Journal of Scientific Computing (2019) 78:1063–1086
https://doi.org/10.1007/s10915-018-0801-z

A Nonconvex Model with Minimax Concave Penalty for Image Restoration

Juntao You[1] · Yuling Jiao[2] · Xiliang Lu[1,3] · Tieyong Zeng[4]

Received: 22 February 2018 / Revised: 26 July 2018 / Accepted: 30 July 2018 /
Published online: 20 August 2018

Abstract
A natural image u is often sparse under a given transformation W, one can use L_0 norm of Wu as a regularisation term in image reconstructions. Since minimizing the L_0 norm is a NP hard problem, the L_1 norm is widely used as an replacement. However, recent studies show that nonconvex penalties, e.g., MCP, enjoy better performance for sparse signal recovery. In this paper, we propose a nonconvex model for image restoration with a minimax concave penalty (MCP). First we establish the existence of a global minimizer for the nonconvex model. Then we solve this model by using the alternating direction method of multipliers algorithm. The convergence of the proposed algorithm is analysed with properly chosen parameters. Numerical experiments show that the MCP model outperforms TV model in terms of efficiency and accuracy.

Keywords Nonconvex model · Image restoration · Regularization · ADMM · Convergence

✉ Tieyong Zeng
zeng@math.cuhk.edu.hk

Juntao You
jyouab@connect.ust.hk

Yuling Jiao
yulingjiaomath@whu.edu.cn

Xiliang Lu
xllv.math@whu.edu.cn

[1] School of Mathematics and Statistics, Wuhan University, Wuhan 430072, Hubei, People's Republic of China

[2] School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan 430072, Hubei, People's Republic of China

[3] Hubei Key Laboratory of Computational Science (Wuhan University), Wuhan 430072, Hubei, People's Republic of China

[4] Department of Mathematics, The Chinese University of Hong Kong, Shatin, NT, Hong Kong

1 Introduction

In this paper we consider the image restoration problem with the form

$$f = Au + \eta,$$

where $f \in \mathbb{R}^{M\times N}$ is the observed image, and u is the image which we want to reconstruct. The linear operator A can be identity (denoising), convolution (deblurring) or other linear transforms and η is the Gaussian additive noise. Due to the unknown noise term η and possible bad condition number of the operator A, finding the desired image u is ill-posed. One common technique to handle the ill-posedness is to consider the variational regularization model

$$\underset{u}{\text{minimize}} \ \frac{\mu}{2}\|Au - f\|_F^2 + R(u), \tag{1}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. The fidelity term $\|Au - f\|_F^2$ concerns the noise, the regularization term $R(u)$ is related to the a priori information of the target image, and $\mu > 0$ is a regularization parameter to balance these two terms. It is known that a natural image is (approximated) sparse after certain transformations, e.g., gradient operator [43] or wavelet transform [4,10,14,34]. The regularization term $R(u)$ should take such properties into account. Suppose W is a given linear operator such that u is sparse under the transformation, i.e. Wu is sparse, then we may choose $R(u) = \|Wu\|_0$ to ensure the sparseness. However to find a global minimizer of $\|\cdot\|_0$ regularized problem is NP hard, one popular approach is to replace $\|\cdot\|_0$ by $\|\cdot\|_1$, see e.g., [7,12,16,48]. If we let $W = \nabla$, where ∇ is the gradient operator, then we have $R(u) = \|\nabla u\|_1$, which is the famous ROF model which is proposed by Rudin–Osher–Fatemi [43].

In this paper, instead of using the L_1 norm in the regularization term $R(u)$, we approximate the L_0 norm by nonconvex model. Some recent studies show that nonconvex approximations of the L_0 norm are more efficient, and the nonconvex penalties in image restoration usually have better performances than L_1 penalty [23,33,38,39]. This is because the L_1 regularized model induces bias for large coefficients [51], and hence lacks of oracle property [19,21]. There are some choices of the nonconvex penalties, such as bridge [22,24], capped-L_1 [52], smoothly clipped absolute deviation (SCAD) [20], minimax concave penalty (MCP) [50]. In particular, MCP has been shown to enjoy a nice performance when it approximates L_0 norm [31]. In this work we will use the MCP penalty to replace the L_0 norm and the corresponding variational model reads

$$\underset{u}{\text{minimize}} \ \frac{\mu}{2}\|Au - f\|_F^2 + \|Wu\|_{\text{MCP}}. \tag{2}$$

First, we will prove the existence of the global minimizer of the proposed model (2), which is not trivial due to lack of coercivity of the cost functional. Most existence results of a global minimizer require the coercivity of the objective functional, including convex models [9, p. 14] and nonconvex models [13]. Our approach to prove the existence of a global minimizer has the advantage that it can be applied to a class of nonconvex penalized models which do not necessarily have coercive penalties.

Second, we give an efficient algorithm to solve the model (2). In literatures several algorithms such as Local Quadratic Approximation (LQA) [20], Local Linear Approximation (LLA) [53], muti-stage convex relaxtion [52] and Difference of Convex Functions Algorithm (DCA) [33,46,47] have been proposed to solve it. All these methods introduce a sequence of convex subproblems to approach the original nonconvex problem, and each subproblem can be approximately solved by the alternating direction method of multipliers (ADMM) [3].

Recently, people use ADMM to solve nonconvex models directly, and the global convergence of the algorithm for a class of nonconvex minimization problems [15,30,49] has been proved with some extra conditions.

Inspired by the previous work, we use ADMM to solve the proposed minimization problem. ADMM, which is also known as Split Bregman method [5,29] in the community of the mathematics image processing [18] was proposed in [26,28] around the mid-1970 and analysis of this algorithm can be found in [17,25,32,41,45]. It has been widely used in signal/image processing, statistics, machine learning, and so on. It's a favorable algorithm in various kinds of models in image processing for its flexibility and efficiency [6,27,44]. To the best of our knowledge, the proof of the convergence of ADMM needs some assumptions on the operator W, i.e., W has a full row rank (e.g., [15]). However, this assumption does not hold for the gradient operator or the framelet transformations, and thus these analysis can not be applied to our model directly. We consider an ADMM similar to [41], and prove the global subsequential convergence of the proposed ADMM for (2) under the mild condition $\ker(A) \cap \ker(W) = \{0\}$ and with properly chosen parameters.

Last, several numerical examples are given to show the efficiency of the MCP model and ADMM algorithm. We fix W as the gradient operator, and compare the MCP model with the TV model. Both models have almost the same computational complexity when invoking the ADMM algorithm. But MCP model has better performance of the restored images.

The following paper is organized as follows. We will propose the MCP-penalized model and then prove the existence of global minimizers in Sect. 2. In Sect. 3 we apply the ADMM algorithm to solve the proposed nonconvex model and discuss the convergence of the algorithm. In Sect. 4, some numerical experiments are given to depict the efficiency of the proposed model and algorithm.

2 MCP-Penalized Model

In this section, we will propose the MCP penalized model and prove the existence of the global minimizer.

2.1 Model Settings

The MCP function introduced in [50] is defined by

$$p_{\alpha,\beta}(x) = \int_0^{|x|} \left(\alpha - \frac{|t|}{\beta}\right)_+ dt \ , \ x \in \mathbb{R} \tag{3}$$

where $\alpha > 0$, $\beta > 1$ are two constants. Figure 1 shows the different pictures of L_0, L_1 and MCP penalties. From the definition we see that the MCP is between L_0 and L_1 penalties. It is clear that the MCP function is continuous, sparsity promoting and unbiasedness, which guarantee the MCP regularized problem have a good statistical property [20,50].

To solve the minimization problem with a penalty function $p_{\alpha,\beta}(x)$, i.e.

$$\min_{x\in\mathbb{R}} p_{\alpha,\beta}(x) + \frac{1}{2}(x-y)^2,$$

$y \in \mathbb{R}$, we recall the important proximity operator introduced by Moreau in [36,37], and similar to the soft-thresholding operator for L_1 penalty [48], the thresholding operator for the MCP penalty is defined as follows, see [50].

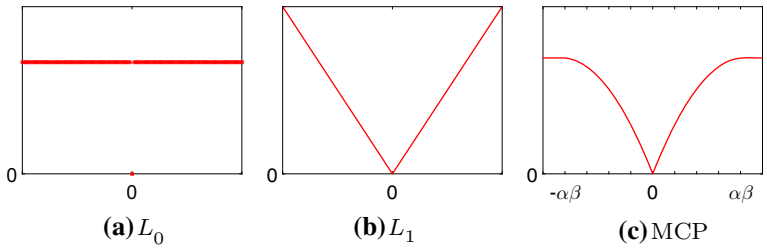

Fig. 1 The penalty functions L_0, L_1 and MCP

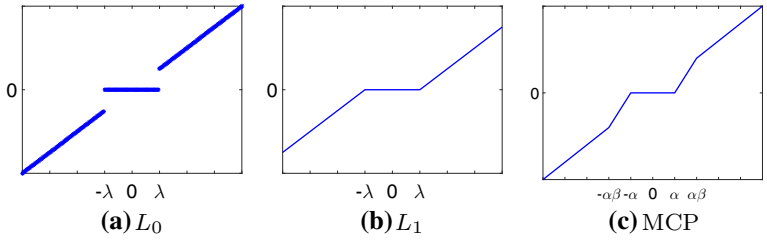

Fig. 2 The thresholding functions of L_0, L_1 and MCP penalties

Definition 1 Let $p_{\alpha,\beta}$ as in (3), the thresholding operator $T_{\alpha,\beta}$ is defined by

$$T_{\alpha,\beta}(y) = Prox_{p_{\alpha,\beta}}(y) = \underset{x}{\operatorname{argmin}} \left\{ p_{\alpha,\beta}(x) + \frac{1}{2}(x-y)^2,\ y \in \mathbb{R} \right\}.$$

One can find that for $y \in \mathbb{R}$,

$$Prox_{p_{\alpha,\beta}}(y) = \begin{cases} 0 & |y| \le \alpha, \\ \operatorname{sgn}(y)\frac{\beta(|y|-\alpha)}{\beta-1} & \alpha < |y| \le \alpha\beta, \\ y & |y| > \alpha\beta. \end{cases}$$

Figure 2 shows the thresholding operators corresponding to the L_0, L_1 and MCP penalties, respectively.

Now we can introduce the MCP-penalized model. Let the image u be a 2-dimensional matrix with size $M \times N$. Given a linear transform $W : \mathbb{R}^{MN} \to \mathbb{R}^{s\times P}$, suppose u is sparse under the transform W, we recall (2) as the MCP-penalized model:

$$\text{minimize} \left\{ \frac{\mu}{2}\|Au - f\|_F^2 + \|Wu\|_{\text{MCP}} \right\} \tag{4}$$

where the function $\|\cdot\|_{\text{MCP}}$ is defined by $\|Wu\|_{\text{MCP}} = \sum_i p_{\alpha,\beta}(\|(Wu)_i\|_2)$, and $\|(Wu)_i\|_2$ is the Euclidean norm: define $v := Wu := (v_{t,i})_{1\le t\le s, 1\le i\le P} \in \mathbb{R}^{s\times P}$, and we note $(Wu)_i := v_i := \begin{pmatrix} v_{1,i} \\ v_{2,i} \\ \cdots \\ v_{s,i} \end{pmatrix} \in \mathbb{R}^s$ in our paper, then

$$\|v\|_{\text{MCP}} := \sum_{1\le i\le P} p_{\alpha,\beta}\left(\sqrt{|v_{1,i}|^2 + |v_{2,i}|^2 + \cdots + |v_{s,i}|^2} \right). \tag{5}$$

By the above definition we know that $\|\cdot\|_{\mathrm{MCP}}$ is separable for $i \in P$, and if we choose $s = 1$, then $\|\cdot\|_{\mathrm{MCP}}$ is separable. The linear operator W can be chosen as the gradient operator, the wavelet (framelet) transform, or other linear transform. We will then use the gradient operator as an example to explain the model.

The standard isotropic TV regularisation as in [8] reads:

$$\|u\|_{\mathrm{TV}} = \sum_{1\le i\le M, 1\le j\le N} \sqrt{|(\nabla_x u)_{i,j}|^2 + |(\nabla_y u)_{i,j}|^2}, \tag{6}$$

where $(\nabla_x u)_{i,j}$ and $(\nabla_y u)_{i,j}$ denote the horizontal and vertical first order finite differences at pixel (i, j) respectively, i.e.,

$$(\nabla_x u)_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j} & \text{if } i < M, \\ u_{1,j} - u_{M,j} & \text{if } i = M, \end{cases}$$

and

$$(\nabla_y u)_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j} & \text{if } j < N, \\ u_{i,1} - u_{i,N} & \text{if } j = N. \end{cases}$$

If we choose $W = \nabla$, then $Wu = \nabla u = \begin{pmatrix} \nabla_x u \\ \nabla_y u \end{pmatrix} \in \mathbb{R}^{2\times MN}$, $\nabla_x u \in \mathbb{R}^{MN}$, $\nabla_y u \in \mathbb{R}^{MN}$, and hence

$$\|\nabla u\|_{\mathrm{MCP}} = \sum_{i,j} p_{\alpha,\beta}\left(\sqrt{|(\nabla_x u)_{i,j}|^2 + |(\nabla_y u)_{i,j}|^2}\right). \tag{7}$$

Notice that in this example, we have $s = 2$, $P = M \times N$ and $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$, $v_1 = \nabla_x u$, $v_2 = \nabla_y u$ as a special case for our general model.

2.2 Existence of a Global Minimizer

Before we propose an algorithm to solve the model (4), we'd like to show the existence of the minimizer in this subsection. We know that our cost functional $F_p(u) = \frac{\mu}{2}\|Au - f\|_F^2 + \|Wu\|_{\mathrm{MCP}}$ is not coercive, in fact, we can choose a sequence $\{u^k\}_{k=0}^{\infty}$ with some nonzero $u^0 \in \ker A$ and $u^k = ku^0$, such that $\|u^k\|_F \to \infty$ as $k \to \infty$, but

$$F_p(u^k) = \frac{\mu}{2}\|f\|_F^2 + \|Wu^k\|_{\mathrm{MCP}} < \infty.$$

This is because $\|Wu\|_{\mathrm{MCP}}$ is always bounded by the definition of $\|\cdot\|_{\mathrm{MCP}}$. Due to the non-coercivity of $F_p(u)$, we can not conclude the boundedness of the minimizing sequence, and thus the existence of the minimizer is not straightforward. Moreover, to extend our result, we would like to give a proof for the result in a more general sense.

Suppose ϕ is a nonnegative lower semicontinuous function defined on $\mathbb{R}$, and it satisfies $\phi(0) = 0$, $\phi(-x) = \phi(x)$, $\phi(x_1) \le \phi(x_2)$ if $|x_1| < |x_2|$.

There are a large class of functions ϕ satisfy the above assumptions, for example:

- bridge [22,24]: $\phi(x) = |x|^\tau$, for $0 < \tau \le 1$;
- capped-L_1 [52]: for $\alpha > 0$, $\beta > \frac{1}{2}$,

$$\phi(x) = \begin{cases} \alpha|x| & \text{if } x \le \alpha\beta, \\ \alpha^2\beta & \text{if } x > \alpha\beta\ ; \end{cases}$$

– logistic penalty [40]: $\phi(x) = \log(1 + \alpha|x|)$, for $\alpha > 0$;
– smoothly clipped absolute deviation (SCAD) [20]: $\phi(x) = \int_0^{|x|} \min(1, (\alpha - \frac{t}{\beta})_+/(\alpha - 1))dt$ for $\alpha > 2$;
– minimax concave penalty (MCP) [50]: $\phi(x) = p_{\alpha,\beta}(x)$.

For some $v \in \mathbb{R}^{s\times P}$, we define a regularization function by $\Phi(v) = \sum_{1\le i\le P} \phi(\|v_i\|_2)$, where $v_i \in \mathbb{R}^s$. Notice that our notation cover the case when $\Phi(v)$ is separable by choosing $s = 1$, then there is $\phi(\|v_i\|_2) = \phi(v_i)$ and $\Phi(v) = \sum_{1\le i\le P} \phi(v_i)$. Now we are going to prove the following theorem.

Theorem 1 *If Φ is a function defined above, and assume* $\ker(A) \cap \ker(W) = \{0\}$*, then there exists a global minimizer to the problem*

$$\text{minimize } \left\{ F_\Phi(u) := \frac{\mu}{2}\|Au - f\|_F^2 + \Phi(Wu) \right\}. \tag{8}$$

To proof the theorem, we begin with the special case where ϕ is coercive.

Proposition 1 *If ϕ is coercive, then problem (8) has a global minimizer.*

Proof The proof is rather standard. Let $\{u^k\}_{k=0}^\infty$ be a minimizing sequence of (8). We know that if $\phi(\cdot)$ is coercive, then $\Phi(\cdot)$ is also coercive. By $F_\Phi(u^k) \le F_\Phi(0) < \infty$ we then know $\{\|Au^k\|_F\}_{k\ge1}$ and $\{\|Wu^k\|_F\}_{k\ge1}$ are bounded, hence $\{\|u^k\|_F\}_{k\ge1}$ is bounded because $\ker(A) \cap \ker(W) = \{0\}$. It follows that there exist a subsequence $\{u^{k_i}\}_{i\ge1}$ which convergent to some $u_{\min}$. By the lower semicontinuity of $\Phi(\cdot)$, we have

$$F_\Phi(u_{\min}) \le \liminf_{i\to\infty} \left\{ \frac{\mu}{2}\|Au^{k_i} - f\|_F^2 + \Phi(Wu^{k_i}) \right\} = \inf F_\Phi(u).$$

So $u_{\min}$ is a minimizer. □

For the case where ϕ is not coercive, we start from a technical Lemma as below.

Lemma 1 *Let $\mathcal{N} \subset \mathbb{R}^{s\times P}$ be a subspace of $\mathcal{H} = \mathbb{R}^{s\times P}$, then for any $q \in \mathbb{R}^{s\times P}$, there exist a minimizer of*

$$\min_{h\in\mathcal{N}} \Phi(q + h), \tag{9}$$

where $q = (q_{t,i})_{1\le t\le s, 1\le i\le P}$, $h = (h_{t,i})_{1\le t\le s, 1\le i\le P} \in \mathbb{R}^{s\times P}$ and

$$\Phi(q + h) = \sum_{1\le i\le P} \phi\left(\sqrt{|q_{1,i} + h_{1,i}|^2 + |q_{2,i} + h_{2,i}|^2 + \cdots + |q_{s,i} + h_{s,i}|^2} \right),$$

for ϕ non-coercive. Moreover, for any $q \in \mathbb{R}^{s\times P}$, we can find a solution $h(q)$ of (9) which is bounded while q is bounded.

Proof Let $\{h^k\}_{k=0}^\infty$ be a minimizing sequence of (9), where $h^k = (h_i^k)_{i\in P} \in \mathbb{R}^{s\times P}$ for $h_i^k \in \mathbb{R}^s$. If there exists a bounded subsequence of $\{h^k\}_{k=0}^\infty$, similar to the proof of Proposition (1), we can immediately conclude the existence of the minimizer. Hence we assume $\|h^k\|_F \to \infty$ when $k \to \infty$.

Since $\mathcal{N}$ is a subspace of $\mathbb{R}^{s\times P}$, let d_0 be the dimension of $\mathcal{N}$, we can find a matrix $\mathcal{M} \in \mathbb{R}^{sP\times d_0}$ whose column vectors form a basis of $\mathcal{N}$. Then any $h \in \mathcal{N}$ can be uniquely represented as $h = \mathcal{M}w$ for some $w \in \mathbb{R}^{d_0}$. Define the index set I and J by

$$I \triangleq \{i \in P : \|h_i^k\|_2^2 \text{ is bounded for any } k\},$$
$$J \triangleq \{i \in P : \|h_i^k\|_2^2 \to \infty \text{ as } k \to \infty\},$$

where $h_i^k \in \mathbb{R}^s$. Define a submatrix $\mathcal{M}_I$, which satisfies $\mathcal{M}_I := \begin{pmatrix} \mathcal{M}_{1,i} \\ \mathcal{M}_{2,i} \\ \cdots \\ \mathcal{M}_{s,i} \end{pmatrix}_{i\in I}, \mathcal{M}_{1,i}, \ldots,$ $\mathcal{M}_{s,i}$ be the rows of $\mathcal{M}$. So it means we choose the rows of $\mathcal{M}$ corresponding to $i \in I$. Let $\{w^k\}_{k=0}^{\infty}$ be the sequence that has the representation $h^k = \mathcal{M}w^k$ for $\{h^k\}_{k=0}^{\infty}$. Then $h_I^k = (h_i^k)_{i\in I}$ can be written as $h_I^k = \mathcal{M}_I w^k$. Now consider a minimization problem

$$\text{minimize } \|z\|_F \quad \text{s.t. } \mathcal{M}_I z = \mathcal{M}_I w^k, \tag{10}$$

whose solution z^k is $\mathcal{M}_I^+(\mathcal{M}_I w^k)$, where $\mathcal{M}_I^+$ is the Moore-Penrose pseudoinverse of $\mathcal{M}_I$. From the definition of index set I, we know $\{\mathcal{M}_I w^k\}_{k=1}^{\infty}$ is bounded. Hence by $\|z^k\|_F \leq \|\mathcal{M}_I^+\|\|\mathcal{M}_I w^k\|_F$, $\{z^k\}_{k=1}^{\infty}$ be bounded as well. For a new sequence $\{\hat{h}^k\}_{k=1}^{\infty}$, where $\hat{h}^k := \mathcal{M}z^k$, we shall see

$$\hat{h}_I^k = \mathcal{M}_I z^k = \mathcal{M}_I w^k = h_I^k.$$

Thus we have

$$\|q_i + \hat{h}_i^k\|_2 = \|q_i + h_i^k\|_2 \text{ for } i \in I,$$

together with $\|q_i + \hat{h}_i^k\|_2 \leq \infty$ for $i \in J$, we have

$$\|q_i + \hat{h}_i^k\|_2 \leq \|q_i + h_i^k\|_2 \text{ for all } i \in P.$$

So by the monotonicity of $\phi(\cdot)$, we conclude that $\hat{h}^k$ is also a minimizing sequence. Hence we can prove the existence of a minimizer by the boundedness of $\{\hat{h}^k\}_{k=1}^{\infty}$.
Next we show that there exist a solution $h(q)$ of (9) such that when q is bounded then $h(q)$ is bounded as well. Because $\phi(\cdot)$ is non-coercive, from the definition of $\phi(\cdot)$, we know there exist a constant $D > 0$, satisfying $\phi(t) \equiv C$ for $t \geq D$, where C is a constant. Define

$$\overline{I} \triangleq \left\{ i \in P : \left\| h_i^k \right\|_2^2 \leq 2\left(\|q_i\|_2^2 + D^2\right) + 1 \right\},$$
$$\overline{J} \triangleq \left\{ i \in P : \left\| h_i^k \right\|_2^2 > 2\left(\|q_i\|_2^2 + D^2\right) + 1 \right\},$$

where $h_i^k, q_i^k \in \mathbb{R}^s$. The definition of $\overline{I}$ and $\overline{J}$ gives: for any $i \in \overline{J}$, we have

$$\left\| h_i^k + q_i \right\|_2^2 \geq \frac{1}{2} \left\| h_i^k \right\|_2^2 - \|q_i\|_2^2 > D^2.$$

Thus $\phi(h_i^k + q_i) \equiv$ constant, for any $i \in \overline{J}$. Similar to (10) we can find $\overline{z}^k$ and $\overline{h}_k := \mathcal{M}\overline{z}^k$ satisfying

$$\|\overline{h}_k\|_F \leq \|\mathcal{M}\|\|\mathcal{M}_{\overline{I}}^+\|\|\mathcal{M}_{\overline{I}} w^k\|_F,$$

notice that the last term $\mathcal{M}_{\overline{I}} w^k$ is $(h_i^k)_{i\in\overline{I}}$, then by the definition of $\overline{I}$, we know $\mathcal{M}_{\overline{I}} w^k$ is bounded when $\|q\|_F$ is bounded. Assume $\|q\|_F$ is bounded by $C(q) < \infty$, then $\|\mathcal{M}\|\|\mathcal{M}_{\overline{I}} w^k\|_F$ is bounded by some $\overline{C}(q) < \infty$. Thus we have

$$\|\overline{h}_k\|_F \leq \overline{C}(q)\|\mathcal{M}_{\overline{I}}^+\|.$$

By the definition of $\overline{I}$, $\overline{J}$ and $\overline{h}^k$, and similar to the discussion of $\hat{h}^k$, we have $\phi(q_i + \overline{h}_i^k) \leq \phi(q_i + h_i^k)$ for all $i \in P$. Thus $\overline{h}^k$ is a minimizing sequence, passage to a convergent subsequence, let $k \to \infty$, the limit point $h(q)$ is the solution of (9), it satisfies

$$\|h(q)\|_F \leq \liminf_{k\to\infty} \|\overline{h}^k\|_F \leq \overline{C}(q) \sup_{\overline{I}} \|\mathcal{M}_{\overline{I}}^{+}\|,$$

which is bounded when q is bounded. □

Now we can prove Theorem 1:

Proof (to Theorem 1) Define $\mathcal{K} = \ker(A)$, and let $\mathcal{N} = W\mathcal{K}$. Suppose $\{u^k\}_{k=1}^{\infty}$ is a minimizing sequence of (8). u^k can be decomposed to

$$u^k = u_{\mathcal{K}}^k + u_{\mathcal{K}^\perp}^k,$$

where $u_{\mathcal{K}}^k \in \ker(A)$, and $u_{\mathcal{K}^\perp}^k \in \ker(A)^\perp$. Thus $Au^k = Au_{\mathcal{K}^\perp}^k$ and $\|u_{\mathcal{K}^\perp}^k\|_F \leq \|A^+\| \|Au^k\|_F$. Therefore, $\{u_{\mathcal{K}^\perp}^k\}_{k=1}^{\infty}$ is bounded, hence $\{Wu_{\mathcal{K}^\perp}^k\}_{k=1}^{\infty}$ is also bounded.
Choose q to be $q^k := Wu_{\mathcal{K}^\perp}^k$ for problem 9, and the corresponding solution is $h(q^k) \in \mathcal{N}$ as in Lemma 1, thus we have $\widetilde{h}^k := h(q^k)$ is bounded. Passage to a subsequence, it follows

$$u_{\mathcal{K}^\perp}^k \to u_{\mathcal{K}^\perp}^{\min} \in \ker(A)^\perp,\ q^k \to q^{\min} = Wu_{\mathcal{K}^\perp}^{\min},\ \widetilde{h}^k \to h^{\min} \in \mathcal{N},$$

when $k \to \infty$. Since $h^{\min} \in \mathcal{N}$, there exist a $u_{\mathcal{K}}^{\min} \in \ker(A)$, satisfying $Wu_{\mathcal{K}}^{\min} = h^{\min}$. Then by $\ker(A) \cap \ker(W) = \{0\}$, we have for some $\infty > c > 0$,

$$\left\|u_{\mathcal{K}}^{\min}\right\|_F^2 \leq c\left(\|Au_{\mathcal{K}}^{\min}\|_F^2 + \|Wu_{\mathcal{K}}^{\min}\|_F^2\right) = c\|Wu_{\mathcal{K}}^{\min}\|_F^2 = c\|h^{\min}\|_F^2.$$

Let $u^{\min} = u_{\mathcal{K}^\perp}^{\min} + u_{\mathcal{K}}^{\min}$, then $u^{\min}$ is bounded, by the lower semicontinuity of $\Phi(\cdot)$ we have

$$\begin{aligned}
\frac{\mu}{2}\|Au^{\min} - f\|_F^2 + \Phi\left(Wu^{\min}\right) &= \frac{\mu}{2}\|Au_{\mathcal{K}^\perp}^{\min} - f\|_F^2 + \Phi\left(Wu_{\mathcal{K}}^{\min} + Wu_{\mathcal{K}^\perp}^{\min}\right) \\
&\leq \liminf_{k\to\infty}\left\{\frac{\mu}{2}\|Au_{\mathcal{K}^\perp}^k - f\|_F^2 + \Phi\left(\widetilde{h}^k + Wu_{\mathcal{K}^\perp}^k\right)\right\} \\
&\leq \liminf_{k\to\infty}\left\{\frac{\mu}{2}\|Au_{\mathcal{K}^\perp}^k + Au_{\mathcal{K}}^k - f\|_F^2 + \Phi\left(\widetilde{h}^k + q^k\right)\right\} \\
&\leq \liminf_{k\to\infty}\left\{\frac{\mu}{2}\|Au_{\mathcal{K}^\perp}^k + Au_{\mathcal{K}}^k - f\|_F^2 + \Phi\left(Wu_{\mathcal{K}}^k + q^k\right)\right\} \\
&\leq \liminf_{k\to\infty}\left\{\frac{\mu}{2}\|Au^k - f\|_F^2 + \Phi(Wu^k)\right\},
\end{aligned}$$

where $q^k = Wu_{\mathcal{K}^\perp}^k$, the second last inequality comes from the definition of the $\widetilde{h}^k$, thus $u^{\min}$ is a minimizer of (8). □

3 Numerical Algorithm

3.1 First Order Necessary Condition

We first recall the definition of subdifferential for a general (nonconvex) function. Let $\phi : \mathbb{R}^{m\times n} \to (-\infty, \infty]$ be an extended-real-valued proper function, i.e. $\psi(x) > -\infty$ for all

$x \in \mathbb{R}^n$ and $\mathrm{dom}\psi := \{x \in \mathbb{R}^n | \ \psi(x) < \infty\}$ is nonempty, the subdifferential of ϕ at $x \in \mathrm{dom} f$ is defined as [49]

$$\partial\psi(x) := \left\{ d \in \mathbb{R}^n : \exists x^k \xrightarrow{\phi} x, d^k \to d \text{ with } \liminf_{y \to x^k} \frac{\psi(y) - \psi(x^k) - < d^k, y - x^k >}{\|y - x^k\|_2} \geq 0 \ \forall k \right\},$$

where $x^k \xrightarrow{\psi} x$ is defined by $x^k \to x$ and $\psi(x^k) \to \psi(x)$. One can find that the definition for the above subdifferential coincides with the classical concept of derivative or convex subdifferential when ψ is continuously differentiable or convex. A mapping $g : \mathbb{R}^{m \times n} \to \mathbb{R}$ single-valued around $\overline{x}$ is called strictly differential at $\overline{x}$ with the derivative $g'(\overline{x})$ if

$$\lim_{x \to \overline{x}, y \to \overline{x}} \frac{g(x) - g(y) - g'(\overline{x})}{\|x - y\|_2} = 0.$$

And by Theorem 3.1 in [35] we know

$$\partial(f + g)(\overline{x}) \subset \partial f(\overline{x}) + g(\overline{x}) \tag{11}$$

for $f : \mathbb{R}^{m \times n} \to \overline{\mathbb{R}}$ lower semicontinuous around $\overline{x}$ and $g : \mathbb{R}^{m \times n} \to \mathbb{R}$ strictly differential at $\overline{x}$.

By applying Theorem 10.1 in [42], we have $0 \in \partial\{\frac{\mu}{2}\|Au - f\|_F^2 + \|Wu\|_{\mathrm{MCP}}\}$ for any local minimizer u^* of problem (4). Then the first-order necessary conditions of (4) reads

$$0 \in \mu A^*(Au^* - f) + \partial\|Wu^*\|_{\mathrm{MCP}}. \tag{12}$$

For $x \in \mathbb{R}$, $p_{\alpha,\beta}$ defined in (3) can be equivalently expressed as

$$p_{\alpha,\beta}(x) = \begin{cases} \alpha|x| - \frac{1}{2\beta}x^2 & \text{if } x \leq \alpha\beta, \\ \frac{\alpha^2\beta}{2} & \text{if } x > \alpha\beta. \end{cases}$$

Thus by (11) we can easily conclude the following proposition:

Proposition 2 *If $d \in \partial p_{\alpha,\beta}(x)$, $x \in \mathbb{R}$, then d is bounded and $|d| \leq \alpha$.*

Proof For $x > \alpha\beta$, $d = 0$. For $x \leq \alpha\beta$, $|x|$ lower semicontinuous and x^2 strictly differential, from 11 we know if $d \in \partial p_{\alpha,\beta}(x)$, then $d \in \partial\alpha|x| - \frac{1}{\beta}x$, which leads to our conclusion that $|d| \leq \alpha$. □

3.2 Alternative Direction Method of Multiplier (ADMM)

Now we employ the ADMM method to solve the model. By introducing a new variable $v = Wu$, model (4) can be equivalently written as the constrained optimization problem:

$$\text{minimize } \left\{ \frac{\mu}{2}\|Au - f\|_F^2 + \|v\|_{\mathrm{MCP}} \right\} \quad \text{s.t. } v = Wu. \tag{13}$$

Then the corresponding argument Lagrangian to (13) is

$$\mathcal{L}_\rho(u, v, \xi) = \frac{\mu}{2}\|Au - f\|_F^2 + \|v\|_{\mathrm{MCP}} + \frac{\rho}{2}\|Wu - v + \frac{\xi}{\rho}\|_F^2 - \frac{\|\xi\|_F^2}{2\rho},$$

where $\xi \in \mathbb{R}^{s \times P}$ is the Lagrange multiplier, $\rho \in \mathbb{R}$ is a positive parameter. Then we employ the following ADMM (similar to [41]) to solve the problem.

Algorithm 1 ADMM solving model (13)

1: Initialize $u^0 = 0,\ v^0 = 0,\ \xi^0 = 0$;
2: **for** $k = 0; k++$ **do**
3: Compute $u^{k+1} = \underset{u}{\operatorname{argmin}} \mathcal{L}_{\rho_k}(u, v^k, \xi^k)$;
4: Compute $v^{k+1} \in \underset{v}{\operatorname{argmin}} \mathcal{L}_{\rho_k}(u^{k+1}, v, \xi^k)$;
5: Update $\xi^{k+1} = \xi^k + \rho_k(Wu^{k+1} - v^{k+1})$ and $\rho_{k+1} = \sigma\rho_k$ with $\sigma > 1$;
6: check stop criteria $\frac{\|u^{k+1}-u^k\|}{\|u^{k+1}\|} \le \delta$ for some small positive δ.
7: **end for**;
8: Output u^{k+1}

The first-order optimality conditions for the subproblems in Algorithm 1 follows:

$$0 = \mu A^*\left(Au^{k+1} - f\right) + \rho_k W^*\left(Wu^{k+1} - v^k\right) + W^*\xi^k, \tag{14}$$

$$0 \in \partial\|v^{k+1}\|_{\mathrm{MCP}} - \rho_k\left(Wu^{k+1} - v^{k+1}\right) - \xi^k, \tag{15}$$

$$\xi^{k+1} = \xi^k + \rho_k\left(Wu^{k+1} - v^{k+1}\right), \rho_{k+1} = \sigma\rho_k, \tag{16}$$

Then by (14), the u-subproblem can be written as

$$(\mu A^*A + \rho_k W^*W)u^{k+1} = \mu A^*f + \rho_k W^*v^k - W^*\xi^k, \tag{17}$$

if there is $\ker(A) \cap \ker(W) = \{0\}$, then the matrix $\mu A^*A + \rho_k W^*W$ is positive definite and hence invertible. For some operator A and transform W (e.g., convolution operator A and wavelet frame transformation W), $\mu A^*A + \rho_k W^*W$ can be diagonalized by the Fourier transform [11]. Consequently, the u-subproblem can be solved efficiently by invoking the fast Fourier transform (FFT).

For the v-subproblem, similar to the shrinkage formula in TV model [29], by invoking the corresponding proximal operator, we can employ the thresholding operator $T_{\frac{\alpha}{\rho_k},\beta}$ defined by Definition 1 for MCP, then the v-subproblem can be explicitly solved in the pointwise formula:

$$v_{t,i}^{k+1} = T_{\frac{\alpha}{\rho_k},\beta}\left(S_i^k\right)\frac{S_{t,i}^k}{S_i^k}, t = 1, 2, \ldots, s, \tag{18}$$

for all $i \in P$, where $S_{t,i}^k = (Wu^{k+1})_{t,i} + \xi_{t,i}^k/\rho_k,\ S_i^k = \sqrt{|S_{1,i}^k|^2 + |S_{2,i}^k|^2 + \cdots + |S_{s,i}^k|^2}$. Therefore, our algorithm can be efficiently implemented.

3.3 Convergence of the Algorithm

Now we discuss the convergence of the sequences generated by algorithm 1. Recently there are a lot of work on this topic. If W has a full row rank, the global convergence of the algorithm can be very similar as in [15,49], which use the Kurdyka–Łojasiewicz (KL) property [1,2] of the MCP function. However, for some transformations such as gradient operator, W is not a full row rank matrix. In this section, for a bounded linear transformation W, we will prove the sequence generated by the ADMM algorithm is subsequential convergent under certain conditions related to the choice of parameters.

Consider a stationary point (u^*, v^*, b^*), i.e., it satisfies the first-order necessary system of (13)

$$\begin{aligned} 0 &= \mu A^*(Au^* - f) + W^*\xi^*, \\ 0 &\in \partial\|v^*\|_{\mathrm{MCP}} - \xi^*, \\ 0 &= Wu^* - v^*. \end{aligned} \tag{19}$$

One can easily find that u^* satisfies

$$0 \in \mu A^*(Au^* - f) + \partial\|Wu^*\|_{\mathrm{MCP}},$$

which is the fist-order necessary conditions for problem 4. In following we will prove the sequence $\{u^k, v^k, \xi^k\}_{k=1}^{\infty}$ generated by the Algorithm 1 has a limit point (u^*, v^*, ξ^*) satisfies the system (19). For convenience, we denote by

$$\mathcal{L}^k := \mathcal{L}_{\rho_k}(u^k, v^k, \xi^k).$$

The following Lemma gives the uniform boundness for the Lagrange multiplier and the augmented Lagragian.

Lemma 2 *Suppose* $\ker(A) \bigcap \ker(W) = \{0\}$. *Let* $\{u^k, v^k, \xi^k\}_{k=1}^{\infty}$ *be a sequence generated by* (14), (15), (16). *Then* $\{\xi^k\}_{k=1}^{\infty}$ *and* $\{\mathcal{L}^k\}_{k=1}^{\infty}$ *are bounded.*

Proof Combine (16) and (15) we have

$$\xi^{k+1} \in \partial\|v^{k+1}\|_{\mathrm{MCP}},$$

Thus by the definition of $\|\cdot\|_{\mathrm{MCP}}$ in (5) and Proposition 2, we have $\|\xi^{k+1}\|_{\infty} \leq \alpha$, which implies that $\|\xi^{k+1}\|_F^2 \leq m_\alpha < \infty$, m_α is some finite constant related to α, this is because ξ^{k+1} is finite dimensional. Hence $\{\xi^k\}_{k=1}^{\infty}$ is bounded. By (16) we notice that

$$\mathcal{L}^{k+1} - \mathcal{L}_{\rho_k}\left(u^{k+1}, v^{k+1}, \xi^{k+1}\right) = \frac{\rho_{k+1} - \rho_k}{2\rho_k^2}\|\xi^{k+1} - \xi^k\|_F^2, \tag{20}$$

$$\mathcal{L}_{\rho_k}\left(u^{k+1}, v^{k+1}, \xi^{k+1}\right) - \mathcal{L}_{\rho_k}\left(u^{k+1}, v^{k+1}, \xi^k\right) = \frac{1}{\rho_k}\|\xi^{k+1} - \xi^k\|_F^2. \tag{21}$$

By the definition of v^{k+1}, we know that

$$\mathcal{L}_{\rho_k}\left(u^{k+1}, v^{k+1}, \xi^k\right) - \mathcal{L}_{\rho_k}\left(u^{k+1}, v^k, \xi^k\right) \leq 0. \tag{22}$$

Since $\ker(A) \bigcap \ker(W) = \{0\}$, there exist a series of nondecreasing constants $\{c_k\}_{k\geq 0}$, $c_k > 0$ for all k, such that

$$\mu\|Au\|_F^2 + \rho_k\|Wu\|_F^2 \geq c_k\|u\|_F^2.$$

Then $u \mapsto \mathcal{L}_{\rho_k}(u, v^k, \xi^k)$ is strongly convex, and the following inequality holds:

$$\mathcal{L}_{\rho_k}\left(u^{k+1}, v^k, \xi^k\right) - \mathcal{L}_{\rho_k}\left(u^k, v^k, \xi^k\right) \leq -\frac{c_k}{2}\|u^{k+1} - u^k\|_F^2. \tag{23}$$

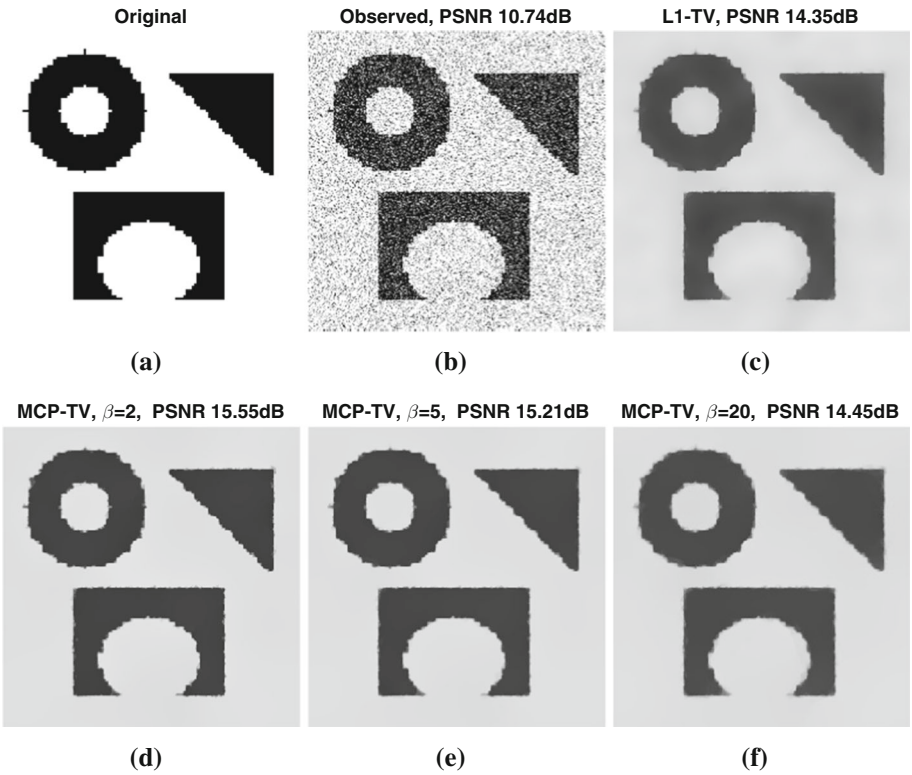

Fig. 3 From **a** to **f**: **a** original image of size 256×256; **b** the noisy image with $\theta = 0.175$; **c** the ADMM reconstruction of L_1-TV model with $\overline{\mu} = 2$; **d**–**f** the ADMM reconstruction of MCP model with $\mu = 2$ and $\beta = 2, 5, 20$ respectively

Summing (20), (21), (22), and (23) we have

$$\begin{aligned}\mathcal{L}^{k+1} - \mathcal{L}^k &\leq -\frac{c_k}{2}\|u^{k+1} - u^k\|_F^2 + \frac{\rho_{k+1} + \rho_k}{2\rho_k^2}\|\xi^{k+1} - \xi^k\|_F^2 \\ &\leq -\frac{c_k}{2}\|u^{k+1} - u^k\|_F^2 + \frac{2(\sigma+1)m_\alpha}{\rho_0\sigma^k}.\end{aligned}$$

Sum up the above inequality from $k = 0$, we get for any $k \geq 0$

$$\mathcal{L}^{k+1} - \mathcal{L}^0 \leq \frac{2m_\alpha(\sigma+1)\left(1 - \frac{1}{\sigma^{k+1}}\right)}{\rho_0\left(1 - \frac{1}{\sigma}\right)} - \sum_{0\leq i\leq k}\frac{c_i}{2}\|u^{i+1} - u^i\|_F^2 \tag{24}$$

Because $\sigma > 1$ and $u^0 = 0, v^0 = 0, \xi^0 = 0$, let $k \to \infty$, it concludes

$$\lim_{k\to\infty}\mathcal{L}^{k+1} < \frac{\mu}{2}\|f\|_F^2 + \frac{2m_\alpha\sigma(\sigma+1)}{\rho_0(\sigma-1)} < \infty \tag{25}$$

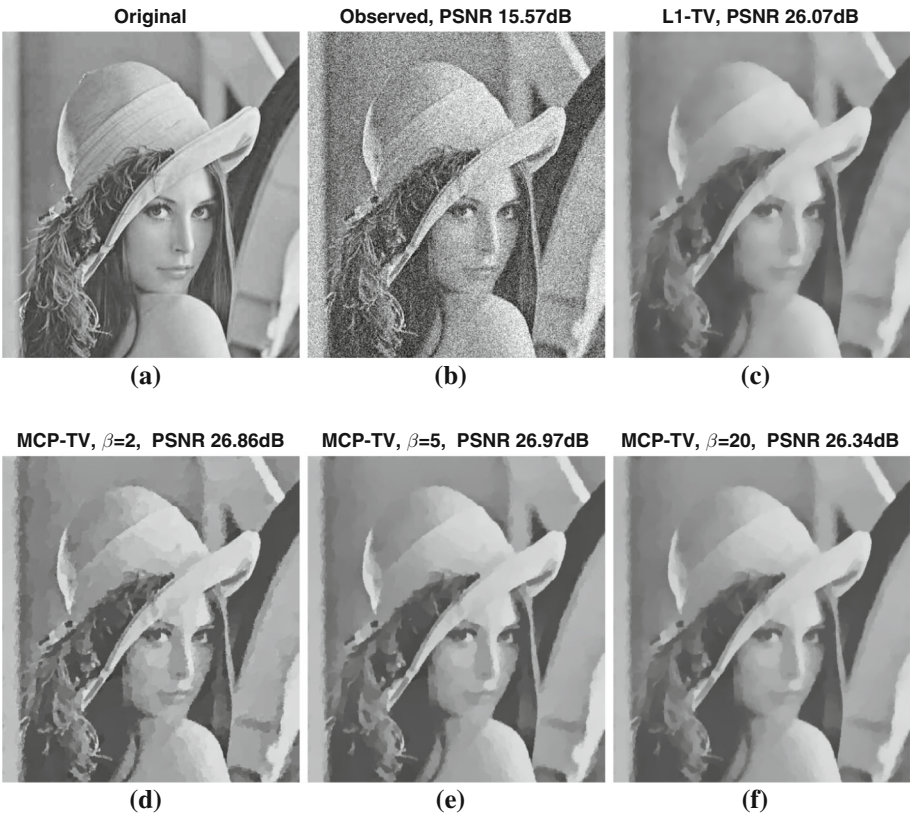

Fig. 4 From **a** to **f**: **a** original image of size 512×512; **b** the noisy image with $\theta = 0.03$; **c** the ADMM reconstruction of L_1-TV model with $\overline{\mu} = 4$; **d**–**f** the ADMM reconstruction of MCP model with $\mu = 4$ and $\beta = 2, 5, 20$ respectively

On the other side, by $\rho_k \to \infty$ and $\{\xi^k\}_{k=1}^{\infty}$ bounded, we know

$$\lim_{k\to\infty} \mathcal{L}^{k+1} \geq \lim_{k\to\infty} \frac{\|\xi^{k+1}\|_F^2}{2\rho_{k+1}} = 0. \tag{26}$$

So $\{\mathcal{L}^k\}_{k=1}^{\infty}$ is bounded. □

Proposition 3 *$\{v^k\}_{k=1}^{\infty}$ is bounded if β satisfies*

$$\frac{1}{\alpha^2}\left(\mu\|f\|_F^2 + \frac{4m_\alpha\sigma(\sigma+1)}{\rho_0(\sigma-1)}\right) < \beta. \tag{27}$$

Proof If the relationship holds, by (25) and $\lim_{k\to\infty} \frac{\|\xi^{k+1}\|_F^2}{2\rho_{k+1}} = 0$ we then know

$$\frac{1}{2}\alpha^2\beta > \lim_{k\to\infty} \mathcal{L}^{k+1} > \lim_{k\to\infty} \|v^{k+1}\|_{\mathrm{MCP}},$$

thus $\{v^k\}_{k=1}^{\infty}$ is bounded. Otherwise, if $\|v^k\|_F \to \infty$, then for some $i \in P$, $\|v_i^k\|_2 \to \infty$, so $p_{\alpha,\beta}(\|v_i^k\|_2) = \frac{1}{2}\alpha^2\beta$, and $\|v^k\|_{\mathrm{MCP}} \geq p_{\alpha,\beta}(\|v_i^k\|_2)$, contradiction. □

Table 1 Comparison of the models and different choice of β in the example, Iter means number of iterations

Example 1	Method	PSNR	Iter	CPU time (s)
Figure 3 (256×256)	ADMM-TV	14.35	14	0.307
	ADMM-MCP, $\beta = 2$	15.55	14	0.360
	ADMM-MCP, $\beta = 5$	15.54	15	0.434
	ADMM-MCP, $\beta = 20$	14.45	14	0.364
Figure 4 (512×512)	ADMM-TV	26.08	20	2.083
	ADMM-MCP, $\beta = 2$	26.86	18	2.356
	ADMM-MCP, $\beta = 5$	26.97	19	2.483
	ADMM-MCP, $\beta = 20$	26.34	20	2.627

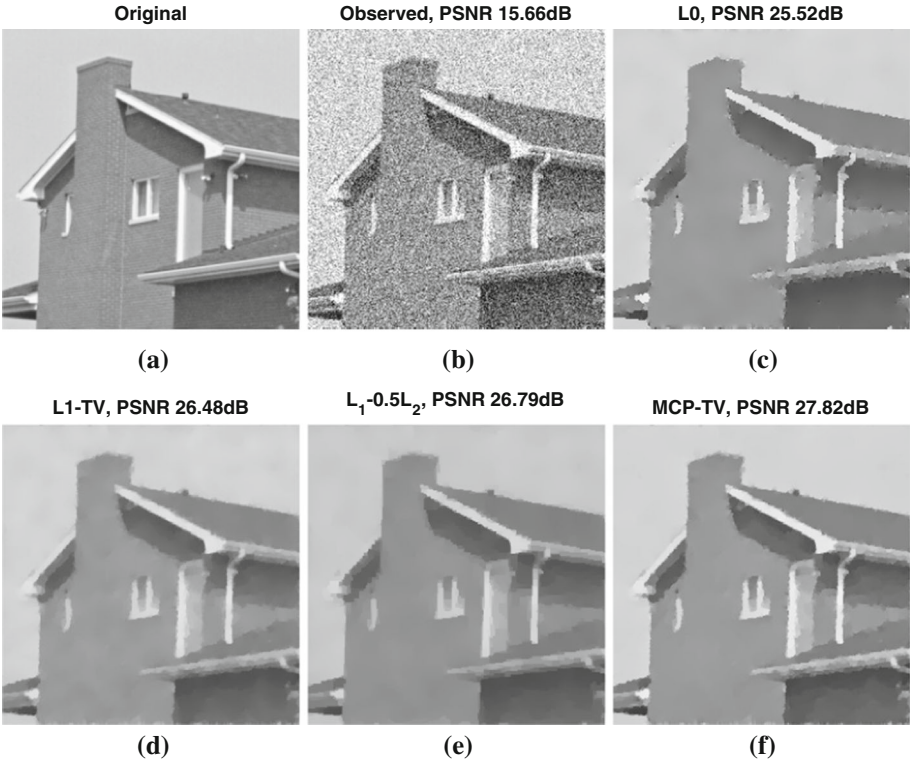

Fig. 5 From **a** to **f**: **a** original House image of size 256×256; **b** the noisy image with $\theta = 0.03$; **c** the ADMM reconstruction of L_0 model; **d** the ADMM reconstruction of TV model with $\overline{\mu} = 5$; **e** $L_1 - 0.5L_2$ [33] by DCA; **f** the ADMM reconstruction of MCP-TV model with $\mu = 5$ and $\alpha = 1$, $\beta = 5$

Theorem 2 *Suppose* $\ker(A) \bigcap \ker(W) = \{0\}$. *Let* $\{u^k, v^k, \xi^k\}_{k=1}^{\infty}$ *be a sequence generated by* (14), (15), (16). *If* β *satisfies* (27), *then* $\{u^k, v^k, \xi^k\}_{k=1}^{\infty}$ *is bounded and*

$$\lim_{k\to\infty} \|u^{k+1} - u^k\|_F + \|v^{k+1} - v^k\|_F = 0. \tag{28}$$

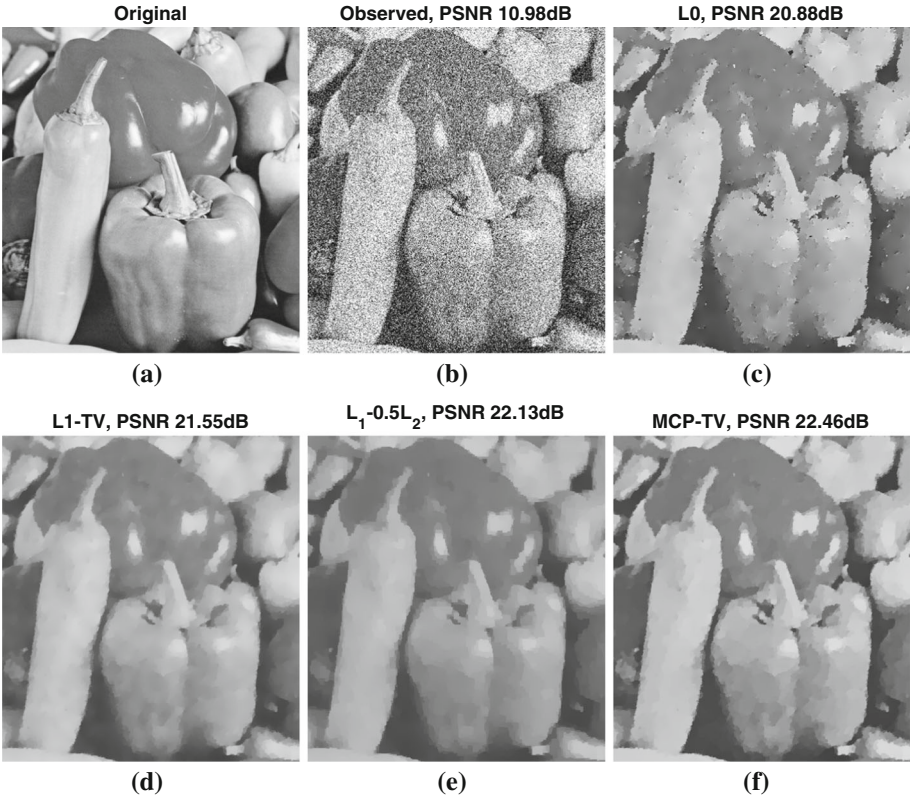

Fig. 6 From **a** to **f**: **a** original Peppers image of size 512×512; **b** the noisy image with $\theta = 0.12$; **c** the ADMM reconstruction of L_0 model; **d** the ADMM reconstruction of TV model with $\overline{\mu} = 3$; **e** $L_1 - 0.5L_2$ [33] by DCA; **f** the ADMM reconstruction of MCP-TV model with $\mu = 3$ and $\alpha = 1, \beta = 5$

Moreover, there exist a subsequence $\{u^{k_j}, v^{k_j}, \xi^{k_j}\}_{k_j=1}^{\infty}$, *which converges to a stationary point which satisfies* (19).

Proof By Proposition 3, $\{v^k\}_{k=1}^{\infty}$ is bounded. And by Lemma 2, $\{\mathcal{L}^k\}_{k=1}^{\infty}$ is bounded, thus $\{Au^k\}_{k=1}^{\infty}$ is bounded. Because $\{\xi^k\}_{k=1}^{\infty}$ bounded and $\rho_k \to \infty$, by (16) we know

$$\|Wu^{k+1} - v^{k+1}\|_F = \|\frac{\xi^{k+1} - \xi^k}{\rho_k}\|_F \to 0, \ k \to \infty \ . \tag{29}$$

So $\{Wu^k\}_{k=1}^{\infty}$ bounded as well. By the assumption that $\ker(A) \bigcap \ker(W) = \{0\}$, we can conclude that $\{u^k\}_{k=1}^{\infty}$ is bounded. So $\{u^k, v^k, \xi^k\}_{k=1}^{\infty}$ is bounded.
By (24) and (26) we have

$$0 \leq \lim_{k\to\infty} \sum_{0\leq i\leq k} \frac{c_i}{2}\|u^{i+1} - u^i\|_F^2 \leq \frac{\mu}{2}\|f\|_F^2 + \frac{2m_\alpha\sigma(\sigma+1)}{\rho_0(\sigma-1)},$$

which means $\lim_{k\to\infty} \sum_{0\leq i\leq k} \frac{c_i}{2}\|u^{i+1} - u^i\|_F^2 < \infty$ and thus

$$\lim_{k\to\infty} \|u^{k+1} - u^k\|_F = 0,$$

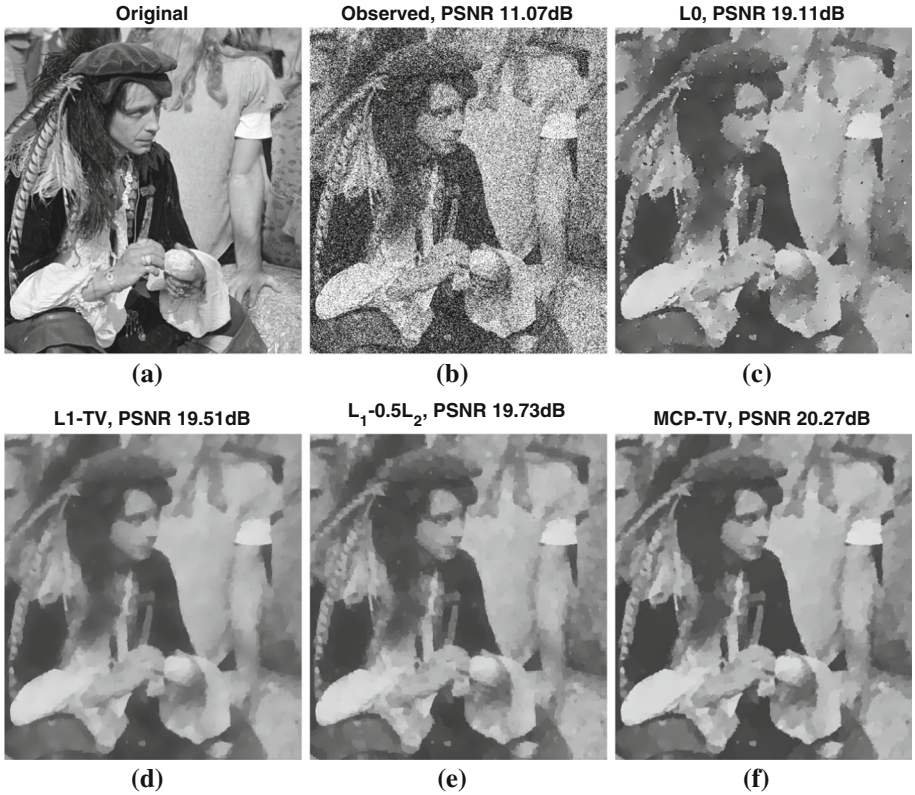

Fig. 7 From **a** to **f**: **a** original Pirate image of size 512×512; **b** the noisy image with $\theta = 0.12$; **c** the ADMM reconstruction of L_0 model; **d** the ADMM reconstruction of TV model with $\overline{\mu} = 3$; **e** $L_1 - 0.5L_2$ [33] by DCA; **f** the ADMM reconstruction of MCP-TV model with $\mu = 3$ and $\alpha = 1$, $\beta = 5$

By (29), we also have

$$\lim_{k\to\infty} \|v^{k+1} - v^k\|_F = 0,$$

Passage to a convergence subsequence $\{u^{k_j}, v^{k_j}, \xi^{k_j}\}_{k_j=1}^{\infty}$, which convergent to a point $\{u^*, v^*, \xi^*\}$. From the lower semicontinuity of $\mathcal{L}_\rho(u, v, \xi)$,

$$\liminf_{i\to\infty} \mathcal{L}_{\rho_{k_i}} \left(u^{k_i+1}, v^{k_i+1}, \xi^{k_i}\right) \geq \mathcal{L}_{\rho_\infty} \left(u^*, v^*, \xi^*\right),$$

and as v^{k_i+1} is a minimizer of function $\mathcal{L}_{\rho_{k_i}}(u^{k_i+1}, v, \xi^{k_i})$ with respect to v, so

$$\limsup_{i\to\infty} \mathcal{L}_{\rho_{k_i}} \left(u^{k_i+1}, v^{k_i+1}, \xi^{k_i}\right) \leq \mathcal{L}_{\rho_\infty} \left(u^*, v^*, \xi^*\right),$$

by the above two inequalities, we get

$$\lim_{i\to\infty} \|v^{k_i+1}\|_{\mathrm{MCP}} = \|v^*\|_{\mathrm{MCP}}.$$

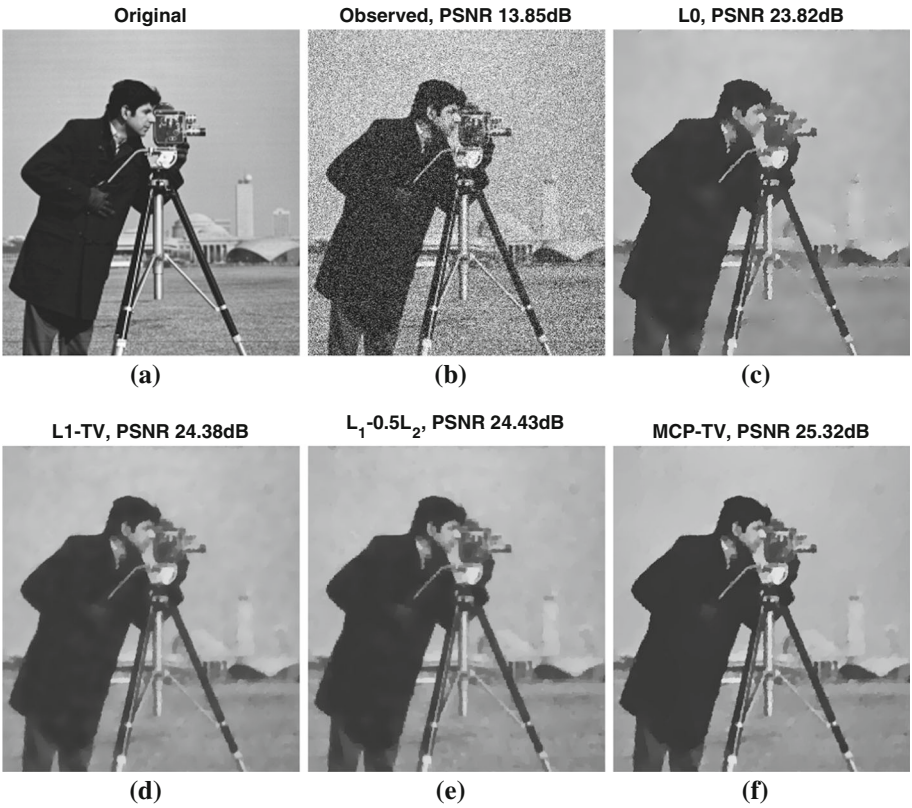

Fig. 8 From **a** to **f**: **a** original Cameraman image of size 512×512; **b** the noisy image with $\theta = 0.05$; **c** the ADMM reconstruction of L_0 model; **d** the ADMM reconstruction of TV model with $\overline{\mu} = 4$; **e** $L_1 - 0.5L_2$ [33] by DCA; **f** the ADMM reconstruction of MCP-TV model with $\alpha = 4$ and $\alpha = 1, \beta = 5$

One can immediately verify that

$$\begin{aligned} 0 &= \mu A^*(Au^* - f) + W^*\xi^*, \\ 0 &\in \partial\|v^*\|_{\mathrm{MCP}} - \xi^*, \\ 0 &= Wu^* - v^*, \end{aligned}$$

which means $\{u^*, v^*, \xi^*\}$ is a KKT point of problem 13. Thus we complete the proof. □

4 Numerical Examples

In this section we present several numerical examples to validate the proposed model (4) and to show the efficiency of the ADMM algorithm. The linear operator A is chosen as the identity operator (denosing) or the convolution operator (debluring), and the linear transformation W is fixed as the gradient operator. All codes are in MATLAB R2015a, and implemented on a laptop with an i7@2.60GHz Quad core and 8.0G memory. The peak signal-to-noise ratio (PSNR) was used to evaluate the performance of the image restoration. PSNR is defined by

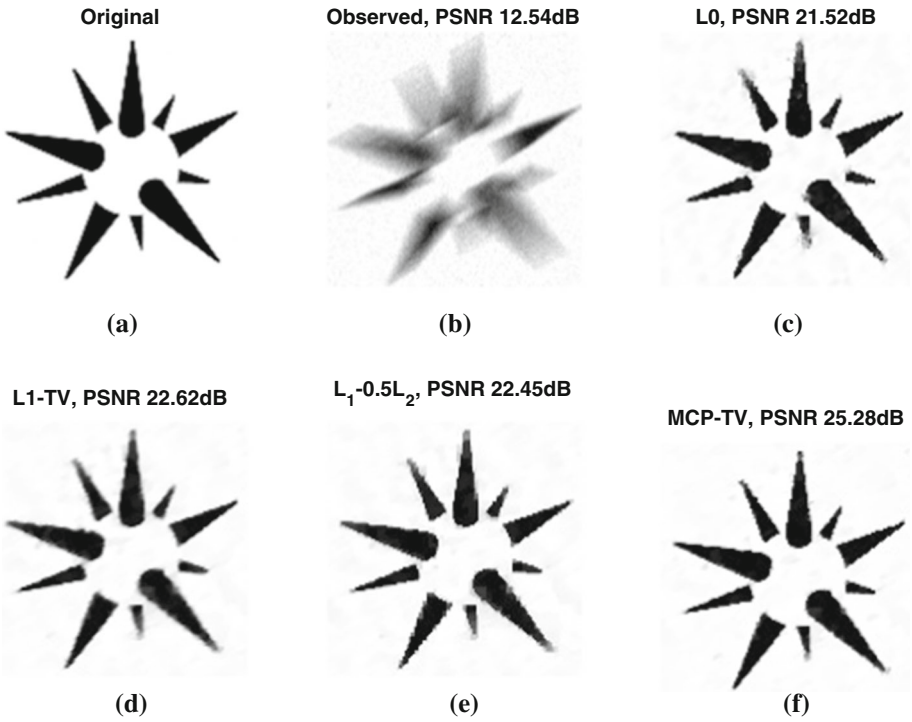

Fig. 9 The motion blur kernel corresponding to A is generated by the Matlab function `fspecial('motion',20,30)`. And add Gaussian noise with $\theta = 0.001$. From **a** to **f**: **a** original image; **b** the noisy image; **c** the ADMM reconstruction of L_0 model; **d** the ADMM reconstruction of TV model with $\overline{\mu} = 400$; **e** $L_1 - 0.5L_2$ [33] by DCA; **f** the ADMM reconstruction of MCP model with $\mu = 400$ and $\alpha = 1, \beta = 5$

$$\text{PSNR} = 20 * \log_{10} \frac{|u_{\max}|}{\text{RMS}} [\text{dB}],$$

where $u_{\max}$ is the largest possible value of the original image u (which is 1 in our experiment settings), and RMS represents the root mean squared difference between the two images. The stopping criterion for the iterations of Algorithm 1 is

$$\frac{\|u^{k+1} - u^k\|_2}{\|u^k\|_2} < \delta, \tag{30}$$

for some given small constant δ.

To compare with the other variational model, we consider the standard isotropic TV model [43] in Examples 1 and 2:

$$\underset{u}{\text{minimize}} \ \frac{\overline{\mu}}{2} \|Au - f\|_2^2 + \|\nabla u\|_1, \tag{31}$$

where $\|\nabla u\|_1 := \|u\|_{\text{TV}}$ as defined in (6). It is well known that the choice of regularization parameter $\overline{\mu}$ is closely related to the noise level. In numerical test, we use the Split Bregman methods to solve the TV model (ADMM-TV).

Example 1 In this example, we illustrate the choice of parameters α and β. From the definition of MCP function, it can be found that the parameters μ and α in (2) together play the role

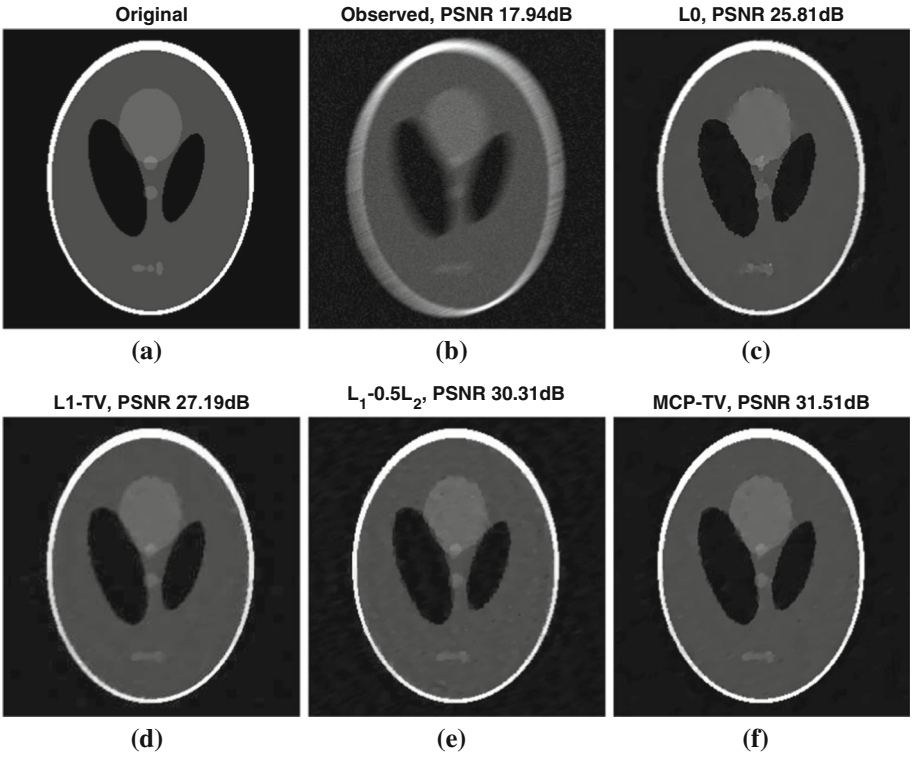

Fig. 10 The motion blur kernel corresponding to A is generated by the Matlab function `fspecial('motion',15,25)`. And add Gaussian noise with $\theta = 0.0005$. From **a** to **f**: **a** original image; **b** the noisy image; **c** the ADMM reconstruction of L_0 model; **d** the ADMM reconstruction of TV model with $\overline{\mu} = 250$; **e** $L_1 - 0.5L_2$ [33] by DCA; **f** the ADMM reconstruction of MCP model with $\mu = 250$ and $\alpha = 1, \beta = 5$

of $\overline{\mu}$ in the TV model (31). To simplify the interpretation, we fix α to be 1 in our numerical experiments, then the choice of μ is similar as $\overline{\mu}$ in (31), i.e., related to the noise level. We then discuss the choice of another parameter β. From the definition of $p_{\alpha,\beta}$, it approximates ℓ_0 when $\beta \to 1$ and approximates ℓ_1 when $\beta \to \infty$. We can see from the experiments Figs. 3 and 4 that the edges of the restored images are sharp when β small, and approaches L_1 model when β is large, but the restored images are relatively smooth. We choose $\beta = 5$ in all the experiments as a balance between the sharp edge and smooth image. The corresponding numerical tests are in Figs. 3 and 4, where the PSNR value, the number of iterations and the CPU time (in second) can be found in Table 1.

Example 2 In Example 2, we compare the proposed model (2) with TV and two nonconvex models, L_0 model and a weighted difference of anisotropic and isotropic model [33] (i.e., $L_1 - \gamma L_2$ with a weighted parameter γ). The last model is solved by DCA [33,47]. According to [33], the parameter γ can be chosen as 0.5.

We divide the second example into two parts: first part for denoising and second part for debluring. The pictures are shown in Figs. 5, 6, 7 and 8 (for denoising) and Figs. 9, 10 and 11 (for debluring), and the PSNR values, numbers of iterations and CPU times are given in Table 2.

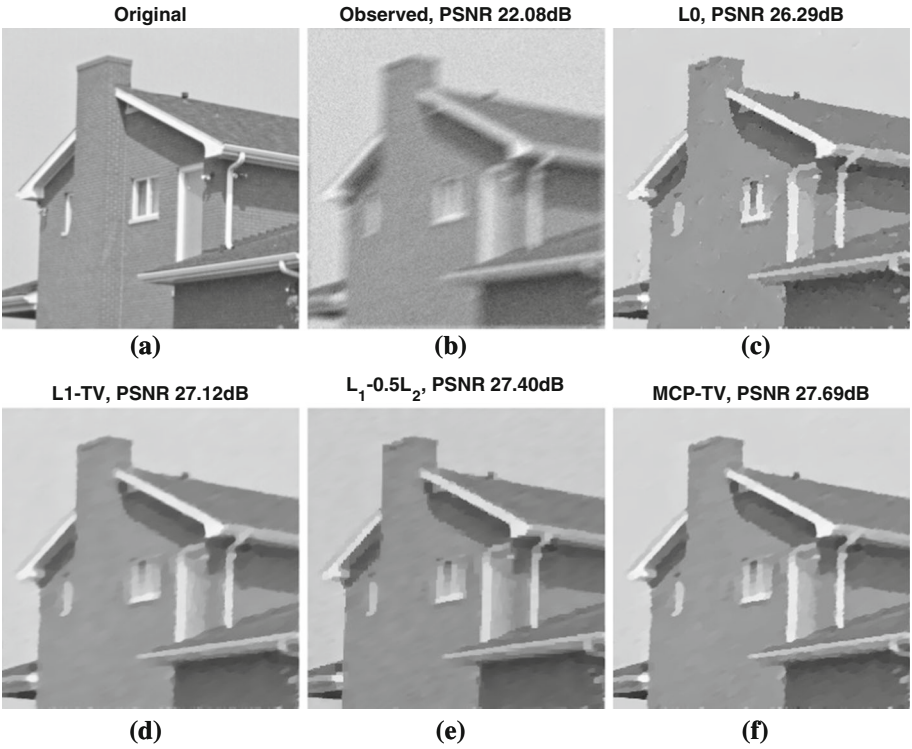

Fig. 11 The motion blur kernel corresponding to A is generated by the Matlab function `fspecial('motion',15,25)`. And add Gaussian noise with $\theta = 0.001$. From **a** to **f**: **a** original image; **b** the noisy image; **c** the ADMM reconstruction of L_0 model; **d** the ADMM reconstruction of TV model with $\overline{\mu} = 100$; **e** $L_1 - 0.5L_2$ [33] by DCA; **f** the ADMM reconstruction of MCP model with $\mu = 100$ and $\alpha = 1, \beta = 5$

Example 2.1 In this part we consider the problem of image denoising, where the noise is Gaussian with variance θ. The parameters are fixed as $\alpha = 1, \beta = 5, \rho_0 = 5, \sigma = 1.25$, and the stopping criteria parameter $\delta = 1e^{-3}$. For the images as House, Peppers, Pirate and Cameraman (see Figs. 5, 6, 7, 8), the noise level θ, parameters μ in MCP and $\overline{\mu}$ in TV model are set to be $(\theta, \overline{\mu}, \mu) = (0.03, 5, 5), (0.12, 3, 3), (0.12, 3, 3)$ and $(0.05, 4, 4)$, respectively. From Figs. 5, 6, 7, 8 and Table 2, we find the proposed MCP model outperform the other models.

Example 2.2 In this part we consider image deblurring problems where a binary image is blurred by motion blurs plus Gaussian noise with variance $\theta = 0.001, 0.0005, 0.001$, respectively. We fix $\rho_0 = 1, \sigma = 1.2$, and the stopping criteria parameter $\delta = 5e^{-4}$. From Figs. 9, 10 and 11, the parameters $(\overline{\mu}, \mu, \alpha, \beta)$ are choose as $(400, 400, 1, 5), (250, 250, 1, 5), (100, 100, 1, 5)$, respectively. From the above pictures we see that the proposed MCP model exhibit advantages over the TV model and other nonconvex models in term of PSNR value.

Example 3 In this example, we will show the efficiency of ADMM algorithm for the proposed model (2). Three different methods: ADMM algorithm 3.2, Local Linear Approximation (LLA, see [53]) and Difference of Convex Functions Algorithm (DCA, c.f. [46,47]), are

Table 2 Comparison of the models and methods in the Examples 2.1 and 2.2, Iter means number of iterations

Examples 2.1 and 2.2	Method	PSNR	Iter	CPU time (s)
Figure 5	DCA-($L_1 - 0.5L_2$) [33]	26.79	10	5.498
House (256×256)	ADMM-L_0	25.52	20	0.419
	ADMM-TV	26.48	20	0.416
	ADMM-MCP	27.82	15	0.400
Figure 6	DCA-($L_1 - 0.5L_2$) [33]	22.13	10	50.793
Peppers (512×512)	ADMM-L_0	20.88	20	2.031
	ADMM-TV	21.55	18	1.879
	ADMM-MCP	22.46	18	2.391
Figure 7	DCA-($L_1 - 0.5L_2$) [33]	19.73	10	62.625
Pirate (512×512)	ADMM-L_0	19.11	20	2.023
	ADMM-TV	19.51	20	2.073
	ADMM-MCP	20.27	20	2.517
Figure 8	DCA-($L_1 - 0.5L_2$) [33]	24.43	10	45.643
Cameraman (512×512)	ADMM-L_0	23.82	20	1.931
	ADMM-TV	24.38	18	1.905
	ADMM-MCP	25.32	16	2.157
Figure 9	DCA-($L_1 - 0.5L_2$) [33]	22.45	20	4.770
(128×128)	ADMM-L_0	21.52	20	0.094
	ADMM-TV	22.62	26	0.113
	ADMM-MCP	25.28	38	0.267
Figure 10	DCA-($L_1 - 0.5L_2$) [33]	30.31	20	52.766
Phantom (256×256)	ADMM-L_0	25.81	40	1.024
	ADMM-TV	27.19	48	1.220
	ADMM-MCP	32.01	32	1.044
Figure 11	DCA-($L_1 - 0.5L_2$) [33]	14.28	20	24.428
House (256×256)	ADMM-L_0	26.29	40	0.900
	ADMM-TV	27.12	31	0.642
	ADMM-MCP	27.69	30	0.803

given to solve the proposed model (2). For the LLA method, one needs to solve a sequence of linear approximation to the model (2):

$$u^{k+1} \leftarrow \operatorname{argmin} \frac{\mu}{2}\|Au - f\|_F^2 + \langle \partial \|Wu^k\|_{\mathrm{MCP}}, |u| - |u^k| \rangle \,, \tag{32}$$

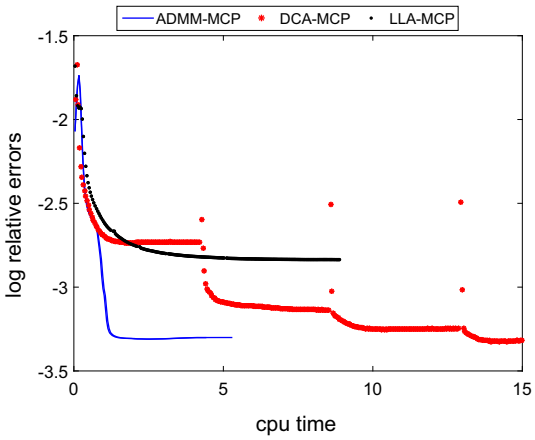

Fig. 12 Log of *relative error* versus cpu time for the LLA, DCA, ADMM solving the proposed model (2) for the problem as in Fig. 10

and the subproblem (32) is solved by ADMM. For DCA method, one can rewrite the objective functional as $G - H$, where $G = \frac{\mu}{2}\|Au - f\|_F^2 + \|Wu\|_{\mathrm{MCP}} + \frac{c}{2}\|u\|_F^2$ and $H = \frac{c}{2}\|u\|_F^2$ with a constant c. Then ADMM is applied to solve a sequence of subproblem 33:

$$u^{k+1} \leftarrow \operatorname{argmin} \frac{\mu}{2}\|Au - f\|_F^2 + \|Wu\|_{\mathrm{MCP}} + \frac{c}{2}\|u\|_F^2 - \langle u, y^k\rangle, \tag{33}$$

and $y^{k+1} = cu^{k+1}$.

Figure 12 shows the log of *relative error* versus cpu time to different methods for the problem as in Fig. 10. The *relative error* is defined to be $\frac{\|u^k - u^\star\|_2}{\|u^\star\|_2}$, where $u^\star$ is the true image. From Fig. 12, one can find that ADMM and DCA get more accurate results than LLA, but ADMM spends less CPU time than DCA.

5 Conclusion

In this paper a MCP-based nonconvex variational model for image restoration is introduced. The existence of the minimizer has been proved. An ADMM algorithm is proposed to solve the model, and convergence analysis is provided. Numerical experiments show the efficiency of the model and the algorithm.

Acknowledgements We thank the reviewers and editor for providing very useful comments and suggestions. We also thank Dr. Yifei Lou for providing the code of $L_1 - 0.5L_2$. The research of Y. Jiao is partially supported by National Science Foundation of China No. 11501579, X. Lu is supported by National Science Foundation of China Nos. 11471253 and 91630313, and T. Zeng is supported in part by National Science Foundation of China No. 11671002, CUHK start-up and CUHK DAG 4053296.

References

1. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka–Łojasiewicz inequality. Math. Oper. Res. **35**(2), 438–457 (2010)
2. Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Math. Program. **146**(1–2), 459–494 (2014)

3. Boyd, S.: Alternating direction method of multipliers. Talk at NIPS Workshop on Optimization and Machine Learning (2011)
4. Cai, J., Chan, R.H., Shen, L., Shen, Z.: Convergence analysis of tight framelet approach for missing data recovery. Adv. Comput. Math. **31**(1), 87–113 (2009)
5. Cai, J.-F., Osher, S., Shen, Z.: Split Bregman methods and frame based image restoration. Multiscale Model. Simul. **8**(2), 337–369 (2009)
6. Cai, X., Chan, R., Zeng, T.: A two-stage image segmentation method using a convex variant of the Mumford–Shah model and thresholding. SIAM J. Imaging Sci. **6**(1), 368–390 (2013)
7. Candes, E.J., Romberg, J.K., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. Commun. Pure Appl. Math. **59**(8), 1207–1223 (2006)
8. Chambolle, A.: An algorithm for total variation minimization and applications. J. Math. Imaging Vis. **20**(1), 89–97 (2004)
9. Chambolle, A., Caselles, V., Cremers, D., Novaga, M., Pock, T.: An introduction to total variation for image analysis. Theor. Found. Numer. Methods Sparse Recovery **9**(263–340), 227 (2010)
10. Chan, R.H., Riemenschneider, S.D., Shen, L., Shen, Z.: Tight frame: an efficient way for high-resolution image reconstruction. Appl. Comput. Harmon. Anal. **17**(1), 91–115 (2004)
11. Chan, R.H., Ng, M.K.: Conjugate gradient methods for Toeplitz systems. SIAM Rev. **38**(3), 427–482 (1996)
12. Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. SIAM Rev. **43**(1), 129–159 (2001)
13. Chouzenoux, E., Jezierska, A., Pesquet, J.-C., Talbot, H.: A majorize–minimize subspace approach for ℓ_2–ℓ_0 image regularization. SIAM J. Imaging Sci. **6**(1), 563–591 (2013)
14. Daubechies, I., Han, B., Ron, A., Shen, Z.: Framelets: MRA-based constructions of wavelet frames. Appl. Comput. Harmon. Anal. **14**(1), 1–46 (2003)
15. Deng, W., Yin, W.: On the global and linear convergence of the generalized alternating direction method of multipliers. J. Sci. Comput. **66**(3), 889–916 (2016)
16. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006)
17. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. **55**(1), 293–318 (1992)
18. Esser, E.: Applications of Lagrangian-based alternating direction methods and connections to split Bregman. CAM Rep. **9**, 31 (2009)
19. Fan, J., Li, R.: Variable selection via nonconcave penalized likelihood and its oracle properties. J. Am. Stat. Assoc. **96**(456), 1348–1360 (2001). [MR 1946581 (2003k:62160)]
20. Fan, J., Li, R.: Variable selection via nonconcave penalized likelihood and its oracle properties. J. Am. Stat. Assoc. **96**(456), 1348–1360 (2001)
21. Fan, J., Peng, H.: Nonconcave penalized likelihood with a diverging number of parameters. Ann. Stat. **32**(3), 928–961 (2004). [MR 2065194 (2005g:62047)]
22. Frank, L.L.E., Friedman, J.H.: A statistical view of some chemometrics regression tools. Technometrics **35**(2), 109–135 (1993)
23. Fu, S.J., Zhang, C.M., Tai, X.C.: Image denoising and deblurring: non-convex regularization, inverse diffusion and shock filter. Sci. China Inf. Sci. **54**(6), 1184–1198 (2011)
24. Fu, W.J.: Penalized regressions: the bridge versus the lasso. J. Comput. Graph. Stat. **7**(3), 397–416 (1998)
25. Gabay, D.: Chapter ix applications of the method of multipliers to variational inequalities. Stud. Math. Appl. **15**, 299–331 (1983)
26. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. **2**(1), 17–40 (1976)
27. Getreuer, P.: Rudin–Osher–Fatemi total variation denoising using split Bregman. Image Process. On Line **2**, 74–95 (2012)
28. Glowinski, R., Marroco, A.: Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires. Revue française d'automatique, informatique, recherche opérationnelle. Anal. Numér. **9**(2), 41–76 (1975)
29. Goldstein, T., Osher, S.: The split Bregman method for L1-regularized problems. SIAM J. Imaging Sci. **2**(2), 323–343 (2009)
30. Hong, M., Luo, Z.-Q., Razaviyayn, M.: Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. SIAM J. Optim. **26**(1), 337–364 (2016)
31. Jiao, Y., Jin, B., Lu, X., Ren, W.: A primal dual active set algorithm for a class of nonconvex sparsity optimization. arXiv preprint arXiv:1310.1147 (2013)
32. Lions, P.-L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**(6), 964–979 (1979)

33. Lou, Y., Zeng, T., Osher, S., Xin, J.: A weighted difference of anisotropic and isotropic total variation model for image processing. SIAM J. Imaging Sci. **8**(3), 1798–1823 (2015)
34. Mallat, S.G.: Multifrequency channel decompositions of images and wavelet models. IEEE Trans. Acoust. Speech Signal Process. **37**(12), 2091–2110 (1989)
35. Mordukhovich, B.S., Shao, Y.: On nonconvex subdifferential calculus in banach spaces1. J. Convex Anal. **2**(1/2), 211–227 (1995)
36. Moreau, J.-J.: Fonctions convexes duales et points proximaux dans un espace hilbertien. CR Acad. Sci. Paris Ser. A Math. **255**, 2897–2899 (1962)
37. Moreau, J.-J.: Proximité et dualité dans un espace hilbertien. Bull. Soc. Math. Fr. **93**(2), 273–299 (1965)
38. Nikolova, M.: Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares. Multiscale Model. Simul. **4**(3), 960–991 (2005)
39. Nikolova, M.: Analytical bounds on the minimizers of (nonconvex) regularized least-squares. Inv. Probl. Imaging **1**(4), 1–677 (2007)
40. 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**(1), 2–25 (2008)
41. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math. Oper. Res. **1**(2), 97–116 (1976)
42. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis, vol. 317. Springer, Berlin (2009)
43. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D Nonlinear Phenom. **60**(1–4), 259–268 (1992)
44. Setzer, S., Steidl, G., Teuber, T.: Deblurring Poissonian images by split Bregman techniques. J. Vis. Commun. Image Represent. **21**(3), 193–199 (2010)
45. Sun, D., Sun, J., Zhang, L.: The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming. Math. Program. **114**(2), 349–391 (2008)
46. Tao, P.D., An, L.T.H.: Convex analysis approach to DC programming: theory, algorithms and applications. Acta Math. Vietnam. **22**(1), 289–355 (1997)
47. Tao, P.D., An, L.T.H.: A DC optimization algorithm for solving the trust-region subproblem. SIAM J. Optim. **8**(2), 476–505 (1998)
48. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B **58**, 267–288 (1996)
49. Yang, L., Pong, T., Chen, X.: Alternating direction method of multipliers for a class of nonconvex and nonsmooth problems with applications to background/foreground extraction. SIAM J. Imaging Sci. **10**(1), 74–110 (2017)
50. Zhang, C.-H.: Nearly unbiased variable selection under minimax concave penalty. Ann. Stat. **38**(2), 894–942 (2010)
51. Zhang, C.-H., Huang, J.: The sparsity and bias of the LASSO selection in high-dimensional linear regression. Ann. Stat. **36**(4), 1567–1594 (2008)
52. Zhang, T.: Analysis of multi-stage convex relaxation for sparse regularization. J. Mach. Learn. Res. **11**(2), 1081–1107 (2010)
53. Zou, H., Li, R.: One-step sparse estimates in nonconcave penalized likelihood models. Ann. Stat. **36**(4), 1509 (2008)