.: Iterative methods for total variation denoising. *SIAM J. Sci. Comput.* **17**, 227–238 (1996)