

A TVSCAD approach for image deblurring with impulsive noise

Guoyong Gu¹, Suhong Jiang² and Junfeng Yang^{1,3}

¹ Department of Mathematics, Nanjing University, People's Republic of China

² School of Accounting, Nanjing University of Finance & Economics, People's Republic of China

E-mail: jfyang@nju.edu.cn

Received 18 March 2017, revised 18 July 2017

Accepted for publication 15 October 2017

Published 10 November 2017



CrossMark

Abstract

We consider image deblurring problem in the presence of impulsive noise. It is known that *total variation* (TV) regularization with L1-norm penalized data fitting (TVL1 for short) works reasonably well only when the level of impulsive noise is relatively low. For high level impulsive noise, TVL1 works poorly. The reason is that all data, both corrupted and noise free, are equally penalized in data fitting, leading to insurmountable difficulty in balancing regularization and data fitting. In this paper, we propose to combine TV regularization with nonconvex *smoothly clipped absolute deviation* (SCAD) penalty for data fitting (TVSCAD for short). Our motivation is simply that data fitting should be enforced only when an observed data is not severely corrupted, while for those data more likely to be severely corrupted, less or even null penalization should be enforced. A *difference of convex functions* algorithm is adopted to solve the nonconvex TVSCAD model, resulting in solving a sequence of TVL1-equivalent problems, each of which can then be solved efficiently by the alternating direction method of multipliers. Theoretically, we establish global convergence to a critical point of the nonconvex objective function. The R-linear and at-least-sublinear convergence rate results are derived for the cases of anisotropic and isotropic TV, respectively. Numerically, experimental results are given to show that the TVSCAD approach improves those of the TVL1 significantly, especially for cases with high level impulsive noise, and is comparable with the recently proposed iteratively corrected TVL1 method (Bai *et al* 2016 *Inverse Problems* 32 085004).

Keywords: total variation, SCAD, deblurring, impulsive noise, DC programming, KL function, alternating direction method of multipliers

(Some figures may appear in colour only in the online journal)

³ Author to whom any correspondence should be addressed.

1. Introduction

We consider the problem of recovering an image degraded by blur and impulsive noise. For simplicity, we focus solely on grayscale images with square domain, and by using the multi-channel *total variation* (TV) regularization introduced in [56] our discussions extend, without essential difficulties, to multichannel images with either square or rectangle domains. Let \bar{u} be a clean image of size n -by- n . By stagnating the columns of \bar{u} in a left-upper and right-lower manner, \bar{u} can be treated equally as a vector in \mathbb{R}^{n^2} . Without loss of generality, we assume that the pixel values of \bar{u} are scaled into $[0, 1]$. The observed blurry and noisy image $f \in \mathbb{R}^{n^2}$ obeys $f = \mathbb{N}_{\text{imp}}(K\bar{u})$, where $K \in \mathbb{R}^{n^2 \times n^2}$ is a blurring operator corresponding to an underlying point spread function, and $\mathbb{N}_{\text{imp}}(\cdot)$ represents a procedure of impulsive noise corruption. Our aim is to approximately recover the clean image \bar{u} from the blurry and noisy observation f , with the blurring operator K given. Apparently, this problem falls into the class of linear inverse problems.

In the rest of this section, we first review briefly some impulsive noise removal methods, either with or without the degradation of blur, and then summarize the notation and the organization of this paper.

1.1. Impulsive noise and filter methods

Impulsive noise is often generated by malfunctioning pixels in camera sensors, faulty memory locations in hardware, or erroneous transmission, see [9]. Two common types of such noise are *salt-and-pepper* (SP) noise and *random-valued* (RV) noise, both of which degrade an image by changing the values of a fraction of randomly selected pixels while leaving the rest ones untouched. When degraded by SP noise, the value of the selected pixel will be replaced with half probability by either the minimum or the maximum pixel value. RV noise degrades images in a similar way, except that intensities of the corrupted pixels are uniformly distributed between the minimum and the maximum pixel values. They are referred to as impulsive noise because corrupted pixels are mostly distinguishable from their neighbors, and intuitively RV noise is harder to remove than SP noise. Based on these features, a number of digital filter methods have been proposed [4], which first detect likely corrupted data entries and then replace them by using the filters. Among others, the median type filters are very popular due to their favorable denoising power and high computational efficiency, e.g. the adaptive median filter [34], the multistate median filter [16], and the median filter based on homogeneity information [24, 41]. We mention that most of the aforementioned filters are merely suitable for denoising in the absence of blur.

1.2. Variational methods

An important class of methods for image deblurring is the variational approach, in which the ground truth \bar{u} is recovered approximately as the minimizer of a certain energy function. It is not uncommon that the energy function is composed of a data fitting term $\Phi(u)$ and a regularization term $\Psi(u)$, i.e. an optimization problem of the form $\min_u \Psi(u) + \mu\Phi(u)$ is solved to recover \bar{u} . Here $\mu > 0$ is a weighting parameter balancing the two terms in minimization.

The purpose of the regularization term $\Psi(u)$ is to enforce certain regularity conditions or prior constraints on the image, such as smoothness and boundedness. Indeed, image deblurring problem is very ill-posed and, as a consequence, the regularization term is indispensable in stabilizing the recovery procedure. The traditional Tikhonov regularization [52] has been

widely used due to its simplicity. However, Tikhonov regularized models, though relatively easy to solve, tend to produce overly smoothed images and are unable to preserve important image attributes such as sharp boundaries. In contrast, the TV regularization pioneered by Rudin *et al* [45] makes image edges and object boundaries, which are generally the most important features of an image, very well preserved due to the linear penalty on the image gradient, see, e.g. [1, 18], where the superiority of TV over Tikhonov-like regularization in recovering image blocky structures was analyzed. Using finite difference operations, one can discretize the TV into different forms, e.g. the well known isotropic and anisotropic discretizations of TV are given by

$$\text{TV}_{\text{iso}}(u) = \sum_{i=1}^{n^2} \|D_i u\|_2 \quad \text{and} \quad \text{TV}_{\text{aniso}}(u) = \sum_{i=1}^{n^2} \|D_i u\|_1, \quad (1.1)$$

respectively, where $D_i \in \mathbb{R}^{2 \times n^2}$ is a local finite difference operator at the i th pixel (boundary conditions will be specified later). Exactly because of the attractive edge-preserving ability, TV regularization has been extremely widely used, see, e.g. a recent review paper up to 2013 [11] and the references therein.

On the other hand, the data fitting term penalizes the deviation of the observed data from the physical model. In the case of additive noise, i.e. $f = Ku + \omega$ for some error $\omega \in \mathbb{R}^n$, or impulsive noise as is the concern of this paper, $\Phi(u)$ usually takes the form $\|Ku - f\|_p^p$ for $p = 1$ or 2 . For additive Gaussian noise, p is usually set to 2 because minimizing $\|Ku - f\|_2^2$ corresponds to seeking the maximum likelihood estimation of \bar{u} . Combined with TV regularization, this leads to the influential TVL2 model [45]. However, practical systems usually suffer from outliers, where only a fraction of data entries are corrupted by noise of some non-Gaussian distribution, e.g. impulsive noise. In such cases, nonsmooth data fitting is preferred due to its ability in fitting the uncorrupted data entries, e.g. ℓ_1 -norm data fitting was originally introduced in [2] for pure denoising problem. The importance of nondifferentiable data fitting such as the ℓ_1 -norm penalization has been examined deeply in [37, 38].

By combining TV regularization with ℓ_1 -norm penalized data fitting $\|Ku - f\|_1$, we arrive at the well known TVL1 model, i.e.

$$\min_u \text{TV}(u) + \mu \|Ku - f\|_1, \quad (1.2)$$

where $\text{TV}(u)$ can be either isotropic or anisotropic discretizations. In practice, the anisotropic TV is slightly easier to process than the isotropic one. In fact, TVL1 model with anisotropic TV can be reformulated as a linear program and has been studied in [27]. Theoretically, various geometrical properties of TVL1 model have been revealed in [15, 17, 58], which well justify its suitability for deblurring in the presence of impulsive noise. Numerically, very efficient algorithms have been designed for solving it, including the structure utilizing split-and-penalty method [57] and the augmented Lagrangian type methods [25, 31] which are also known as split Bregman methods. These methods can deal with both isotropic and anisotropic TV. We note that simple bound constraint can be incorporated to improve the quality of recovery, for which efficient structure-utilizing algorithms can still be designed, e.g. TVL1 model with constraint $0 \leq u \leq 1$ was considered in [14].

1.3. Two-phase method and corrected TVL1 method

The TVL1 model (1.2) fits all data entries, both corrupted and noise-free, via entry-wise absolute difference and with a unified weighting parameter μ . On the one hand, μ should be reasonably small in order to weaken the influence of fitting the corrupted data. On the other hand,

if μ is not big enough, the recovered image is mostly over regularized and very blocky. As a consequence, it becomes critical to choose an appropriate weighting parameter μ so that regularization and data fitting are well balanced, particularly for the cases with high level of noise. As a remedy, a two-phase method was proposed in [12] for image deblurring with impulsive and Gaussian noise. In the first stage, the outliers are approximately identified using median filters and removed from the data set. In the second stage, the image is restored by solving a regularized model with ℓ_1 -norm data fitting that applies to the remaining data entries. In [12], a regularizer that approximates the Mumford–Shah regularizer [36] was applied, and the non-smooth ℓ_1 -norm was approximated by a smooth function, see [12, equation (7)]. A similar two phase approach was proposed in [13], where anisotropic TV regularizer was used to replace the nonconvex regularizer used in [12]. The resulting problem was again solved by smoothing methods. We note that solving the variational models arising from the second stage is generally much harder than solving the TVL1, which, in contrast, has very efficient algorithms [14, 25, 31, 57].

Very recently, an iteratively *corrected TVL1* (abbreviated as CTVL1) method was proposed in [8]. The CTVL1 method accomplishes deblurring and impulsive noise removal simultaneously via solving a sequence of TVL1-equivalent problems. In particular, given the current point u^k , the CTVL1 method [8] generates the next iterate u^{k+1} via

$$u^{k+1} = \arg \min_u \text{TV}(u) + \mu (\|Ku - f\|_1 - \ell^k(u)), \quad (1.3)$$

where $\ell^k(u) = \langle s^k, Ku - f \rangle$, $s^k = (s_1^k, \dots, s_{n^2}^k)^T \in \mathbb{R}^{n^2}$,

$$s_i^k = \phi \left(\frac{(Ku^k - f)_i}{\|Ku^k - f\|_\infty} \right), \quad i = 1, 2, \dots, n^2, \quad \text{and} \quad \phi(t) = \text{sign}(t)(1 + \varepsilon^\tau) \frac{|t|^\tau}{|t|^\tau + \varepsilon^\tau}. \quad (1.4)$$

Here we adopt the convention $0/0 = 0$, and $\varepsilon, \tau > 0$ are given parameters. It has been demonstrated via extensive numerical results in [8] that the CTVL1 method performs very competitive and, in particular, outperforms the two-phase method [12] for RV noise.

1.4. Notation

Given a vector v , its dimension and i th component are denoted by $\dim(v)$ and v_i , respectively. The superscript T denotes matrix or vector transpositions. The standard inner product in \mathbb{R}^n is denoted by $\langle \cdot, \cdot \rangle$, i.e. $\langle x, y \rangle = x^T y$ for $x, y \in \mathbb{R}^n$. We follow standard notation in convex analysis [43]. The set of extended-real-valued, lower-semicontinuous, proper and convex functions on \mathbb{R}^n is denoted by $\Gamma_0(\mathbb{R}^n)$. The effective domain of an extended-real-valued function f on \mathbb{R}^n is denoted by $\text{dom} f$. The conjugate function f^* of a given $f \in \Gamma_0(\mathbb{R}^n)$ is defined by $f^*(x) = \sup_y \langle y, x \rangle - f(y)$. The indicator function of a set Ω is denoted by χ_Ω , i.e. $\chi_\Omega(x) = 0$ if $x \in \Omega$, and $+\infty$ otherwise. The Euclidean distance of a point x to a set S is denoted by $\text{dist}(x, S)$. The gradient of a multivariate differentiable function f is denoted by ∇f . If unspecified, $\|\cdot\|$ denotes the ℓ_2 -norm. Whenever there is no confusion, we let $\sum_i := \sum_{i=1}^{n^2}$.

1.5. Organization

The paper is organized as follows. The TVSCAD model and a *difference of convex functions* (DC) algorithm are proposed in section 2, with global convergence and convergence rate to a critical point established in section 3. Extensive numerical results, with comparisons to both TVL1 and CTVL1, are demonstrated in section 4. Finally, some concluding remarks are given in section 5.

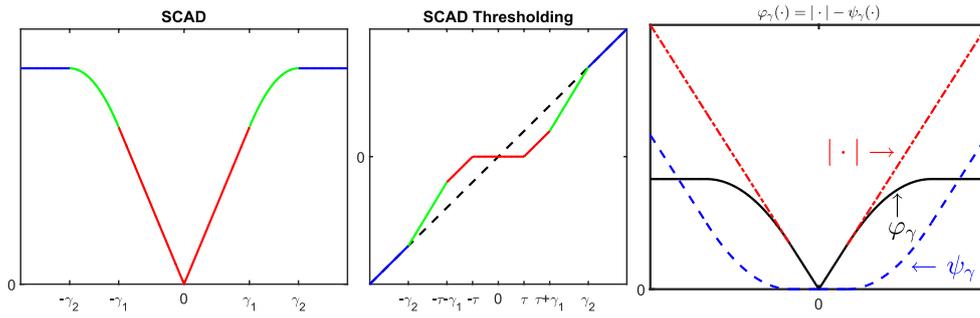


Figure 1. Left: the SCAD function φ_γ . Middle: the thresholding operator $\mathcal{T}(x, \varphi_\gamma, \tau)$. Right: A DC decomposition of the SCAD function: $\varphi_\gamma(\cdot) = |\cdot| - \psi_\gamma(\cdot)$.

2. A TVSCAD approach

In this section, we present a TVSCAD model, reformulate it as a DC program, and propose a *DC algorithm* (DCA). A brief overview on DC programming is also included.

2.1. A TVSCAD model

The SCAD function was originally introduced in [26] as a penalty function for sparse variable selection. Let $\gamma := (\gamma_1, \gamma_2)$ be a pair of given parameters satisfying $\gamma_2 > \gamma_1 > 0$. The one-dimensional SCAD function φ_γ is defined as

$$\varphi_\gamma(x) = \begin{cases} |x|, & \text{if } |x| \leq \gamma_1, \\ \frac{-x^2 + 2\gamma_2|x| - \gamma_1^2}{2(\gamma_2 - \gamma_1)}, & \text{if } \gamma_1 < |x| < \gamma_2, \\ \frac{\gamma_1 + \gamma_2}{2}, & \text{if } |x| \geq \gamma_2, \end{cases} \quad (2.1)$$

and its graph is given in the first plot of figure 1. It can be seen, from either the definition or the graph, that φ_γ coincides with the absolute value function if $|x| \leq \gamma_1$ and takes the constant value $(\gamma_1 + \gamma_2)/2$ if $|x| \geq \gamma_2$. For $\gamma_1 < |x| < \gamma_2$, the unique quadratic function, which connects $(\pm\gamma_1, \gamma_1)$ and $(\pm\gamma_2, (\gamma_1 + \gamma_2)/2)$ and makes the connected parts smoothly linked, were inserted. The thresholding operator, also known as shrinkage or proximity operator, of φ_γ is defined as

$$\mathcal{T}(x, \varphi_\gamma, \tau) := \arg \min_{y \in \mathbb{R}} \varphi_\gamma(y) + \frac{1}{2\tau}(y - x)^2, \quad x \in \mathbb{R}, \quad (2.2)$$

where $\tau > 0$ is a parameter. We note that $\tau \leq \gamma_2 - \gamma_1$ must be satisfied to guarantee that the objective function in (2.2) has a unique minimizer, see [26]. In this case, the graph of $\mathcal{T}(x, \varphi_\gamma, \tau)$ is given in the second plot of figure 1. It is clear from the graph that SCAD penalty function makes the resulting estimator possess three desired properties, namely, continuity, sparsity, and unbiasedness. In contrast, the widely used hard- and soft-thresholding estimators lack continuity and unbiasedness, respectively, see [26] for detailed discussions.

To present our TVSCAD model, we define the componentwise extension of φ_γ as follows

$$\Phi_\gamma(v) = \sum_{i=1}^{\dim(v)} \varphi_\gamma(v_i), \quad v \in \mathbb{R}^{\dim(v)}. \quad (2.3)$$

We propose the following TVSCAD model for image deblurring with impulsive noise:

$$\min_u \{ \text{TV}(u) + \mu \Phi_\gamma(Ku - f) \mid 0 \leq u \leq 1 \}, \quad (2.4)$$

where Φ_γ is defined in (2.3), $\text{TV}(u)$ can be either isotropic or anisotropic, and $\mu > 0$. The motivation of using SCAD function here is simply to enforce less or even null data fitting and more regularization whenever $(Ku)_i$ deviates significantly from f_i . This is quite reasonable as in such case the i th pixel is more likely to be corrupted. For those i such that $(Ku - f)_i$ is sufficiently small, the absolute deviation penalty is kept, as indicated by the definition of φ_γ . Note that here we also include in the model the bound constraint $0 \leq u \leq 1$, which usually improves the recovery quality [14]. It is clear that the TVSCAD model is nonconvex and nonsmooth since φ_γ is so. The use of nonconvex and nonsmooth functions in image recovery usually results in images with better contrasts and sharper edges, see [20, 21, 39, 40], which reveal the theoretical advantages of least squares problems regularized by nonconvex and nonsmooth functions. In our setting, the nonconvex and nonsmooth SCAD function is adopted for data fitting.

2.2. Reformulation as a DC program

The TVSCAD problem (2.4) is nonconvex and nonsmooth and, in general, very challenging to solve. We reformulate the TVSCAD model as a DC program by decomposing the SCAD function (2.1) as the difference of two convex functions, where the first is the absolute value function $|\cdot|$ and the second is given by

$$\psi_\gamma(x) := |x| - \varphi_\gamma(x) = \begin{cases} 0, & \text{if } |x| \leq \gamma_1, \\ \frac{x^2 - 2\gamma_1|x| + \gamma_1^2}{2(\gamma_2 - \gamma_1)}, & \text{if } \gamma_1 < |x| \leq \gamma_2, \\ |x| - \frac{\gamma_1 + \gamma_2}{2}, & \text{if } |x| > \gamma_2, \end{cases} \quad x \in \mathfrak{R}. \quad (2.5)$$

The decomposition $\varphi_\gamma = |\cdot| - \psi_\gamma$ is illustrated in the last plot of figure 1. The component-wise extension of ψ_γ , denoted by Ψ_γ , is given by

$$\Psi_\gamma(v) = \sum_{i=1}^{\dim(v)} \psi_\gamma(v_i), \quad v \in \mathfrak{R}^{\dim(v)}. \quad (2.6)$$

It then follows from (2.3) and (2.6) that $\Phi_\gamma(v) = \|v\|_1 - \Psi_\gamma(v)$ for all $v \in \mathfrak{R}^{\dim(v)}$. As a result, the TVSCAD model (2.4) can be reformulated as the following DC programming problem

$$\min_u \{ \text{TV}(u) + \mu (\|Ku - f\|_1 - \Psi_\gamma(Ku - f)) \mid 0 \leq u \leq 1 \}. \quad (2.7)$$

The idea of decomposing a nonconvex function as the difference of two convex functions is not new, see, e.g. [29], where it was applied to sparse recovery problems. Before deriving our algorithm for solving the reformulated problem (2.7), we next give a very briefly overview on DC programming.

2.3. DC programming

DC programming refers to optimizing a function that can be written as the difference of two convex functions. This class of problems covers generic convex optimization and many real world nonconvex problems as special cases. The study of subgradient method for convex maximization problem in [46, 47] can be viewed as early works on DC programming. Later,

DC duality theory and DCAs were developed in [50, 51] based on the nonconvex duality results in [53, 54]. Now, DC programming, including model, theory and algorithms have been widely used in many applications. Interested readers can refer to the review papers [3, 33, 48].

Let $g, h \in \Gamma_0(\mathbb{R}^n)$. A generic DC programming problem takes the form

$$\alpha := \inf_x g(x) - h(x). \quad (2.8)$$

It is standard result in convex analysis that the dual conjugate of $h \in \Gamma_0(\mathbb{R}^n)$ is itself, i.e. $h^{**} := (h^*)^* = h$, see, e.g. [43, theorem 12.2]. Therefore, the DC program (2.8) can be reformulated as

$$\alpha = \inf_x g(x) - h^{**}(x) = \inf_x (g(x) - \sup_y \langle x, y \rangle - h^*(y)) = \inf_y m(y), \quad (2.9)$$

where $m(y) := \inf_x g(x) - \langle x, y \rangle + h^*(y) = h^*(y) - g^*(y)$ if $y \in \text{dom } h^*$, and $+\infty$ otherwise. Plug into (2.9), we obtain

$$\alpha = \inf_y \{h^*(y) - g^*(y) \mid y \in \text{dom } h^*\}. \quad (2.10)$$

We assume that α is finite, which implies that $\text{dom } g \subset \text{dom } h$ and $\text{dom } h^* \subset \text{dom } g^*$, and adopt the convention $+\infty - (+\infty) = +\infty$. Then, (2.10) can be equivalently simplified as

$$\alpha = \inf_y h^*(y) - g^*(y), \quad (2.11)$$

which is clearly also a DC programming problem and is known as the dual problem of (2.8). Note that there exists perfect symmetry between the primal and the dual DC programming problems (2.8) and (2.11). Interested readers can refer to [49, sections 3.1 and 3.2] and references therein for more discussions on duality results, global and local optimality conditions of (2.8) and (2.11).

The classic DCAs (see [46, 47] and [49, section 3.3]) aim at solving the DC program (2.8) and its dual problem (2.11) to their global or local optimality. A simplified form of DCA solves (2.8) via linearizing the second part and solving a sequence of convex problems. Specifically, starting at an initial point $x^0 \in \text{dom } g$, DCA iterates as

$$x^{k+1} = \arg \min_x [g(x) - (h(x^k) + \langle y^k, x - x^k \rangle)], \quad y^k \in \partial h(x^k), \quad k = 0, 1, 2, \dots \quad (2.12)$$

Under certain conditions, the DCA is well defined [49, lemma 3.6], and its convergence properties were summarized in [49, theorem 3.7] and [48, theorem 3]. In particular, the sequence of function values $\{g(x^k) - h(x^k)\}$ is guaranteed to be decreasing, and if α is finite and the sequence $\{(x^k, y^k)\}$ is bounded, then every limit point of $\{x^k\}$ (resp. $\{y^k\}$) is a critical point of $g - h$ (resp. $h^* - g^*$), a notion that is slightly weaker than local minimizer [49, theorem 3.2]. We emphasize that the theoretical results derived for our DCA in section 2.4 are much stronger than those existing ones for generic DC programming, see section 3.

2.4. A DCA

A DCA for solving the proposed TVSCAD model (2.7) is derived in this subsection. It is easy to verify that ψ_γ defined in (2.5) is smooth, and so is Ψ_γ in (2.6). Given u^k satisfying $0 \leq u^k \leq 1$, our DCA generates the next iterate by

$$u^{k+1} = \arg \min_u \left\{ \text{TV}(u) + \mu (\|Ku - f\|_1 - \ell^k(u)) + \frac{\eta}{2} \|u - u^k\|^2 \mid 0 \leq u \leq 1 \right\}, \quad (2.13)$$

where $\eta > 0$ is a given parameter, and $\ell^k(u)$ is the linear part of $\Psi_\gamma(Ku - f)$ at u^k , i.e.

$$\ell^k(u) := \Psi_\gamma(Ku^k - f) + \langle K^T \nabla \Psi_\gamma(Ku^k - f), u - u^k \rangle. \quad (2.14)$$

It is straightforward to verify that the DCA defined in (2.13) and (2.14) is a special case of the classic DCA (2.12) for DC programming (2.8) with

$$g(u) = \text{TV}(u) + \mu \|Ku - f\|_1 + \chi_{0 \leq u \leq 1}(u) + \frac{\eta}{2} \|u\|^2 \text{ and } h(u) = \mu \Psi_\gamma(Ku - f) + \frac{\eta}{2} \|u\|^2. \quad (2.15)$$

DCA resulting from the above decomposition makes the objective function of (2.13) strongly convex, and thus u^{k+1} is well defined. Indeed, it is exactly because of the resulting proximal term $\frac{\eta}{2} \|u - u^k\|^2$ that makes our theoretical results in section 3 much stronger than those in [49]. In particular, the theory of DC programming for strongly convex g and h can only guarantee that $\lim_{k \rightarrow \infty} (u^{k+1} - u^k) = 0$, see, e.g. [49, theorem 3.7 (iii)], while we guarantee that the sequence $\{u^k\}$ converges globally to a critical point of the problem (2.4). Moreover, the addition of the proximal term also ensures R-linear convergence rate in the case of anisotropic TV. In computation, a small value of η can always be used to enhance robustness.

Next, we adapt the versatile and efficient *alternating direction method of multipliers* (ADMM, [28, 30]) to solving the subproblem (2.13). ADMM can be viewed as a practical variant of the classic augmented Lagrangian method [32, 42] or a dual application of the Douglas-Rachford splitting [19, 23] for monotone inclusion problem. ADMM has been applied to numerous applications, see, e.g. [10, 22], among others, it has been applied to TV based image deblurring problem [31] and its variants [25]. Given the encouraging performance of ADMM, especially for TV regularized image recovering problems, it is appropriate here to adopt it as the subproblem solver.

ADMM has been applied to constrained TVL1 problem in [14]. With the additional proximal term $\frac{\eta}{2} \|u - u^k\|^2$ and the linear term $\ell^k(u)$ added to the TVL1 problem, the resulting ADMM is similar. Here we present the algorithm only for completeness. We take the isotropic discretization of TV as an example, i.e. $\text{TV}(u) = \sum_i \|D_i u\|$, and the discussion for the case of anisotropic TV is completely analogous. First, we introduce a set of auxiliary variables $\{\mathbf{w}_i \in \mathbb{R}^2 : i = 1, \dots, n^2\}$ and $v \in \mathbb{R}^{n^2}$ to transfer $\{D_i u : i = 1, \dots, n^2\}$ and $Ku - f$ out of the nondifferentiable norms. The set of \mathbf{w}_i 's is also denoted by $w \in \mathbb{R}^{2n^2}$. In order to treat the bound constraint, we also need an additional auxiliary variable $x \in \mathbb{R}^{n^2}$. Define

$$\theta(w) := \sum_i \|\mathbf{w}_i\|, \quad v^k := Ku^k - f, \quad p^k := \nabla \Psi_\gamma(v^k) \quad \text{and} \quad \Omega := \{u \mid 0 \leq u \leq 1\}. \quad (2.16)$$

By omitting constant values, (2.13) can be reformulated as

$$\min_{u, v, w, x} \left\{ \theta(w) + \mu (\|v\|_1 - \langle p^k, v \rangle) + \frac{\eta}{2} \|u - u^k\|^2 \mid \mathbf{w}_i = D_i u, \forall i, v = Ku - f, x = u, x \in \Omega \right\}. \quad (2.17)$$

The augmented Lagrange function associated with (2.17) is

$$\begin{aligned} \mathcal{L}(w, v, x, u, \lambda_w, \lambda_v, \lambda_x) = & \theta(w) + \sum_i (\langle (\lambda_w)_i, D_i u - \mathbf{w}_i \rangle + \frac{\beta_w}{2} \|D_i u - \mathbf{w}_i\|^2) \\ & + \mu (\|v\|_1 - \langle p^k, v \rangle) - \langle \lambda_v, v - (Ku - f) \rangle + \frac{\beta_v}{2} \|v - (Ku - f)\|^2 \\ & + \frac{\eta}{2} \|u - u^k\|^2 - \langle \lambda_x, x - u \rangle + \frac{\beta_x}{2} \|x - u\|^2, \end{aligned}$$

where $\lambda_w \in \mathfrak{R}^{2n^2}$, $\lambda_v, \lambda_x \in \mathfrak{R}^{n^2}$ are Lagrange multipliers, and $\beta_w, \beta_v, \beta_x > 0$ are penalty parameters. Given u^0 and $(\lambda_w^0, \lambda_v^0, \lambda_x^0)$, the classic 2-block ADMM iterates as

$$(w^{j+1}, v^{j+1}, x^{j+1}) = \arg \min_{w, v, x} \{ \mathcal{L}(w, v, x, u^j, \lambda_w^j, \lambda_v^j, \lambda_x^j) \mid x \in \Omega \}, \quad (2.18a)$$

$$u^{j+1} = \arg \min_u \{ \mathcal{L}(w^{j+1}, v^{j+1}, x^{j+1}, u, \lambda_w^j, \lambda_v^j, \lambda_x^j) \}, \quad (2.18b)$$

$$\begin{pmatrix} \lambda_w^{j+1} \\ \lambda_v^{j+1} \\ \lambda_x^{j+1} \end{pmatrix} = \begin{pmatrix} \lambda_w^j - \beta_w(w^{j+1} - Du^{j+1}) \\ \lambda_v^j - \beta_v(v^{j+1} - Ku^{j+1} + f) \\ \lambda_x^j - \beta_x(x^{j+1} - u^{j+1}) \end{pmatrix}. \quad (2.18c)$$

We note that both (2.18a) and (2.18b) have closed form solutions and hence can be computed efficiently. In fact, the optimization of w, v and x in (2.18a) can be carried out in parallel since they are separable from each other. Moreover, w^{j+1} and v^{j+1} are given explicitly by the proximity operators of θ and $\|\cdot\|_1$, respectively, the computations of which have linear cost. On the other hand, x^{j+1} can be computed via a projection onto Ω . For convenience of the readers, here we present the formulas for computing these variables. Denote the proximity operator of any $\vartheta \in \Gamma_0(\mathfrak{R}^n)$ by $\text{prox}_{\vartheta}(\cdot) := \arg \min_y \vartheta(y) + \frac{1}{2}\|y - \cdot\|^2$ and the Euclidean projection onto Ω by proj_{Ω} . Then, the solutions of (2.18a) are given by

$$\begin{cases} w^{j+1} &= \text{prox}_{\frac{\theta}{\beta_w}}(Du^j + \lambda_w^j/\beta_w), \\ v^{j+1} &= \text{prox}_{\frac{\mu}{\beta_v}\|\cdot\|_1}(Ku^j - f + (\lambda_v^j + \mu p^k)/\beta_v), \\ x^{j+1} &= \text{proj}_{\Omega}(u^j + \lambda_x^j/\beta_x). \end{cases}$$

The u -subproblem (2.18b) is a least-squares problem, whose normal equations have a coefficient matrix $(\beta_w D^T D + \beta_v K^T K + (\beta_x + \eta)I)$. Under the assumption of periodic boundary conditions, this matrix has a BCCB (block circulant matrix with circulant blocks) structure and thus can be diagonalized by two-dimensional discrete Fourier matrix. Therefore, the u -subproblem can be solved very efficiently via two fast Fourier transforms. The problem structures of this type were first recognized and fully exploited in [55], where an alternating minimization algorithm was designed based on quadratic penalty method for image deblurring with Gaussian noise. Later, this splitting and alternating minimization idea was extended in [56, 57] to solve TVL1 problem and multichannel image deblurring.

2.5. Connections with the CTVL1 method

The connection between our DCA (2.13) and the CTVL1 method (1.3) recently proposed in [8] is explained in this subsection. By removing in (2.13) the bound constraint $0 \leq u \leq 1$ and the proximal term $\frac{\eta}{2}\|u - u^k\|^2$, we see that (2.13) differs from (1.3) only in the choice of $\ell^k(u)$. By letting $v^k = Ku^k - f$ and throwing away some constants, one can see that $\ell^k(u)$ for both methods has the form $\ell^k(u) = \langle s^k, Ku - f \rangle$ with $s^k = (s_1^k, \dots, s_{n^2}^k)^T$ defined by

$$s_i^k = \begin{cases} \phi(v_i^k/\|v^k\|_{\infty}), & \text{for (1.3),} \\ \psi'_{\gamma}(v_i^k), & \text{for (2.13),} \end{cases} \quad i = 1, \dots, n^2. \quad (2.19)$$

Here ϕ and ψ_{γ} are defined in (1.4) and (2.5), respectively. It was suggested in [8] that $\varepsilon^2 = 10^{-3}$ and $\tau = 2$ should be used in ϕ , and the derivative of ψ_{γ} is given by

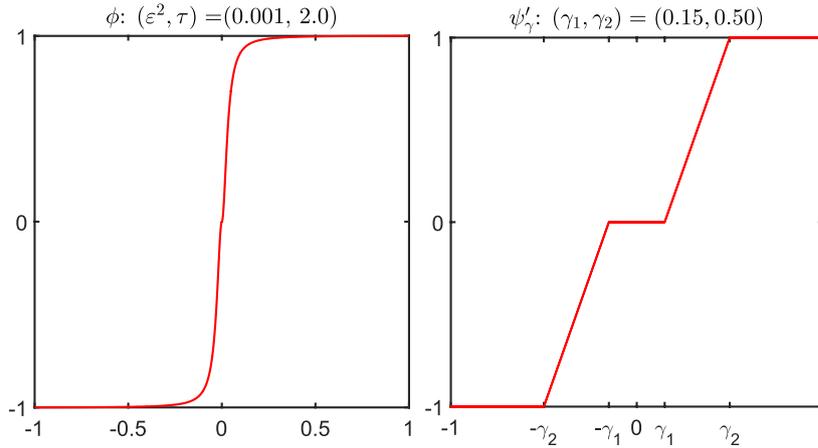


Figure 2. Left: the ϕ function used by the CTVL1 method [8]. Right: the derivative of ψ_γ .

$$\psi'_\gamma(x) = \begin{cases} 0, & \text{if } |x| \leq \gamma_1, \\ \frac{x - \gamma_1 \text{sign}(x)}{\gamma_2 - \gamma_1}, & \text{if } \gamma_1 < |x| \leq \gamma_2, \\ \text{sign}(x), & \text{if } |x| > \gamma_2, \end{cases} \quad x \in \mathfrak{R}.$$

Plots of ϕ and ψ'_γ (with $\gamma_1 = 0.15$, $\gamma_2 = 0.5$, and the practical choice of γ_1 and γ_2 will be discussed in section 4) are given in figure 2. From the figure, one may observe that if $v_i^k = (Ku^k - f)_i$ is positive and relatively large, the values of s_i^k will tend to be 1 and equal to 1 in (1.3) and (2.13), respectively. Consequently, this means the data fitting term $\mu (\|Ku - f\|_1 - \ell^k(u))$ being nearly or totally cancelled out from (1.3) and (2.13), respectively. In other words, nearly no or just no penalization being enforced on $v_i^k = (Ku^k - f)_i$. This coincides with our motivation as large $(Ku^k - f)_i$ means the i th pixel is more likely to be corrupted by the impulsive noise, and thus should not be fitted. When $v_i^k = (Ku^k - f)_i$ is negative with its absolute value relatively large, similar explanation applies. On the other hand, when the absolute value of $v_i^k = (Ku^k - f)_i$ is relatively small, $\mu |(Ku - f)_i|$ corresponding to ℓ_1 -norm penalized data fitting is used in (2.13) as opposed to $\mu |(Ku - f)_i|$ minus a linear term used in (1.3).

We also mention that the outer loop convergence of the CTVL1 method is unknown in [8], while our algorithm is guaranteed to converge globally to a critical point of the TVSCAD problem (2.4), as will be shown in section 3.

3. Theoretical analysis

A major contribution of this paper lies in the theoretical side. In this section, we establish global convergence of the DCA (2.13). In comparison, the outer loop convergence of the CTVL1 method [8] is not guaranteed. Moreover, by further exploring the structures of the SCAD function, results on convergence rate are also established for the DCA (2.13).

We start with some definitions. Let F be an extended-real-valued, proper and lower semi-continuous function on \mathfrak{R}^n . The limiting subdifferential of F at $x \in \text{dom } F$ is defined by

$$\partial F(x) := \left\{ v \in \mathbb{R}^n : \exists x' \rightarrow x, F(x') \rightarrow F(x), v' \rightarrow v, \liminf_{z \neq x', z \rightarrow x'} \frac{F(z) - F(x') - \langle v', z - x' \rangle}{\|z - x'\|} \geq 0 \right\}.$$

A necessary condition for $x \in \mathbb{R}^n$ to be a minimizer of F is $0 \in \partial F(x)$, and a point satisfying this condition is called a critical or stationary point of F . Readers are referred to [44] for these basics. The function F is said to have the KL property at $\bar{x} \in \text{dom } \partial F$ with an exponent of σ (see [6, 7]) if there exist $c, \epsilon > 0$ and $\nu \in (0, \infty]$ such that

$$\text{dist}(0, \partial F(x)) \geq c(F(x) - F(\bar{x}))^\sigma \quad (3.1)$$

for all x satisfying $\|x - \bar{x}\| \leq \epsilon$ and $F(\bar{x}) < F(x) < F(\bar{x}) + \nu$. If F has the KL property with the exponent σ at any $x \in \text{dom } \partial F$, then F is called a KL function with an exponent of σ . We continue using the notation of v^k and Ω as defined in (2.16), and according to (2.4) and (2.7), we define

$$F(u) := \text{TV}(u) + \mu \Phi_\gamma(Ku - f) + \chi_\Omega(u) = \text{TV}(u) + \mu (\|Ku - f\|_1 - \Psi_\gamma(Ku - f)) + \chi_\Omega(u).$$

Lemma 3.1 (Sufficient decrease property). *For any $u^0 \in \Omega$ and $\eta > 0$, the sequence $\{u^k\}$ generated by (2.13) satisfies*

$$F(u^k) - F(u^{k+1}) \geq \eta \|u^{k+1} - u^k\|^2, \quad \forall k \geq 0. \quad (3.2)$$

Proof. By definition, (2.13) implies that $u^k \in \Omega$ and thus $\chi_\Omega(u^k) = 0$ for all $k \geq 0$. Let $k \geq 0$ be fixed. The convexity of Ψ_γ implies that $\Psi_\gamma(v^{k+1}) \geq \Psi_\gamma(v^k) + \langle \nabla \Psi_\gamma(v^k), v^{k+1} - v^k \rangle$. Thus,

$$F(u^k) - F(u^{k+1}) \geq (\text{TV}(u^k) + \mu \|v^k\|_1) - (\text{TV}(u^{k+1}) + \mu \|v^{k+1}\|_1) + \mu \langle \nabla \Psi_\gamma(v^k), v^{k+1} - v^k \rangle. \quad (3.3)$$

It follows from (2.13) that

$$s^{k+1} := \mu K^T \nabla \Psi_\gamma(v^k) + \eta(u^k - u^{k+1}) \in \partial [\text{TV}(u) + \mu \|Ku - f\|_1 + \chi_\Omega(u)] \Big|_{u=u^{k+1}}.$$

Further considering the convexity of $\text{TV}(u) + \mu \|Ku - f\|_1 + \chi_\Omega(u)$, we obtain from (3.3) that

$$F(u^k) - F(u^{k+1}) \geq \langle s^{k+1}, u^k - u^{k+1} \rangle + \mu \langle \nabla \Psi_\gamma(v^k), v^{k+1} - v^k \rangle = \eta \|u^{k+1} - u^k\|^2,$$

which completes the proof. \square

Lemma 3.2 (KL property). *If the TV is anisotropic as given in (1.1), then for any fixed $\mu > 0$ the function F is a KL function with exponent $1/2$.*

Proof. According to the definitions of anisotropic TV (1.1) and Ψ_γ (2.6), F is a piecewise linear-quadratic function, and every piece is defined on a closed polyhedral set. Denote by m the number of pieces of F , F_i the linear-quadratic function on the i th piece, and C_i the domain of F_i . Then, it holds that $F(u) = \min_{1 \leq i \leq m} \{F_i(u) + \chi_{C_i}(u)\}$. Since χ_{C_i} is a proper closed polyhedral function, it follows from [35, corollary 5.2] that F is a KL function with exponent $1/2$. \square

Now we are ready to establish the global convergence and convergence rate results for our DCA. In the general theory of DC programming, see, e.g. [49, theorem 3.7 (iv)], it is only guaranteed that, if α in (2.8) is finite and the generated sequence is bounded, then any accumulation point is a critical point. The same result is derived in part (i) of the following theorem

3.3, while the rest of the theorem, namely part (ii)–(iv), are much stronger than existing results for generic DC programming.

Theorem 3.3 (Global convergence and rate of convergence). *Let $u^0 \in \Omega$, $\eta > 0$ and $\{u^k\}$ be the sequence generated by (2.13). Then*

- (i) any accumulation point of $\{u^k\}$ is a critical point of (2.4);
- (ii) $\{u^k\}$ converges globally to a critical point of (2.4), and furthermore $\sum_{k=0}^{\infty} \|u^{k+1} - u^k\| < +\infty$;
- (iii) if the TV is anisotropic, then there exist $h > 0$ and $\tau \in (0, 1)$ such that $\|u^k - u^*\| \leq h\tau^k$ for all $k \geq 0$, i.e. the convergence rate is R -linear;
- (iv) if the TV is isotropic, then $\{u^k\}$ converges to u^* at least sublinearly.

Proof.

- (i) It follows from (2.13) that

$$d^{k+1} := \mu K^T (\nabla \Psi_\gamma(v^k) - \nabla \Psi_\gamma(v^{k+1})) - \eta(u^{k+1} - u^k) \in \partial F(u^{k+1}). \quad (3.4)$$

Since $u^k \in \Omega$ for all $k \geq 0$ and $\text{TV}(u) + \mu \Phi_\gamma(u)$ is bounded below (by 0), it follows from (3.2) that

$$\sum_{k=0}^{\infty} \|u^{k+1} - u^k\|^2 \leq F(u^0)/\eta < +\infty.$$

It thus follows that $\lim_{k \rightarrow \infty} (u^{k+1} - u^k) = 0$. Let u^* be any accumulation point of $\{u^k\}$ and $\{u^{n_k}\}$ be a subsequence such that $\lim_{k \rightarrow \infty} u^{n_k} = u^*$. Then, $\lim_{k \rightarrow \infty} u^{n_k+1} = \lim_{k \rightarrow \infty} [(u^{n_k+1} - u^{n_k}) + u^{n_k}] = u^*$. By replacing k by n_k and letting $k \rightarrow +\infty$ on the both sides of (3.4), we obtain immediately from the upper semicontinuity of ∂F and the continuity of $\nabla \Psi_\gamma$ that $0 \in \partial F(u^*)$, which completes the proof of part (i).

- (ii) According to [7, theorem 2.9], it suffices to guarantee that the sufficient decrease condition of $\{F(u^k)\}$, the relative error condition on $\{d^k\}$ and the boundedness of $\{u^k\}$ hold for all k . In fact, $\{F(u^k)\}$ is sufficiently decreasing due to (3.2), and $\{u^k\}$ is bounded since $0 \leq u^k \leq 1$. Moreover, since $\nabla \Psi_\gamma$ is Lipschitz continuous, we know from (3.4) that there must exist $M > 0$ sufficiently large such that

$$\|d^{k+1}\| \leq M \|u^{k+1} - u^k\|. \quad (3.5)$$

Thus, all the conditions required to guarantee the global convergence of $\{u^k\}$ to a critical point have been fulfilled. The global convergence of $\{u^k\}$, as well as the inequality $\sum_{k=0}^{\infty} \|u^{k+1} - u^k\| < +\infty$, follows directly from [7, theorem 2.9].

- (iii) From (3.2), we see immediately that

$$F(u^{k+1}) - F(u^*) \geq \sum_{i=k+1}^{\infty} \eta \|u^{i+1} - u^i\|^2. \quad (3.6)$$

By lemma 3.2, F has the KL property at u^* with $\sigma = 1/2$. It thus follows from (3.1) and (3.5) that there exists a sufficiently large K such that

$$F(u^{k+1}) - F(u^*) \leq \frac{1}{c^2} \text{dist}^2(0, \partial F(u^{k+1})) \leq \frac{M^2}{c^2} \|u^{k+1} - u^k\|^2, \quad \forall k \geq K. \quad (3.7)$$

Define $A_k = \sum_{i=k}^{\infty} \|u^{i+1} - u^i\|^2$. Combing (3.6) and (3.7), we get $A_{k+1} \leq \frac{M^2}{\eta c^2} (A_k - A_{k+1})$, or equivalently $A_{k+1} \leq \tau^2 A_k$ for all $k \geq K$, where $\tau := M/\sqrt{M^2 + \eta c^2} \in (0, 1)$. Hence, there exist a constant $\rho > 0$ sufficiently large such that $A_k \leq \rho^2 \tau^{2k}$ for all $k \geq 0$, and hence $\|u^{k+1} - u^k\| \leq \rho \tau^k$, for all $k \geq 0$. Thus,

$$\|u^k - u^*\| \leq \sum_{i=k}^{\infty} \|u^{i+1} - u^i\| \leq \frac{\rho}{1-\tau} \tau^k := h \tau^k, \quad \forall k \geq 0.$$

- (iv) Since the isotropic TV is semialgebraic, so is F . It then follows that F is a KL function with certain exponent $\sigma \in [1/2, 1)$, see [6] for details. If $\sigma > 1/2$, then it is straightforward to show by following [5, theorem 2] that $\|u^k - u^*\| \leq h k^{-\frac{1-\sigma}{2\sigma-1}}$ for some $h > 0$ and all $k \geq 1$, i.e. the convergence rate is sublinear. If $\sigma = 1/2$, then similar to the proof of (iii), we can derive that the sequence $\{u^k\}$ converges to u^* with the faster R-linear rate. \square

4. Numerical results

In this section, numerical results are presented to demonstrate the performance of the proposed TVSCAD model (2.4) and the corresponding DCA (2.13). The superior performance of CTVL1 compared to the two-phase method [12] in terms of recovery quality has been demonstrated in [8] via extensive numerical results. Therefore, we only compare TVSCAD with CTVL1. For reference purpose, we also present results of TVL1. We note that the per iteration computational cost of TVSCAD and CTVL1 are roughly the same since both methods solve a TVL1-equivalent problem. The overall computational cost of both methods are also comparable since they consume roughly the same number of outer iterations. Apparently, both methods are more expensive than solving a single TVL1 problem.

Evaluating the quality of a recovered image is a challenging task, especially when the ground truth is unknown. Though several measures are available when the ground truth is given, e.g. MSE (means square error), SNR (signal-to-noise ratio), PSNR (peak signal-to-noise ratio) and SSIM (structural similarity index), the evaluation process has to be aided by human eyes in many situations. Here we choose SNR as the measure, which is commonly used in the literature. Let \bar{u} and u be the original and the recovered images, respectively, and denote the mean intensity of \bar{u} by $\text{mean}(\bar{u})$. The SNR (in dB) of u is defined by

$$\text{SNR} := 20 \times \log_{10} (\|\bar{u} - \text{mean}(\bar{u})\| / \|\bar{u} - u\|). \quad (4.1)$$

According to our experiments, adding the bound constraint $0 \leq u \leq 1$ can generally stabilize the compared algorithms and improve the recovery quality. Therefore, we incorporate this constraint into all the compared algorithms, and the resulting algorithms will still be referred as TVL1 and CTVL1. In our experiments, constrained TVL1-equivalent problems, e.g. (2.18) for TVSCAD, were always solved by ADMM with the same set of parameters. It is also our experience that the isotropic and anisotropic discretizations of TV do not result in significantly different recoveries in terms of both visual quality and SNR. Thus, we simply chose the anisotropic TV in our tests, which has favorable theoretical guarantee for TVSCAD.

All algorithms were implemented in Matlab, and the experiments were executed on a Lenovo Thinkpad laptop with an Intel Core i7-3667U CPU at 2.00 GHz and 8 GB of memory, running 64 bit Microsoft Windows 8 and Matlab v8.6 (R2015b).

4.1. Details of setting

The details of our experiments are summarized below.

- *Tested images.* In our experiments, we chose two images to test, i.e. House (256-by-256) and Pepper (512-by-512), both are obtained from the USC-SIPI Image Database⁴. The two images are named `house` and `4.2.07` in the image data base, both of which have been reused a lot in the academic image processing community. The original images are given in figure 3. In fact, a few other nonabnormal images were tested as well. Since the results remain largely similar provided that relevant parameters are properly set and also due to the limitation of space, only results of these two images are included. The reason that we choose to present results of these two images is because they are widely used in academic research works of imaging sciences.
- *Tested blur.* We tested three types of blur, i.e. Gaussian, motion and average blurs, the kernel of which were generated by the Matlab function `fspecial`. A Gaussian kernel obeys two-dimensional truncated and normalized Gaussian distribution and thus is determined by the kernel size and the standard deviation, an average kernel obeys uniform distribution, while a motion blur is determined by the length and angle of the motion. For Gaussian and average blurs, the tested kernel size, denoted by `hsize`, was set to be 9. For Gaussian blur, the standard deviation, denoted by `std`, of the tested kernel was set to be 10. For motion blur, we tested `len = 7` and `angle = 45°`. Readers are referred to the MathWorks documentation for details about what `fspecial` does and how it is done. We have also tested disk and pillbox blurs. Since the comparison results remain largely alike, we only present recovered results of Gaussian, motion and average blurs.
- *Impulsive noise.* We tested the two common types of impulsive noise, i.e. SP and RV noise. In general, RV noise is much harder to remove than SP noise. In our experiments, we tested 90% SP noise and 70% RV noise.
- *Parameters: μ , τ , ε , γ_1 , γ_2 , η , and others.* Now we specify the model and algorithmic parameters used in our tests.
 - (i) The ‘best’ choice of the parameter μ is known to be problem dependent and very hard to find. In fact, this is largely an open problem, even for the simpler TVL2 model [45]. In our experiments, based on the widely used Lena image, which can also be found at the USC-SIPI Image Database, we first solved a sequence of bound constrained TVL1 problems to search for each specific case the ‘best’ choice of μ , i.e. a value that gives nearly the highest possible SNR, for bounded constrained TVL1 problems. These ‘best’ values of μ will then be used to determine the parameters for TVSCAD and CTVL1.
 - (ii) For TVSCAD and CTVL1, the model parameter μ is set by $\mu = \frac{c\mu^*}{1-r}$. Here μ^* denotes the ‘best’ μ found based on the Lena image for bound constrained TVL1 model, $c > 0$ is a constant, and r denotes the noise level defined by $(\# \text{ corrupted pixels})/(\# \text{ total pixels})$. This simple rule for setting μ is based on numerical experiments. We set $c = 5$ and $c = 3$ for SP and RV noise, respectively, which may be suboptimal but performed favourably in our tests. We emphasize that both TVSCAD and CTVL1 are much less sensitive in the choice of μ than TVL1 due to the additional correction term $\ell^k(u)$ at each iteration, see detailed discussions in [8]. Continuation on μ , i.e. starting from a small value and increasing it gradually, is also applied to enhance the performance and stability of the algorithms.

⁴<http://sipi.usc.edu/database/>



Figure 3. Tested images. Left: House (256-by-256). Right: Pepper (512-by-512).

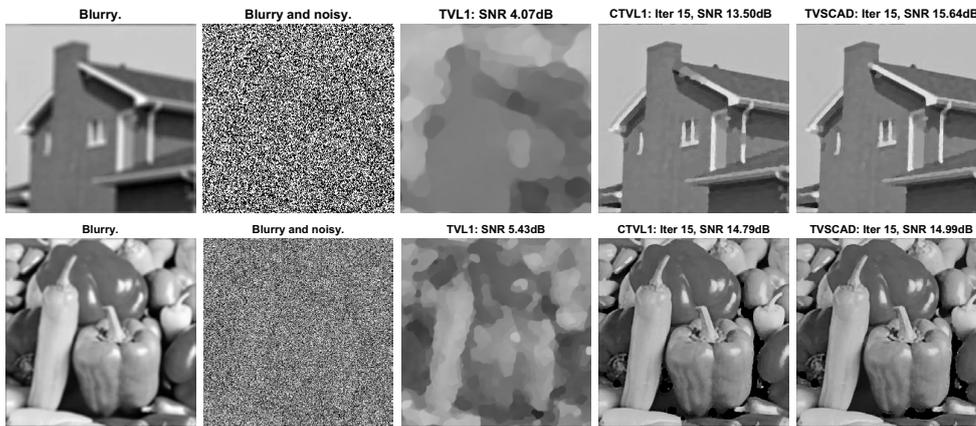


Figure 4. Gaussian blur ($hsize = 9$, $std = 10$), 90% SP noise.

- (iii) Parameters τ and ε , which define ϕ in (1.4) for CTVL1, and parameters γ_1 and γ_2 , which define the SCAD function φ_γ in (2.1), are also problem dependent. Consequently, the ‘best’ choices of these parameters are very difficult to determine. For CTVL1, we set $\tau = 2$ and $\varepsilon = 10^{-3}$, as recommended by the authors of [8]. For TVSCAD, we set $\gamma_2 = \max(0.2 \times 0.85^{k-1}, 0.1)$ and $\gamma_1 = 0.08/k$, where $k \geq 1$ denotes the iteration counter. Our choice of γ_2 is based on experiments, which performed favourably in all the tested cases. It is definitely possible to adjust the rule for choosing γ_2 case by case. However, a uniformly defined γ_2 is preferred.
- (iv) The proximal term $\frac{\eta}{2} \|u - u^k\|^2$ is added mainly to facilitate the theoretical analysis. In practice, the presence of this term also prevents u^{k+1} from deviating too much from u^k , and thus stabilizes the algorithm. We set η to be 10^{-5} throughout. For solving the constrained TVL1 problem like (2.17), we set $\beta_w = 50$, $\beta_v = 100$ and $\beta_x = 100$ uniformly for all cases, and ADMM performs very stable and efficient.

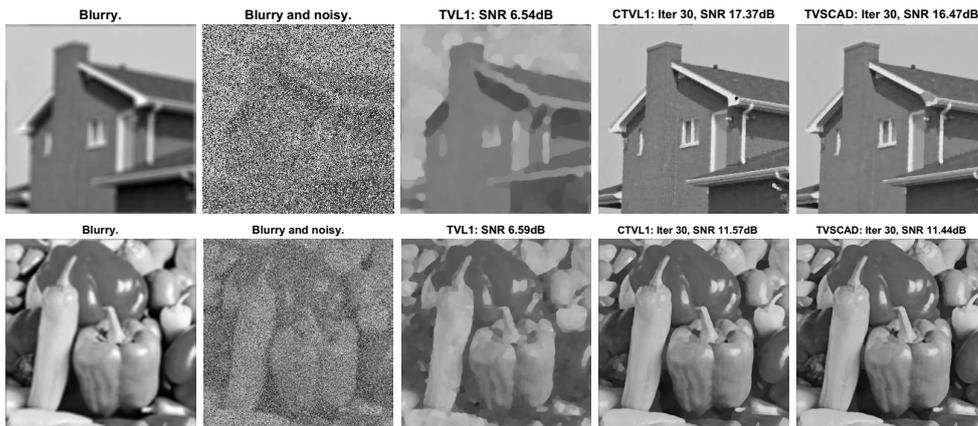


Figure 5. Gaussian blur ($hsize = 9$, $std = 10$), 70% RV noise.

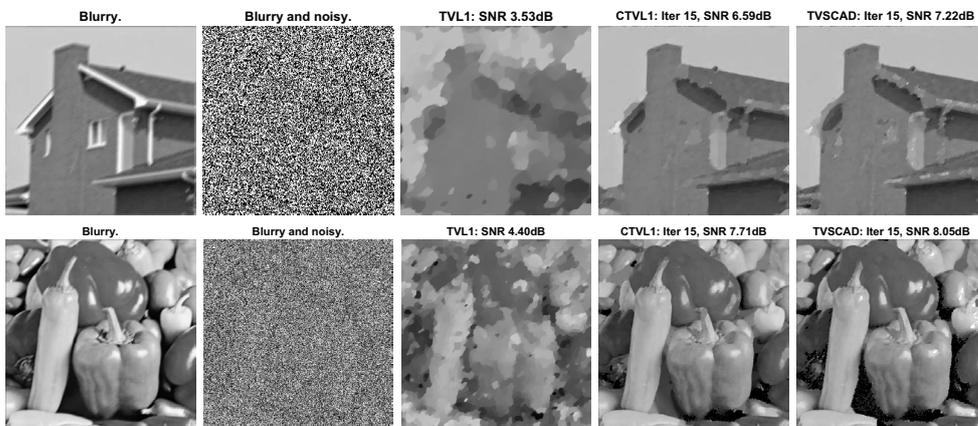


Figure 6. Motion blur ($len = 7$, $angle = 45^\circ$), 90% SP noise.

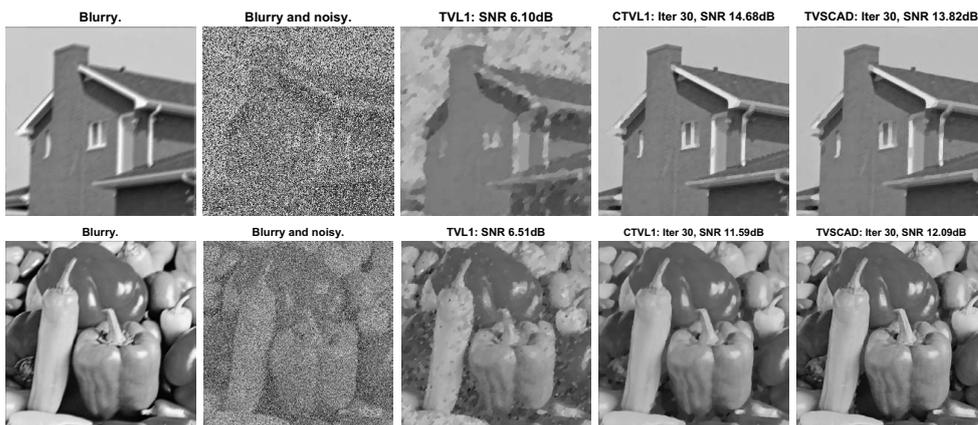


Figure 7. Motion blur ($len = 7$, $angle = 45^\circ$), 70% RV noise.

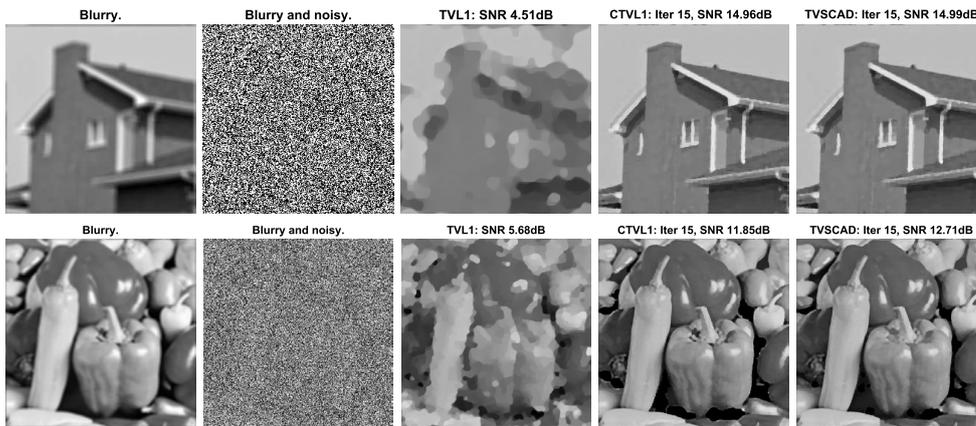


Figure 8. Average blur (9×9), 90% SP noise.

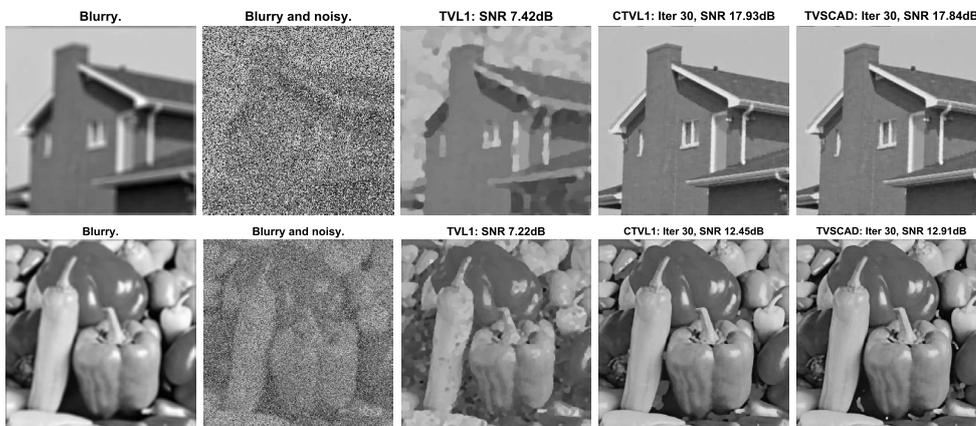


Figure 9. Average blur (9×9), 70% RV noise.

- *Initialization.* For solving TVL1 problem by ADMM, we set $u^0 = f$ always. For TVSCAD and CTVL1, u^0 is set to be the solution obtained from solving TVL1. To compute u^{k+1} from u^k by solving the TVL1-equivalent problem like (2.13), ADMM (2.18) is initialized from u^k . Each time we launched ADMM, all the starting Lagrange multipliers are set to zero.
- *Stopping rule.* For solving each TVL1-equivalent subproblem, the ADMM (2.18) was terminated by $\|u^{j+1} - u^j\| / (1 + \|u^j\|) \leq 10^{-4}$. According to our experiments, in the case of SP noise, the quality of recovered images does not change significantly after no more than 15 iterations for both TVSCAD and CTVL1, while in the case of RV noise, a few more iterations are usually required, see [8] for a similar conclusion. In our experiments, we terminated both methods after 15 and 30 iterations for SP and RV noise, respectively. For TVSCAD, a more practical stopping rule could be introduced depending on detecting the stagnation of objective function values in (2.4). However, this is not possible for CTVL1 since the convergence of its outer loop is unknown. Nonetheless, a practically effective stopping rule that performs uniformly favorable in terms of SNR deserves further research.

4.2. Comparison results with TVL1 and CTVL1

In this section, we present comparison results with CTVL1. For reference purpose, results of TVL1 are also presented. We emphasize that we have modified TVL1 and CTVL1 by adding the bound constraint $0 \leq u \leq 1$ to improve the recovery quality. Without this constraint, pixel values of recovered images may fall out of $[0, 1]$. In such cases, truncation or rescaling technique needs to be incorporated, which influences the SNR values of recovered results. We tested three blurs, i.e. Gaussian blur (`hsize` = 9, `std` = 10), motion blur (`len` = 7, `angle` = 45°) and average blur (`hsize` = 9). For each blur, 90% SP noise and 70% RV noise were tested. Detailed experimental results are given in figures 4–9, including the blurry images, the blurry and noisy images and the recovered images by all the compared algorithms.

It can be seen from figures 4–9 that both TVSCAD and CTVL1 outperform TVL1 significantly for all the tested cases. It is also observed in our experiments that for low noise level TVL1 performs reasonably well, while TVSCAD and CTVL1 perform better. For high noise level, TVSCAD and CTVL1 can improve the results of TVL1, and the improvements are mostly significant. This is desirable and easy to understand because both methods enforce less fitting on likely noisy data, as a result of which the likely uncorrupted data can be fitted more sufficiently by attaching a parameter μ larger than that of TVL1. From figures 4–9, the images recovered by both CTVL1 and TVSCAD are apparently much better than those recovered by TVL1 in terms of both visual quality and SNR. As we observe from the numerical experiments, in general both TVSCAD and CTVL1 improve SNR as the algorithms proceed. Moreover, the objective function value of (2.4) also decreases monotonically and quickly. Comparing the results of CTVL1 and TVSCAD, we see that TVSCAD performs competitively with CTVL1. In particular, TVSCAD performs mostly better than CTVL1 in the case of SP noise. For RV noise, TVSCAD is also competitive with slightly inferior performance.

We remark that the performances of both TVSCAD and CTVL1 depend on a few parameters, as specified at the beginning of this section. In addition, the adaptive choices of and the continuation strategies on these parameters also influence the behavior of the algorithms. As a consequence, different parameters may lead to opposite results in performances of TVSCAD and CTVL1. It is definitely possible to tune relevant parameters case by case to obtain much better results for each test. However, for nonconvex problems it is generally very challenging to tune an optimization algorithm to its best performance by using a unified set of parameters and adaptive rules. The numerical results presented here are only to illustrate the feasibility and the potential superiority of the proposed model and algorithm. Issues including the choice of model and algorithm parameters, the adaptive choices of these parameters and a practically useful stopping rule definitely deserve further investigations given the promising performance of the proposed TVSCAD model and the DCA.

5. Concluding remarks

We proposed, analyzed and tested a nonconvex TVSCAD model for image deblurring with impulsive noise corruption. To solve the nonconvex TVSCAD model, we proposed a DCA, which enjoys favorable convergence properties, namely, global convergence to a critical point of the nonconvex objective function, R-linear rate of convergence in the case of anisotropic TV, and at-least-sublinear convergence rate for isotropic TV regularization case. These results are much stronger than existing results for general DC programming. Extensive numerical results demonstrated that TVSCAD performs favorably. In particular, it performs much better than TVL1 and very competitive with the CTVL1 method [8].

We note that the idea of this paper can be extended to solve more problems. For example, color image processing based on the multichannel TV regularization proposed in [57]. The idea of DCA can also be extended to solve problems with other nonconvex penalty function, such as folded concave functions [29]. At present, it remains largely unexplored on how to adaptively choose the weighting parameter μ and the SCAD parameters γ_1 and γ_2 , which are clearly problem dependent. Our choice in this paper could be far from optimal for certain specific instances. Another problem is how to determine a ‘better’ DC decomposition of the objective function. Our choice in (2.15) was motivated by the easiness of subproblem and better convergence properties of the whole algorithm. In fact, this problem is known as regularization technique in DC programming and kept open for a long time. Moreover, the theoretical properties of the TVSCAD model is definitely very important in understanding its performance and deserves further investigations. Related results for regularized least squares problem have been presented in [20, 21, 39, 40]. We leave these issues to further research.

Acknowledgments

We thank Prof. Minru Bai of Hunan University for providing the CTVL1 codes of [8] and the helpful discussions. We also thank two anonymous referees for their helpful comments. G. Gu and J. Yang are supported by National Natural Science Foundation of China (NSFC-11371192, 11771208, 11671195) and Fundamental Research Funds for the Central Universities. S. Jiang is supported by National Natural Science Foundation of China (NSFC-71671080, 71602083).

References

- [1] Acar R and Vogel C R 1994 Analysis of bounded variation penalty methods for ill-posed problems *Inverse Problems* **10** 1217–29
- [2] Alliney S 1992 Digital filters as absolute norm regularizers *IEEE Trans. Signal Process.* **40** 1548–62
- [3] An L T H and Tao P D 2005 The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems *Ann. Oper. Res.* **133** 23–46
- [4] Astola J and Kuosmanen P 1997 *Fundamentals of Nonlinear Digital Filtering* (Boca Raton, FL: CRC Press)
- [5] Attouch H and Bolte J 2009 On the convergence of the proximal algorithm for nonsmooth functions involving analytic features *Math. Program.* **116** 5–16
- [6] Attouch H, Bolte J and Svaiter B F 2013 Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss–Seidel methods *Math. Program.* **137** 91–129
- [7] Attouch H, Bolte J, Redont P and Soubeyran A 2010 Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka–Łojasiewicz inequality *Math. Oper. Res.* **35** 438–57
- [8] Bai M, Zhang X and Shao Q 2016 Adaptive correction procedure for TVL1 image deblurring under impulse noise *Inverse Problems* **32** 085004, 23
- [9] Bovik A C 2010 *Handbook of Image and Video Processing* (New York: Academic)
- [10] Boyd S, Parikh N, Chu E, Peleato B and Eckstein J 2011 Distributed optimization and statistical learning via the alternating direction method of multipliers *Found. Trends® Mach. Learn.* **3** 1–122
- [11] Burger M and Osher S 2013 *A Guide to the TV Zoo. In Level Set and PDE Based Reconstruction Methods in Imaging* (Berlin: Springer) pp 1–70
- [12] Cai J F, Chan R H and Nikolova M 2008 Two-phase approach for deblurring images corrupted by impulse plus Gaussian noise *Inverse Problems Imaging* **2** 187–204
- [13] Cai J F, Chan R H and Nikolova M 2010 Fast two-phase image deblurring under impulse noise *J. Math. Imaging Vis.* **36** 46–53

- [14] Chan R H, Tao M and Yuan X 2013 Constrained total variation deblurring models and fast algorithms based on alternating direction method of multipliers *SIAM J. Imaging Sci.* **6** 680–97
- [15] Chan T F and Esedoğlu S 2005 Aspects of total variation regularized L^1 function approximation *SIAM J. Appl. Math.* **65** 1817–37
- [16] Chen T and Wu H R 2001 Space variant median filters for the restoration of impulse noise corrupted images *IEEE Trans. Circuits Syst. II* **48** 784–9
- [17] Darbon J and Sigelle M 2006 Image restoration with discrete constrained total variation. I. Fast and exact optimization *J. Math. Imaging Vis.* **26** 261–76
- [18] Dobson D C and Santosa F 1996 Recovery of blocky images from noisy and blurred data *SIAM J. Appl. Math.* **56** 1181–98
- [19] Douglas J Jr and Rachford H H Jr 1956 On the numerical solution of heat conduction problems in two and three space variables *Trans. Am. Math. Soc.* **82** 421–39
- [20] Durand S and Nikolova M 2006 Stability of the minimizers of least squares with a non-convex regularization. I. Local behavior *Appl. Math. Optim.* **53** 185–208
- [21] Durand S and Nikolova M 2006 Stability of the minimizers of least squares with a non-convex regularization. II. Global behavior *Appl. Math. Optim.* **53** 259–77
- [22] Eckstein J and Yao W 2011 Augmented Lagrangian and alternating directions methods for convex optimization: a tutorial and some illustrative computational results *RUTCOR Research Reports*
- [23] Eckstein J and Bertsekas D P 1992 On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators *Math. Program.* **55** 293–318
- [24] Eng H L and Ma K K 2001 Noise adaptive soft-switching median filter *IEEE Trans. Image Process.* **10** 242–51
- [25] Esser E 2009 Applications of lagrangian-based alternating direction methods and connections to split bregman *Technical Report* CAM 09-31 UCLA
- [26] Fan J and Li R 2001 Variable selection via nonconcave penalized likelihood and its oracle properties *J. Am. Stat. Assoc.* **96** 1348–60
- [27] Fu H, Ng M K, Nikolova M and Barlow J L 2006 Efficient minimization methods of mixed l_2 - l_1 and l_1 - l_1 norms for image restoration *SIAM J. Sci. Comput.* **27** 1881–902
- [28] Gabay D and Mercier B 1976 A dual algorithm for the solution of nonlinear variational problems via finite element approximation *Comput. Math. Appl.* **2** 17–40
- [29] Gasso G, Rakotomamonjy A and Canu S 2009 Recovering sparse signals with a certain family of nonconvex penalties and DC programming *IEEE Trans. Signal Process.* **57** 4686–98
- [30] Glowinski R and Marrocco A 1975 Sur l’approximation, par éléments finis d’ordre un, et la résolution, par pénalisation-dualité, d’une classe de problèmes de Dirichlet non linéaires *RAIRO* **9** 41–76
- [31] Goldstein T and Osher S 2009 The split bregman method for l_1 -regularized problems *SIAM J. Imaging Sci.* **2** 323–43
- [32] Hestenes M R 1969 Multiplier and gradient methods *J. Optim. Theory Appl.* **4** 303–20
- [33] Horst R and Thoai N V 1999 DC programming: overview *J. Optim. Theory Appl.* **103** 1–43
- [34] Hwang H and Haddad R A 1995 Adaptive median filters: new algorithms and results *IEEE Trans. Image Process.* **4** 499–502
- [35] Li G Y and Pong T K 2016 Calculus of the exponent of the kurdyka-lojasiewicz inequality and its applications to linear convergence of first-order methods *Found. Comput. Math.* submitted (<https://doi.org/10.1007/s10208-017-9366-8>)
- [36] Mumford D and Shah J 1989 Optimal approximations by piecewise smooth functions and associated variational problems *Commun. Pure Appl. Math.* **42** 577–685
- [37] Nikolova M 2002 Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers *SIAM J. Numer. Anal.* **40** 965–94 (electronic)
- [38] Nikolova M 2004 A variational approach to remove outliers and impulse noise *J. Math. Imaging Vis.* **20** 99–120 (Special issue on mathematics and image analysis)
- [39] Nikolova M 2005 Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares *Multiscale Model. Simul.* **4** 960–91 (electronic)
- [40] Nikolova M 2008 Analytical bounds on the minimizers of (nonconvex) regularized least-squares *Inverse Problems Imaging* **2** 133–49
- [41] Pok G, Liu J C and Nair A S 2003 Selective removal of impulse noise based on homogeneity level information *IEEE Trans. Imag. Process.* **12** 85–92
- [42] Powell M J D 1969 A method for nonlinear constraints in minimization problems *Optimization* (London: Academic) pp 283–98 (*Symp. University Keele (Keele, 1968)*)

- [43] Rockafellar R T *Convex Analysis (Princeton Mathematical Series vol 28)* (Princeton: Princeton University Press)
- [44] Rockafellar R T and Wets R J B 1998 *Variational Analysis (Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences) vol 317)* (Berlin: Springer)
- [45] Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica D* **60** 259–68
- [46] Tao P D 1984 Convergence of a subgradient method for computing the bound norm of matrices *Linear Algebr. Appl.* **62** 163–82
- [47] Tao P D 1985 Algorithmes de calcul du maximum des formes quadratiques sur la boule unité de la norme du maximum *Numer. Math.* **45** 377–401
- [48] Tao P D and An L T H 1997 Convex analysis approach to dc programming: theory, algorithms and applications *Acta Math. Vietnamica* **22** 289–355
- [49] Tao P D and An L T H 1998 A d.c. optimization algorithm for solving the trust-region subproblem *SIAM J. Optim.* **8** 476–505
- [50] Tao P D *et al* 1986 Algorithms for solving a class of nonconvex optimization problems. Methods of subgradients *North-Holland Math. Stud.* **129** 249–71
- [51] Tao P D *et al* 1988 Duality in dc (difference of convex functions) optimization. Subgradient methods *Trends in Mathematical Optimization* (Basel: Birkhäuser) pp 277–93
- [52] Tikhonov A N, Arsenin V I and John F 1977 *Solution of Ill-Posed Problems* vol 14 (Washington, DC: Winston)
- [53] Toland J F 1978 Duality in nonconvex optimization *J. Math. Anal. Appl.* **66** 399–415
- [54] Toland J F 1979 On subdifferential calculus and duality in nonconvex optimization *Bull. Soc. Math. France Mém.* **177–83** (Analyse non convexe (Proc. Colloq., Pau, 1977))
- [55] Wang Y, Yang J, Yin W and Zhang Y 2008 A new alternating minimization algorithm for total variation image reconstruction *SIAM J. Imaging Sci.* **1** 248–72
- [56] Yang J, Yin W, Zhang Y and Wang Y 2009 A fast algorithm for edge-preserving variational multichannel image restoration *SIAM J. Imaging Sci.* **2** 569–92
- [57] Yang J, Zhang Y and Yin W 2009 An efficient TVL1 algorithm for deblurring multichannel images corrupted by impulsive noise *SIAM J. Sci. Comput.* **31** 2842–65
- [58] Yin W, Goldfarb D and Osher S 2007 The total variation regularized L^1 model for multiscale decomposition *Multiscale Model. Simul.* **6** 190–211 (electronic)

