Electronic Research Archive

https://www.aimspress.com/journal/era

ERA, 32(5): 3433–3456.
DOI: 10.3934/era.2024159
Received: 20 November 2023
Revised: 24 February 2024
Accepted: 09 May 2024
Published: 28 May 2024

Research article

An MTL1TV non-convex regularization model for MR Image reconstruction using the alternating direction method of multipliers

Xuexiao You[1,2], **Ning Cao**[1,*] **and Wei Wang**[3]

[1] College of Information Science and Engineering, Hohai University, Nanjing 210098, China

[2] School of Mathematics and Statistics, Hubei Normal University, Huangshi 435002, China

[3] Key Laboratory of Clinical Engineering, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing 211166, China

* **Correspondence:** Email: caoning@vip.163.com.

Abstract: The acquisition time of magnetic resonance imaging (MRI) is relatively long. To achieve high-quality and fast reconstruction of magnetic resonance (MR) images, we proposed a non-convex regularization model for MR image reconstruction with the modified transformed l_1 total variation (MTL1TV) regularization term. We addressed this new model using the alternating direction method of multipliers (ADMM). To evaluate the proposed MTL1TV model, we performed numerical experiments on several MR images. The numerical results showed that the proposed model gives reconstructed images of improved quality compared with those obtained from state of the art models. The results indicated that the proposed model can effectively reconstruct MR images.

Keywords: MR Image reconstruction; transformed l_1 penalty; non-convex regularization; total variation; ADMM

1. Introduction

Magnetic resonance imaging (MRI) is an approach broadly used for clinical diagnosis because of its ability to effectively reflect changes in human organs [1–3]. However, a long-standing challenge is that the acquisition time of MRI is relatively long. To address this problem, researchers have proposed numerous techniques, such as sparse sampling [4, 5] and parallel imaging [6, 7]. After the theory of compressed sensing (CS) [8] was proposed, many scholars applied it to recover magnetic resonance (MR) images from highly undersampled measurements [9, 10]. Research on MR image compression sensing has thus become increasingly important.

In the MRI domain, the data acquisition of MRI can be modeled as follows:

$$\mathrm{y} = \mathrm{Ax} + \varepsilon \tag{1.1}$$

where $\mathrm{x} \in \mathbb{R}^n$ denotes the desired MRI data, $\mathrm{y} \in \mathbb{R}^m$ is the observed undersampled k-space MRI data, and the matrix $\mathrm{A} = \mathrm{RF} \in \mathbb{R}^{m\times n}(m < n)$. R represents the undersampling operator, such as the radial sampling operator and Cartesian sampling operator. F is the discrete Fourier transform (DFT) operator, and $\varepsilon \in \mathbb{R}^m$ is the additive Gaussian noise. The objective of MR image reconstruction is to recover x from y. Consequently, the reconstruction problem of (1) can be formulated as

$$\min_{\mathrm{x}} \lambda\varphi(\mathrm{x}) + \frac{1}{2}\left\|\mathrm{y} - \mathrm{Ax}\right\|_2^2 \tag{1.2}$$

in which $\left\|\mathrm{y} - \mathrm{Ax}\right\|_2^2$ is the data fidelity term, $\lambda > 0$ is the regularization parameter, and $\varphi(x)$ is some regularizing functional exploiting image prior knowledge. In CS-MRI, $\varphi(\mathrm{x})$ is the standard total variation (TV) norm (or l_1-based regularization) [11], and the corresponding reconstruction model can be expressed as

$$\min_{\mathrm{x}} \lambda\|\mathrm{x}\|_{TV} + \frac{1}{2}\left\|\mathrm{y} - \mathrm{Ax}\right\|_2^2 \tag{1.3}$$

where $\|\mathrm{x}\|_{TV} = \|D\mathrm{x}\|_1$ (D is the gradient operator). Although standard TV regularization is convex and widely used, the l_1-norm regularized model induces bias for large coefficients [12] and, hence, lacks the oracle property. To overcome the shortcomings of l_1-norm regularization, researchers have proposed some non-convex regularization methods based on the capped-L1 (CaP) [13], smoothly clipped absolute deviation (SCAD) [14], minimax concave (MC) penalty [15–18], arctangent penalty (ATAN) [19], logarithm function (Log) [20], transformed l_1 (TL1) [21, 22], fraction function penalty [23], and non-convex graph total variation (GTV) regularization [23]. For instance, Liu et al. [24] introduced the minimax concave total variation (MCTV) penalty as a regularization term for MR brain image reconstruction. Further, Luo et al. [25] used the arctangent function as the non-convex TV regularize term (AtanTV) for MR image reconstruction. The same team [26] introduced the non-convex MR image reconstruction model via the SCAD penalty function. Lu et al. [27] also used the non-convex Cauchy total variation (CauchyTV) as the regularization term for MR image reconstruction. These studies have shown that non-convex penalties in image restoration usually have better performance than the l_1-norm (or standard TV).

The contributions of this study are as follows. First, we slightly modify the transformed l_1 (TL1) function and propose a modified TL1 penalty function (MTL1). Second, inspired by the previous work, we use the MTL1 as the regularization term (MTL1TV) to construct a non-convex regularization model for MR image reconstruction. Third, the proposed model can be solved by the alternating direction method of multipliers (ADMM) [28–30]. Finally, numerical experiments on several MR images showed that compared with the traditional TV and state of the art models, the performance of the MTL1TV model was significantly improved. These results confirm the effectiveness of the proposed model.

The remainder of this paper is organized as follows. In Section 2, we formally introduce the MTL1 function and present a non-convex MTL1TV model, which is based on the MTL1 function. In Section 3, we utilize the ADMM algorithm to address the proposed non-convex model. The convergence result is provided in Section 4. The experimental results are then presented in Section 5. Finally, we conclude the paper in Section 6.

2. Preliminaries and MTL1TV MRI model

In this section, we provide the definition of the the modified TL1 penalty function (MTL1) and give an explicit expression of the proximity operator of MTL1. Then, we use MTL1 to define MTL1TV regularization and structure a non-convex MTL1TV model for MR image reconstruction.

2.1. Preliminaries

Definition 1. *The modified transformed l_1 function (MTL1) is defined as*

$$\phi_a(x) = \frac{a\,|x|}{a + |x|}, \tag{2.1}$$

where parameter $a > 0$. It is obtained by slightly modifying the following TL1 penalty function [20]:

$$p_a(x) = \frac{(a+1)\,|x|}{a + |x|}. \tag{2.2}$$

Figure 1 shows that the MTL1 is a good alternative to the l_1 norm. The larger the value of parameter a is, the closer the behavior of MTL1 is to the l_1 norm.

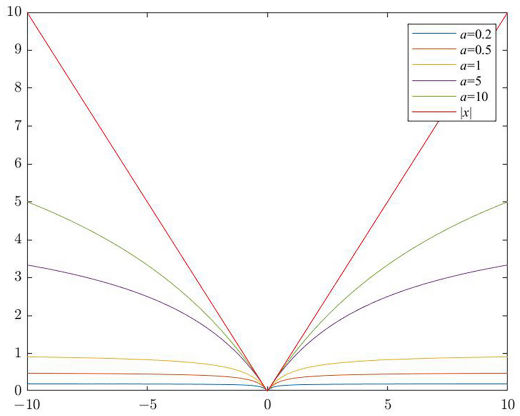


Figure 1. $|x|$ and MTL1 function plots with $a = \{0.2, 0.5, 1, 5, 10\}$.

The proximal operator of $\phi_a(x)$ at $t \in \mathbb{R}$ is defined as

$$prox_{\lambda}^{\phi}(t) = \arg\min_{x\in\mathbb{R}} \left\{ \theta_{\lambda,t}(x) := \lambda\phi_a(x) + \frac{1}{2}(x-t)^2 \right\}. \tag{2.3}$$

Similar to Theorem III.1 of [20], we can derive that for the MTL1 function, there exists a closed-formed representation of the optimal solution to (2.3).

Theorem 1. *The optimal solution $x^* = \arg\min \theta_{\lambda,t}(x)$ is a threshold function defined as*

$$x^* = \begin{cases} 0, & |t| \le \delta_\lambda; \\ g_\lambda(t), & |t| > \delta_\lambda. \end{cases} \tag{2.4}$$

where

$$g_\lambda(t) = \mathrm{sgn}(t)\left(\frac{2}{3}(a+|t|)\cos\left(\frac{\varphi(t)}{3}\right) - \frac{2a}{3} + \frac{|t|}{3}\right), \tag{2.5}$$

with $\varphi(t) = \arccos(1 - \frac{27\lambda a^2}{2(a+|t|)^3})$, *and the threshold value* δ_λ *satisfies*

$$\delta_\lambda = \begin{cases} \lambda & \lambda \le \frac{a}{2}; \\ \sqrt{2\lambda a} - \frac{a}{2} & \lambda > \frac{a}{2}. \end{cases}$$

To prove Theorem 1, we first give the following lemmas.

Lemma 1. *Define three parameters*

$$t_1^* = \sqrt[3]{\frac{27\lambda a^2}{4}} - a, t_2^* = \lambda, t_3^* = \sqrt{2\lambda a} - \frac{a}{2}$$

for any parameters $\lambda > 0$ *and* $a > 0$, *then the inequalities* $t_1^* \le t_3^* \le t_2^*$ *hold. Furthermore, they are equal to* $\frac{a}{2}$ *when* $\lambda = \frac{a}{2}$.

Lemma 2. *For any given t, the roots of the following two cubic polynomials of x satisfy the following properties:*

1) If $t > t_1^*$, *then the cubic polynomial*

$$x(a+x)^2 - t(a+x)^2 + \lambda a^2 \tag{2.6}$$

has three distinct real roots, and the largest root x_0 *is given by* $x_0 = g_\lambda(t)$, $|g_\lambda(t)| \le |t|$.

2) If $t < -t_1^*$, *then the cubic polynomial*

$$x(a-x)^2 - t(a-x)^2 - \lambda a^2 \tag{2.7}$$

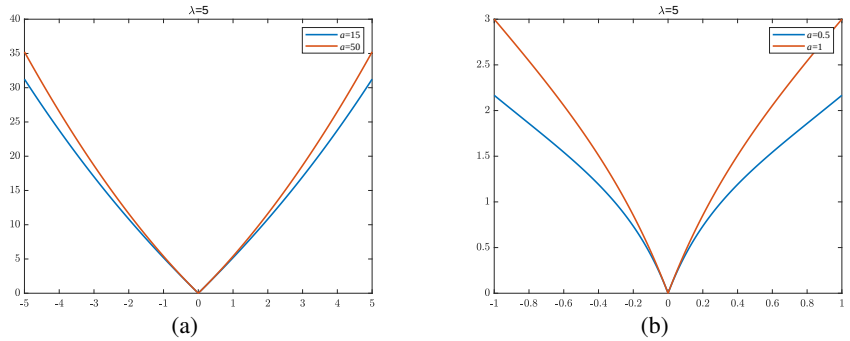
has also three distinct real roots, and the smallest root x_0 *is given by* $x_0 = g_\lambda(t)$.

The proof of Lemma 2 is essentially the same as Lemma III.1 in [20] and Lemma 8 in [22], with only minor changes required due to the different values of t_1^*, t_2^*, and t_3^* and different cubic polynomials. The proof of Theorem 1 is similar to Theorem III.1 in [20] and Lemma 9 in [22]. The detailed proofs of Lemma 2 and Theorem 1 can be found in Appendixes A and B, respectively. Based on the Theorem 1, the following corollary holds.

Corollary 1. *For any* $a \ge 2\lambda$, *the function* $\theta_{\lambda,t}(x)$ *defined in (2.6) is strictly convex.*

The proof of Corollary 1 can be found in Appendix C.

The graph presented in Figure 2(a) shows the plots of the function $\theta_{\lambda,t}(x) = \lambda\phi_a(x) + \frac{1}{2}(x-t)^2$ with $t = 0$, $\lambda = 5$, and $a = 15, 50$. It can be seen in Figure 2(a) that the function $\theta_{\lambda,t}(x)$ is strictly convex for $a \ge 2\lambda$. However, when $\lambda = 5$, and $a = 0.5, 1$, the graph presented in Figure 2(b) shows that the function $\theta_{\lambda,t}(x)$ is non-convex.



(a) (b)

Figure 2. The behavior of the function $\theta_{\lambda,t}(x)$ with a, $\lambda = 5$, and $t = 0$.

From Theorem 1, the convexity of the function $\theta_{\lambda,t}(x)$ can be ensured by constraining the parameter a as $a \geq 2\lambda$, which means that there exists the unique global minimizer to $\theta_{\lambda,t}(x)$ when $a \geq 2\lambda$. The unique global minimizer to $\theta_{\lambda,t}(x)$ can be expressed in the following theorem.

According to (2.1), we consider another function $\psi_a(x)$, which is induced by $\phi_a(x)$:

$$\psi_a(x) := |x| - \phi_a(x). \tag{2.8}$$

Figure 3 shows the curves of the functions $|x|$, $\phi_a(x)$, $\psi_a(x)$ and $\frac{x^2}{a}(a = 10)$. Clearly, the function $\psi_a(x)$ is a convex, continuously differentiable function that satisfies $0 \leq \psi_a(x) \leq |x|$ and $\psi_a(x) \leq \frac{x^2}{a}$ for all $x \in \mathbb{R}$.

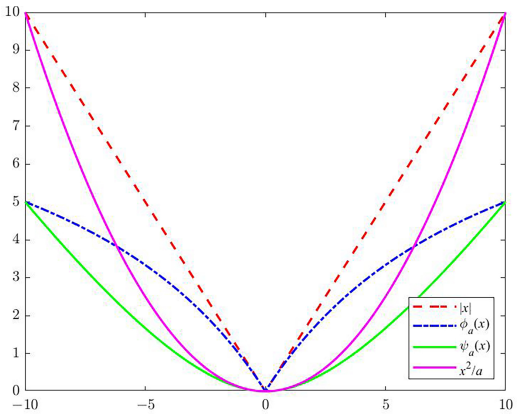


Figure 3. Plots of the $|x|$, $\phi_a(x)$, $\psi_a(x)$, and $\frac{x^2}{a}(a = 10)$ functions.

Then, the MTL1 function can be written as

$$\phi_a(x) = |x| - \psi_a(x). \tag{2.9}$$

To establish our MTL1TV MRI model, we define a multivariate penalty $\Phi_a : \mathbb{R}^n \to \mathbb{R}$ based on the above penalty ϕ_a. For this purpose, we first define a multivariate generalization form of the corresponding function ψ_a in (2.8).

Definition 2. *Let ψ_a be correspondingly defined in (2.8); then $\Psi_a : \mathbb{R}^n \to \mathbb{R}$ is defined as*

$$\Psi_a(\mathrm{v}) = \sum_{i=1}^{n} \psi_a(v_i), \mathrm{v} \in \mathbb{R}^n. \tag{2.10}$$

Remark 1. *The function Ψ_a satisfies $0 \leq \Psi_a(\mathrm{v}) \leq \|\mathrm{v}\|_1$ and $\Psi_a(\mathrm{v}) \leq \frac{1}{a}\|\mathrm{v}\|_2^2$ for all $\mathrm{v} \in \mathbb{R}^n$.*

Then, the multivariate penalty of ϕ_a is given by

$$\Phi_a(\mathrm{v}) = \sum_{i=1}^{n} (|v_i| - \psi_a(v_i)) = \|v\|_1 - \Psi_a(\mathrm{v}), \mathrm{v} \in \mathbb{R}^n \tag{2.11}$$

Finally, by replacing v with gradient Dx in (2.11) (D is the first-order difference operator/matrix), we obtain the definition of MTL1TV below.

Definition 3 (MTL1TV)**.** *The MTL1TV regularizer:* $\|x\|_{MTL1TV} : \mathbb{R}^n \to \mathbb{R}$,

$$\|\mathrm{x}\|_{MTL1TV} := \Phi_a(\mathrm{Dx}) = \|\mathrm{Dx}\|_1 - \Psi_a(\mathrm{Dx}). \tag{2.12}$$

Now, the MR image reconstruction model can be formulated as

$$\min_{\mathrm{x}} \lambda\|\mathrm{x}\|_{MTL1TV} + \frac{1}{2}\left\|\mathrm{y} - \mathrm{Ax}\right\|_2^2, \tag{2.13}$$

where $\lambda > 0$ is called the regularization parameter. Because the MTL1 penalty function is non-convex, the model in (2.13) is non-convex. However, the global convexity of the objective function can be controlled by adjusting parameter a, as shown in Theorem 2 below.

Theorem 2. *Let* $\lambda > 0$, $a > 0$. *Define* $F : \mathbb{R}^n \to \mathbb{R}$ *as*

$$F_a(\mathrm{x}) = \frac{1}{2}\left\|\mathrm{y} - \mathrm{Ax}\right\|_2^2 + \lambda\|\mathrm{x}\|_{MTL1TV}, \tag{2.14}$$

If $a \geq 2\lambda$ *and* $\mathrm{D^TD} \preceq \mathrm{A^TA}$*, then* $F_a(\mathrm{x})$ *is a convex function.*

Proof. Substituting (2.12) into (2.14), we have

$$\begin{aligned}
F_a(\mathrm{x}) &= \frac{1}{2}\left\|\mathrm{y} - \mathrm{Ax}\right\|_2^2 + \lambda\|\mathrm{x}\|_{MTL1TV} \\
&= \frac{1}{2}\left\|\mathrm{y} - \mathrm{Ax}\right\|_2^2 + \lambda\left(\|\mathrm{Dx}\|_1 - \Psi_a(\mathrm{Dx})\right) \\
&= \frac{1}{2}\left[\left\|\mathrm{y}\right\|_2^2 - 2\mathrm{y^TAx} + \|\mathrm{Ax}\|_2^2\right] + \lambda\|\mathrm{Dx}\|_1 - \lambda\Psi_a(\mathrm{Dx}) \\
&= \left[\frac{1}{2}\left\|\mathrm{y}\right\|_2^2 - \mathrm{y^TAx} + \lambda\|\mathrm{Dx}\|_1\right] + \left[\frac{1}{2}\|\mathrm{Ax}\|_2^2 - \lambda\Psi_a(\mathrm{Dx})\right].
\end{aligned}$$

Note that $\frac{1}{2}\left\|\mathrm{y}\right\|_2^2$ does not depend on x and that $\mathrm{y^TAx}$ is linear in x. Hence, the function $F_a(\mathrm{x})$ is convex if $F_2(\mathrm{x})$ is convex, where $F_2(\mathrm{x})$ is defined as

$$\begin{aligned}
F_2(\mathrm{x}) &= \frac{1}{2}\|\mathrm{Ax}\|_2^2 - \lambda\Psi_a(\mathrm{Dx}) \\
&= \frac{1}{2}\left(\|\mathrm{Ax}\|_2^2 - \frac{2\lambda}{a}\|\mathrm{Dx}\|_2^2\right) + \lambda\left(\frac{1}{a}\|\mathrm{Dx}\|_2^2 - \Psi_a(\mathrm{Dx})\right) \\
&= \frac{1}{2}\mathrm{x^T}\left(\mathrm{A^TA} - \frac{2\lambda}{a}\mathrm{D^TD}\right)\mathrm{x} + \lambda\left(\frac{1}{a}\|\mathrm{Dx}\|_2^2 - \Psi_a(\mathrm{Dx})\right).
\end{aligned}$$

The first term is convex if $\mathrm{A^TA} \geq \frac{2\lambda}{a}\mathrm{D^TD}$. Since $a \geq 2\lambda$ and are $\mathrm{D^TD} \preceq \mathrm{A^TA}$ given, it follows that the first term is convex. From Remark 1, the last term is convex. Hence, $F_2(\mathrm{x})$ is convex. Therefore, $F_a(\mathrm{x})$ is convex.

3. ADMM for MTL1TV model

In this section, the optimization algorithm is presented in detail. From Theorem 2, (2.13) is a convex optimization problem under certain conditions, which can be solved via the convex optimization

algorithms. Hence, the ADMM is used to effectively solve the model in (2.13). First, according to the definition of $\|\mathrm{x}\|_{MTL1TV}$, we rewrite (2.13) as follows:

$$\min_{\mathrm{x}} \lambda\Phi_a(\mathrm{Dx}) + \frac{1}{2}\left\|\mathrm{y}-\mathrm{Ax}\right\|_2^2. \tag{3.1}$$

Next, we introduce the auxiliary variable z with the constraint z = Dx. Then, the optimization problem (3.1) is rewritten as

$$\min_{\mathrm{x}} \lambda\Phi_a(\mathrm{z}) + \frac{1}{2}\|y-Ax\|_2^2, \quad s.t.\ z = Dx. \tag{3.2}$$

Under the ADMM framework, the augmented Lagrangian function of (3.2) is

$$L(\mathrm{x},\mathrm{z},\mathrm{w},\beta) = \lambda\Phi_a(\mathrm{z}) + \frac{1}{2}\left\|\mathrm{y}-\mathrm{Ax}\right\|_2^2 - \langle \mathrm{w}, \mathrm{z}-\mathrm{Dx}\rangle + \frac{\beta}{2}\|\mathrm{z}-\mathrm{Dx}\|_2^2 \tag{3.3}$$

$$= \lambda\Phi_a(\mathrm{z}) + \frac{1}{2}\left\|\mathrm{y}-\mathrm{Ax}\right\|_2^2 + \frac{\beta}{2}\left\|\mathrm{z}-\mathrm{Dx}-\frac{\mathrm{w}}{\beta}\right\|_2^2 - \frac{\|\mathrm{w}\|_2^2}{2\beta}, \tag{3.4}$$

where $\mathrm{w} \in \mathbb{R}^n$ is the Lagrange multiplier, and $\beta > 0$ is a penalty parameter. Then, we invoke the ADMM by iterating the variable updates in (3.5) to (3.7):

$$\mathrm{x}^{k+1} = \arg\min_{\mathrm{x}} L(\mathrm{x},\mathrm{z}^{\mathrm{k}},\mathrm{w}^{\mathrm{k}},\beta^{\mathrm{k}}) = \arg\min_{\mathrm{x}}\left\{\frac{1}{2}\left\|\mathrm{y}-\mathrm{Ax}\right\|_2^2 + \left\langle \mathrm{w}^k, \mathrm{Dx}\right\rangle + \frac{\beta^k}{2}\left\|\mathrm{z}^k-\mathrm{Dx}\right\|_2^2\right\}, \tag{3.5}$$

$$\mathrm{z}^{k+1} = \arg\min_{\mathrm{z}} L(\mathrm{x}^{k+1},\mathrm{z},\mathrm{w}^k,\beta^k) = \arg\min_{\mathrm{z}}\left\{\lambda\Phi_a(\mathrm{z}) - \left\langle \mathrm{w}^k, \mathrm{z}\right\rangle + \frac{\beta^k}{2}\left\|\mathrm{z}-\mathrm{Dx}^{k+1}\right\|_2^2\right\}, \tag{3.6}$$

$$\mathrm{w}^{k+1} = \mathrm{w}^k + \beta^k(\mathrm{Dx}^{k+1}-\mathrm{z}^{k+1}), \quad \beta^{k+1} = \theta\beta^k. \tag{3.7}$$

Now, we give the detailed steps for solving the subproblems in (3.5) and (3.6) alternatively.

Step 1. Update x^{k+1} with z^{k}, and keep w^{k} fixed. Let the gradient of this objective function be zero, and obtain the following equation:

$$(\beta\mathrm{D}^{\mathrm{T}}\mathrm{D}+\mathrm{A}^{\mathrm{T}}\mathrm{A})\mathrm{x}^{k+1} = \beta\mathrm{D}^{\mathrm{T}}\mathrm{z}^k+\mathrm{A}^{\mathrm{T}}\mathrm{y}-\mathrm{D}^{\mathrm{T}}\mathrm{w}^k. \tag{3.8}$$

The matrix A = RF, where F is the Fourier operator with the property $\mathrm{F}^{\mathrm{T}} = \mathrm{F}^{-1}$. Since $\mathrm{D}^{\mathrm{T}}\mathrm{D}$ is a cyclic matrix, it can be diagonalized by Fourier transform. Therefore, x^{k+1} can be solved by using two Fourier transforms:

$$\mathrm{x}^{k+1} = (\beta\mathrm{D}^{\mathrm{T}}\mathrm{D}+\mathrm{A}^{\mathrm{T}}\mathrm{A})^{-1}(\beta\mathrm{D}^{\mathrm{T}}\mathrm{z}^k+\mathrm{A}^{\mathrm{T}}\mathrm{y}-\mathrm{D}^{\mathrm{T}}\mathrm{w}^k). \tag{3.9}$$

Step 2. Update z^{k+1} with x^{k+1}, and keep w^{k} fixed. From (3.6), we have

$$z^{k+1} = \arg\min_{z}\left\{\lambda\Phi_a(z) + \frac{\beta}{2}\left\|z-\left(D\mathrm{x}^{k+1}+\frac{\mathrm{w}^k}{\beta}\right)\right\|_2^2\right\}.$$

According to Theorem 1, z^{k+1} is given by the following expression:

$$\mathrm{z}^{k+1} = prox_{\lambda/\beta}^{\Phi_a}(D\mathrm{x}^{k+1}+\frac{\mathrm{w}^k}{\beta}). \tag{3.10}$$

Based on the above analysis, the specific algorithm for solving (3.1) can be summarized as Algorithm 1.

Algorithm 1 ADMM for solving the MTL1TV model

Input: $a > 0, \beta > 0, \lambda > 0, \theta > 1, k = 0, \mathrm{u}^0 = (\mathrm{x}^0, \mathrm{z}^0, \mathrm{w}^0)$, *maxiter*
while not converged, **do**
    Update x^{k+1} via (3.9).
    Update z^{k+1} via (3.10).
    Update w^{k+1} via (3.7).
    Set $\mathrm{u}^{k+1} = (\mathrm{x}^{k+1}, \mathrm{z}^{k+1}, \mathrm{w}^{k+1})$.
    $k = k + 1$.
end while

4. Convergence analysis

In this section, we discuss the convergence of the proposed alternating minimization algorithm. Through this paper, we assume that $ker(A) \cap ker(D) = 0$. The assumption is very reasonable for the imaging problem [31, 32]. In Algorithm 1, the dominant computation is the steps to solve the two minimization subproblems (3.5) and (3.6). Inspired by [31, 32], we have the following convergence result for Algorithm 1.

Assume that $(\mathrm{x}^*, \mathrm{z}^*, \lambda^*)$ is a stationary point that satisfies the first order optimality conditions of (3.6):

$$\begin{aligned} 0 &= \mathrm{A}^{\mathrm{T}}(\mathrm{Ax}^* - \mathrm{y}) + \mathrm{D}^{\mathrm{T}}\mathrm{w}^*, \\ 0 &\in \partial\lambda\Phi_{\mathrm{a}}(\mathrm{z}^*) - \mathrm{w}^*, \\ 0 &= \mathrm{Du}^* - \mathrm{z}^*. \end{aligned} \tag{4.1}$$

Obviously, one can easily find that x^* satisfies

$$0 \in \partial\lambda\Phi_a(\mathrm{z}^*) + \mathrm{A}^{\mathrm{T}}(\mathrm{Ax}^* - \mathrm{y}), \tag{4.2}$$

which is the first order necessary condition of (2.13) about stationary point. Otherwise, by Algorithm 1, each iteration step about the subproblems (3.5)–(3.7) follows:

$$0 = \mathrm{A}^{\mathrm{T}}(\mathrm{Ax}^{k+1} - \mathrm{y}) + \mathrm{D}^{\mathrm{T}}\mathrm{w}^k + \beta^k\mathrm{D}^{\mathrm{T}}(\mathrm{Dx}^{k+1} - \mathrm{z}^k), \tag{4.3}$$

$$0 \in \partial\lambda\Phi_a(\mathrm{z}^{k+1}) - \mathrm{w}^k - \beta^k(\mathrm{Dx}^{k+1} - \mathrm{z}^{k+1}), \tag{4.4}$$

$$\mathrm{w}^{k+1} = \mathrm{w}^k + \beta^k(\mathrm{Dx}^{k+1} - \mathrm{z}^{k+1}). \tag{4.5}$$

In the following, we will prove the sequence $\{\mathrm{x}^k, \mathrm{z}^k, \lambda^k\}$ generated by the Algorithm 1 has a limit point $(\mathrm{x}^*, \mathrm{z}^*, \lambda^*)$ that satisfies (4.1). The following two lemmas state that the metric $L(\mathrm{x}^k, \mathrm{z}^k, \mathrm{w}^k, \beta^k)$ and the sequences x^k and $\{\mathrm{z}^k\}$ are bounded.

Lemma 3. *Let $L(\mathrm{x}^k, \mathrm{z}^k, \mathrm{w}^k, \beta^k)$ be the sequence generated by Algorithm 1, then $\{\mathrm{w}^k\}$ and $\{L(\mathrm{x}^k, \mathrm{z}^k, \mathrm{w}^k, \beta^k)\}$ are bounded.*

Proof. First, we prove the Lagrange multiplier $\mathrm{w}^k \in \mathbb{R}^n$ is bounded and $\|\mathrm{w}^k\| \leq \sqrt{n}\lambda$. Let $d \in \partial\lambda\phi_a(x)$, then, if $x > 0$, $d = \frac{\lambda a^2}{(a+x)^2} \in (0, \lambda)$. If $x < 0$, $d = \frac{-\lambda a^2}{(a-x)^2} \in (-\lambda, 0)$. If $x = 0$, $d \in [\lambda, \lambda]$. Therefore, $|d| \leq \lambda$.

On the other hand, combine (4.4) and (4.5), and we have $\mathrm{w}^{k+1} \in \partial\lambda\Phi_a(\mathrm{z}^{k+1})$. Thus, by $|d| \le \lambda$, we have $\|\mathrm{w}^{k+1}\|_\infty \le \lambda$. Because w^{k+1} is finite dimensional, it implies that $\{\mathrm{w}^k\}$ is bounded and $\|\mathrm{w}^k\| \le \sqrt{n}\lambda < \infty$.

Then, we prove $\{L(\mathrm{x}^k, \mathrm{z}^k, \mathrm{w}^k, \beta^k)\}$ is bounded. By (4.5), we notice that

$$\mathrm{D}\mathrm{x}^{k+1} - \mathrm{z}^{k+1} = \frac{\mathrm{w}^{k+1} - \mathrm{w}^k}{\beta^k}. \tag{4.6}$$

From the above formula and (3.7), we have

$$\begin{aligned} L(\mathrm{x}^{k+1}, \mathrm{z}^{k+1}, \mathrm{w}^{k+1}, \beta^{k+1}) - L(\mathrm{x}^{k+1}, \mathrm{z}^{k+1}, \mathrm{w}^{k+1}, \beta^k) &= \frac{\beta^{k+1}}{2}\|\mathrm{z}^{k+1} - \mathrm{D}\mathrm{x}^{k+1}\|_2^2 - \frac{\beta^k}{2}\|\mathrm{z}^{k+1} - \mathrm{D}\mathrm{x}^{k+1}\|_2^2 \\ &= \frac{\beta^{k+1} - \beta^k}{2(\beta^k)^2}\|\mathrm{w}^{k+1} - \mathrm{w}^k\|_2^2, \end{aligned} \tag{4.7}$$

$$\begin{aligned} L(\mathrm{x}^{k+1}, \mathrm{z}^{k+1}, \mathrm{w}^{k+1}, \beta^k) - L(\mathrm{x}^{k+1}, \mathrm{z}^{k+1}, \mathrm{w}^k, \beta^k) &= \langle \mathrm{w}^k - \mathrm{w}^{k+1}, \mathrm{z}^{k+1} - \mathrm{D}\mathrm{x}^{k+1}\rangle \\ &= \frac{1}{\beta^k}\|\mathrm{w}^{k+1} - \mathrm{w}^k\|_2^2. \end{aligned} \tag{4.8}$$

By the definitions of z^{k+1} and x^{k+1}, we know that

$$L(\mathrm{x}^{k+1}, \mathrm{z}^{k+1}, \mathrm{w}^k, \beta^k) - L(\mathrm{x}^{k+1}, \mathrm{z}^k, \mathrm{w}^k, \beta^k) \le 0, \tag{4.9}$$

$$L(\mathrm{x}^{k+1}, \mathrm{z}^k, \mathrm{w}^k, \beta^k) - L(\mathrm{x}^k, \mathrm{z}^k, \mathrm{w}^k, \beta^k) \le 0. \tag{4.10}$$

Summing (4.7)–(4.10), we have

$$\begin{aligned} L(\mathrm{x}^{k+1}, \mathrm{z}^{k+1}, \mathrm{w}^{k+1}, \beta^{k+1}) - L(\mathrm{x}^k, \mathrm{z}^k, \mathrm{w}^k, \beta^k) &\le \frac{\beta^{k+1} + \beta^k}{(2\beta^k)^2}\|\mathrm{w}^{k+1} - \mathrm{w}^k\|_2^2 \\ &\le \frac{2(\theta+1)n\lambda^2}{\beta^0\theta^k}, \end{aligned} \tag{4.11}$$

Summing up the above inequality from $k = 0$, we obtain

$$L(\mathrm{x}^{k+1}, \mathrm{z}^{k+1}, \mathrm{w}^{k+1}, \beta^{k+1}) - L(\mathrm{x}^0, \mathrm{z}^0, \mathrm{w}^0, \beta^0) \le \frac{2n(\theta+1)\lambda^2(1 - \frac{1}{\theta^{k+1}})}{\beta^0(1 - \frac{1}{\theta})}, \tag{4.12}$$

Because $\theta > 1$, let $k \to \infty$, and we have

$$\lim_{k\to\infty} L(\mathrm{x}^{k+1}, \mathrm{z}^{k+1}, \mathrm{w}^{k+1}, \beta^{k+1}) < L(\mathrm{x}^0, \mathrm{z}^0, \mathrm{w}^0, \beta^0) + \frac{2n\theta(\theta+1)\lambda^2}{\beta^0(\theta-1)} < \infty, \tag{4.13}$$

On the other hand, from $\lim_{k\to\infty} \beta^k = \infty$, and the fact that $\{\mathrm{w}^{k+1}\}$ is bounded, we know

$$\lim_{k\to\infty} \|\mathrm{D}\mathrm{x}^{k+1} - \mathrm{z}^{k+1}\|_2^2 = \lim_{k\to\infty} \left\| \frac{\mathrm{w}^{k+1} - \mathrm{w}^k}{\beta^k} \right\|_2^2 = 0, \tag{4.14}$$

and then

$$\lim_{k\to\infty} L(\mathrm{x}^{k+1}, \mathrm{z}^{k+1}, \mathrm{w}^{k+1}, \beta^{k+1}) \ge \lim_{k\to\infty} \langle \mathrm{w}^{k+1}, \mathrm{D}\mathrm{x}^{k+1} - \mathrm{z}^{k+1}\rangle = 0. \tag{4.15}$$

By (4.13) and (4.15), we know

$$0 \le \lim_{k\to\infty} L(\mathrm{x}^{k+1}, \mathrm{z}^{k+1}, \mathrm{w}^{k+1}, \beta^{k+1}) < \infty. \tag{4.16}$$

So, $\{L(\mathrm{x}^{k+1}, \mathrm{z}^{k+1}, \mathrm{w}^{k+1}, \beta^{k+1})\}$ is bounded.

Lemma 4. *Let* $\{(\mathrm{x}^k, \mathrm{z}^k)\}$ *be the sequence generated by Algorithm 1, then the sequences* $\{\mathrm{x}^k\}$ *and* $\{\mathrm{z}^k\}$ *are bounded.*

Proof. From (4.9) and (4.10), we obtain that $L(\mathrm{x}^{k+1}, \mathrm{z}^{k+1}, \mathrm{w}^k, \beta^k)$ is upper bounded. Next, we know that $L(\mathrm{x}^k, \mathrm{z}^k, \mathrm{w}^k, \beta^k)$ is strongly convex about the variable x, and the following inequality holds:

$$L(\mathrm{x}^{k+1}, \mathrm{z}^k, \mathrm{w}^k, \beta^k) - L(\mathrm{x}^k, \mathrm{z}^k, \mathrm{w}^k, \beta^k) \leq -\frac{c^k}{2}\|\mathrm{x}^{k+1} - \mathrm{x}^k\|_2^2. \tag{4.17}$$

Summing (4.7)–(4.9) and (4.17), we have

$$L(\mathrm{x}^{k+1}, \mathrm{z}^{k+1}, \mathrm{w}^{k+1}, \beta^{k+1}) - L(\mathrm{x}^k, \mathrm{z}^k, \mathrm{w}^k, \beta^k) \leq -\frac{c^k}{2}\|\mathrm{x}^{k+1} - \mathrm{x}^k\|_2^2 + \frac{2(\theta+1)n\lambda^2}{\beta^0\theta^k}. \tag{4.18}$$

Summing the above inequality from $k = 0$, we have

$$L(\mathrm{x}^{k+1}, \mathrm{z}^{k+1}, \mathrm{w}^{k+1}, \beta^{k+1}) - L(\mathrm{x}^0, \mathrm{z}^0, \mathrm{w}^0, \beta^0) \leq \frac{2(\theta+1)n\lambda^2}{\beta^0(\theta-1)} - \sum_{0\leq i\leq k}\frac{c^i}{2}\|\mathrm{x}^{i+1} - \mathrm{x}^i\|_2^2, \tag{4.19}$$

which together with (4.16) yields

$$\lim_{k\to\infty}\|\mathrm{x}^{k+1} - \mathrm{x}^k\|_2^2 = 0. \tag{4.20}$$

Otherwise, if $\lim_{k\to\infty}\|\mathrm{x}^{k+1} - \mathrm{x}^k\|_2^2 \neq 0$, then $\sum_{0\leq i\leq k}\frac{c^i}{2}\|\mathrm{x}^{i+1} - \mathrm{x}^i\|_2^2 = +\infty$; thus,

$$\lim_{k\to\infty} L(\mathrm{x}^{k+1}, \mathrm{z}^{k+1}, \mathrm{w}^{k+1}, \beta^{k+1}) = -\infty,$$

and (4.16), a contradiction. By (4.20), we know that $\{x^k\}$ is a Cauchy sequence and, thus, is convergent and bounded.

Furthermore, we show that $\{\mathrm{z}^k\}$ is bounded. By (3.4) and (4.12), we have

$$\begin{aligned}\lambda\Phi_a(\mathrm{z}^k) + \frac{1}{2}\left\|\mathrm{y} - \mathrm{A}\mathrm{x}^k\right\|_2^2 &= L(\mathrm{z}^k, \mathrm{x}^k, \mathrm{w}^{k-1}, \beta^{k-1}) + \frac{\|\mathrm{w}^{k-1}\|_2^2}{2\beta^{k-1}} - \frac{\beta^{k-1}}{2}\left\|\mathrm{z}^k - \mathrm{D}\mathrm{x}^k - \frac{\mathrm{w}^{k-1}}{\beta^{k-1}}\right\|_2^2 \\ &= L(\mathrm{z}^k, \mathrm{x}^k, \mathrm{w}^{k-1}, \beta^{k-1}) + \frac{1}{2\beta^{k-1}}(\|\mathrm{w}^{k-1}\|_2^2 - \|\mathrm{w}^k\|_2^2).\end{aligned}$$

Because $\lim_{k\to\infty}\beta^k = \infty$ and $\{\mathrm{w}^k\}$ are bounded, we also know that $\{\mathrm{z}^k\}$ and $\{\mathrm{A}\mathrm{x}^k\}$ are bounded.

Now, we are ready for proving the following convergence result.

Theorem 3. *Let* $\{\mathrm{x}^k, \mathrm{z}^k, \mathrm{w}^k\}$ *be the sequence generated by Algorithm 1, then, there exists a subsequence* $\{\mathrm{x}^{k_j}, \mathrm{z}^{k_j}, \mathrm{w}^{k_j}\}$*, which converges to a stationary point* $(\mathrm{x}^*, \mathrm{z}^*, \mathrm{w}^*)$ *and satisfies (4.1).*

Proof. By Lemmas 3 and 4, the sequence $\{\mathrm{x}^k, \mathrm{z}^k, \mathrm{w}^k\}$ is bounded, so there exists a subsequence $\{\mathrm{x}^{k_j}, \mathrm{z}^{k_j}, \mathrm{w}^{k_j}\}$ that converges to a cluster point $(\mathrm{x}^*, \mathrm{z}^*, \mathrm{w}^*)$. From the lower semi-continuity of $L(\mathrm{x}, \mathrm{z}, \mathrm{w}, \beta)$,

$$\liminf_{j\to\infty} L(\mathrm{x}^{k_j+1}, \mathrm{z}^{k_j+1}, \mathrm{w}^{k_j}, \beta^{k_j}) \geq L(\mathrm{x}^*, \mathrm{z}^*, \mathrm{w}^*, \beta^\infty)$$

and as z^{k_j+1} is a minimizer of function $L(x^{k_j+1}, z^{k_j+1}, w^{k_j}, \beta^{k_j})$ with respect to z,

$$\limsup_{j\to\infty} L(x^{k_j+1}, z^{k_j+1}, w^{k_j}, \beta^{k_j}) \le L(x^*, z^*, w^*, \beta^\infty).$$

By the above two inequalities, we get

$$\lim_{j\to\infty} \lambda\Phi_a(z^{k_j}) = \lambda\Phi_a(z^*).$$

One can immediately verify that

$$\begin{aligned} 0 &= A^T(Ax^* - y) + D^T w^*, \\ 0 &\in \partial\lambda\Phi_a(z^*) - w^*, \\ 0 &= Du^* - z^*. \end{aligned}$$

Therefore, (x^*, z^*, w^*) is a Karush-Kuhn-Tucker point of the problem (3.2).

Remark 2. *From Corollary 1, if $a \ge 2\lambda$, the two subproblems (3.5) and (3.6) are continuous, coercive, and strictly convex. Consequently, every subproblem has a global solution. Additionally, from Theorem 2, we know that the objective function is a convex function, which implies the stationary point is the global minimum point.*

5. Experiments

In this section, we present experiments conducted to verify the performance of the MTL1TV method in MATLAB 2019a on a laptop with 32 GB RAM and an Intel Core i7-12700H CPU at 2.70 GHz. We compared our proposed method with the standard TV [11], MCTV [24], and transformed total variation (TTV) [33] methods. For fair performance comparison, the parameters used in comparing algorithms have been appropriately tuned to yield a better reconstruction effect. We used three metrics to assess the quality and accuracy of image reconstruction: the relative error (RE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM). These metrics are defined as follows:

$$\mathrm{PSNR} = 10\lg\frac{255^2}{\|x^k - x\|_2}, \quad \mathrm{RE} = \frac{\|x^k - x\|_2}{\|x\|_2}.$$

where x, x^k are the original image and reconstructed image, respectively, and

$$\mathrm{SSIM} = \frac{(2\mu_x\mu_{x^k} + C_1)(2\sigma_{xx^k} + C_2)}{(\mu_x^2 + \mu_{x^k}^2 + C_1)(\sigma_x^2 + \sigma_{x^k}^2 + C_2)},$$

where μ_x and μ_{x^k} are the mean of x and x^k, k is the number of iterations in the algorithm, σ^2 and $\sigma_{x^k}^2$ denote the variance of x and x^k, and σ_{xx^k} denotes the covariance of x and x^k. The positive constants C_1 and C_2 are stabilization constants. Generally, the MR image reconstruction is better if it has a lower RE and a higher PSNR and SSIM.

The sampling templates and the experimental data are shown in Figure 4. The test data (a)–(c) was three typical MR images: Shepp-Logan, Brain, and Brain 2; (c) and (d) are two different brain MR images. The sampling templates (d)–(f) were radial, random, and Cartesian sampling, respectively. The size of sampling masks and MRI data was 256×256. In the experiments, the stopping criterion was usually as follows:

$$\frac{\|\mathrm{x}^{k+1} - \mathrm{x}^{k}\|_2}{\|\mathrm{x}^{k+1}\|_2} \leq 10^{-4},$$

or the maximum number of iteration 200.

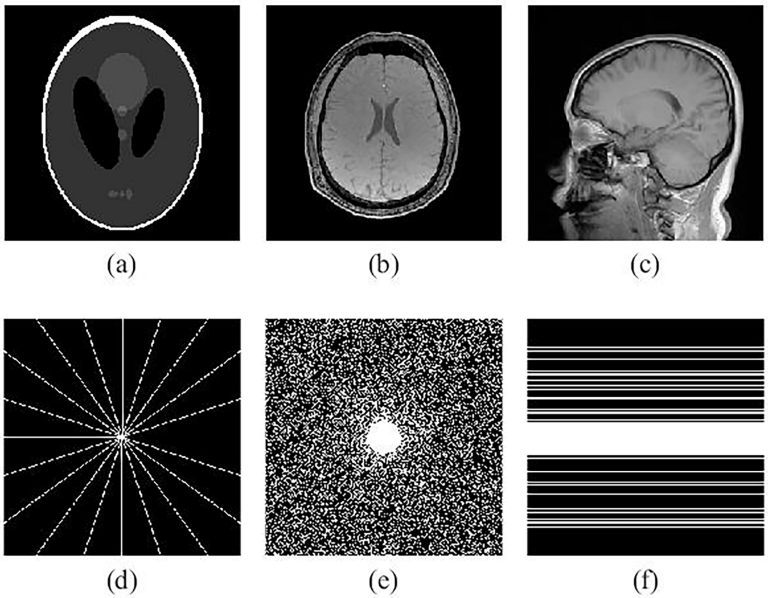


Figure 4. Experimental data and sampling patterns. (a) Shepp-Logan Phantom; (b) Brain; (c) Brain 2; (d) Radial sampling; (e) Random sampling; (f) Cartesian sampling.

5.1. Experiments on noiseless data

The sampling masks and MR data shown in Figure 4 are used to evaluate our proposed MTL1TV method. Table 1 displays the PSNR, SSIM, RE, and CPU time required to reconstruct various MR images using different methods with varying sampling templates. As observed in Table 1, the time required to reconstruct an MR image using TV is shorter than other methods, but its performance is inferior. Furthermore, when compared with TV, MCTV, and TTV, it is evident that MTL1TV exhibits higher PSNR and SSIM values with lower RE. This suggests that MTL1TV offers improved MR image reconstruction quality, simultaneously maintaining a reasonable runtime as shown in Table 1.

The visual comparison of reconstruction results and the error images are shown in Figures 5−7. In Figure 5, the Shepp-Logan phantom is used to demonstrate the performance of the proposed method. A Cartesian sampling at a sampling rate of 34% is employed to compare with the four reconstruction models proposed above. The regularization parameters λ for TV, MCTV, TTV, and MTL1TV are set to 0.0001, 0.005, 0.001, and 0.005, respectively. For TTV and MTL1TV, the parameter a is set to 1 and 0.05, respectively. From the Figure 5, it is observed that compared with other methods, MTL1TV has the best reconstruction capability, and the reconstructed image is most similar to the original image in

the visual effects, which is reflected by the comparison of the difference images.

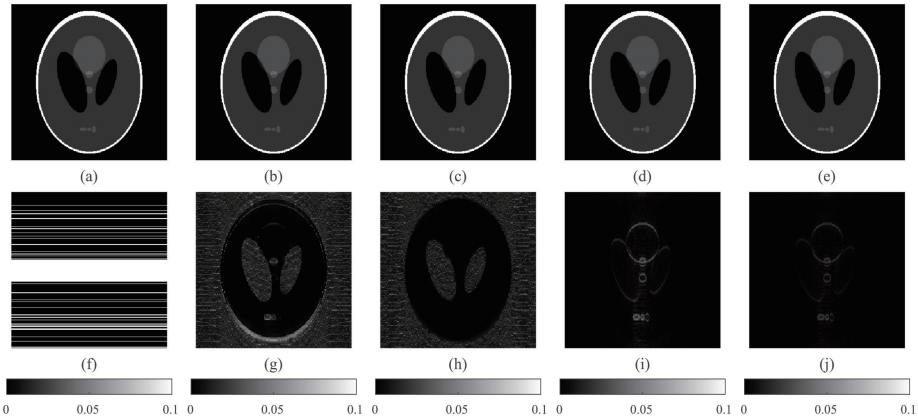


Figure 5. Results of Shepp-Logan reconstruction using a Cartesian mask with a sampling rate of 34%. (a) Original image; (f) Cartesian sampling; (b)–(e) reconstructed images using TV, MCTV, TTV, and MTL1TV, respectively; (g)–(i) difference images between (a) and (b)–(e).

Figure 6 displays the reconstruction results of brain data under radial sampling with 10 trajectory lines and a 3% sampling ratio. We set the regularization parameters λ of TV, MCTV, TTV, and MTL1TV to 0.000001, 0.000005, 0.001, and 0.001, respectively. The value of parameter a is set as 0.5 for TTV, while it is set as 0.1 for MTL1TV. Visually, there is no significant difference. The residual images illustrate that the proposed method can visually gain better reconstructions.

For the Brain 2 image, random sampling with a 30% sampling rate and a 0.1 sampling radius was employed to test the image. The regularization parameters λ were 0.00001, 0.00001, 0.0001, and 0.0005. The value of a in TTV was $a = 0.5$, while in MTL1TV, $a = 0.1$. The reconstructed images and errors are shown in Figure 7. It shows that the reconstruction results of the four methods are similar, and MTL1TV is slightly better than the other three methods.

Additionally, the convergence of the proposed method is empirically tested. In Figure 8, we illustrate the behavior of the proposed MTL1TV model in the iteration process. It respectively shows the curves of the RE, PSNR, and SSIM values on Shepp-Logan phantom data under a radial mask with a 3% sampling ratio versus iteration numbers. It is shown that as the number of iterations increased, the RE value generated by MTL1TV gradually decreased, while the PSNR and SSIM values increased, which exhibits the stable convergence trend in the subsequent iterations.

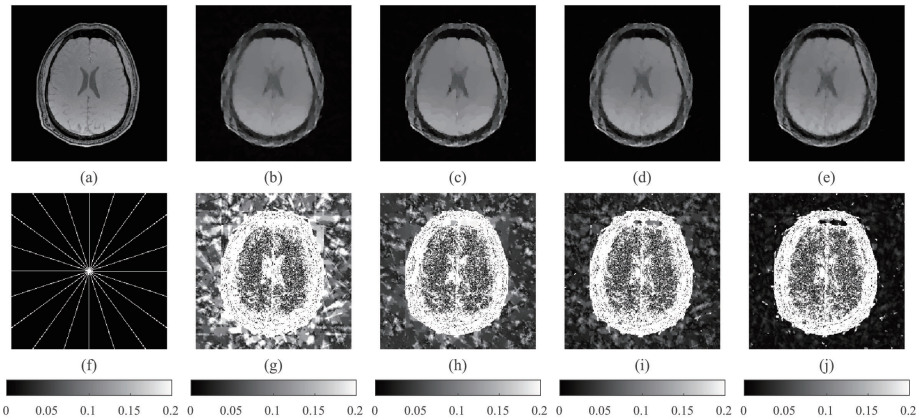


Figure 6. Results of brain reconstruction using a radial mask with a sampling rate of 3%. (a) Original image; (f) Radial sampling; (b)–(e) reconstructed images using TV, MCTV, TTV, and MTL1TV, respectively; (g)–(i) difference images between (a) and (b)–(e).

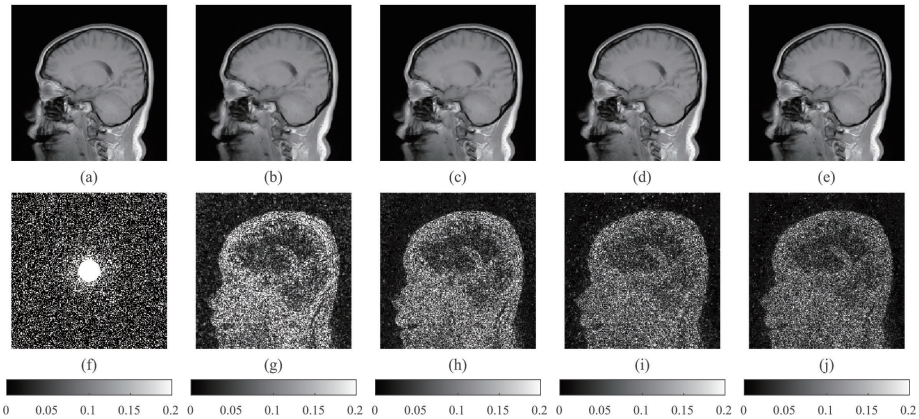


Figure 7. Results of Brain 2 reconstruction using a random mask with a sampling rate of 30%. (a) Original image; (f) Random sampling; (b)–(e) reconstructed images using TV, MCTV, TTV, and MTL1TV, respectively; (g)–(i) difference images between (a) and (b)–(e).

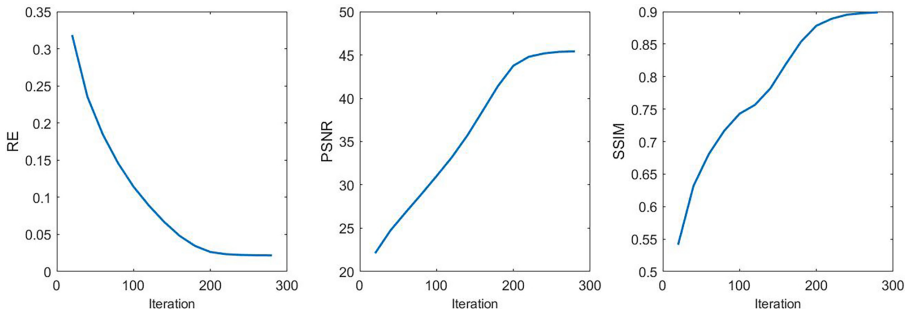


Figure 8. Metric curves versus iteration number.

Table 1. Numerical results of the four algorithms.

Test image	Template	Method	RE (%)	PSNR (dB)	SSIM	CPU time (s)
Shepp-Logan	Radial sampling 3%	TV	15.71	28.2474	0.6794	**1.649427**
		MCTV	5.44	37.4617	0.7120	9.922404
		TTV	4.36	39.3778	0.7731	2.089988
		MTL1TV	**2.74**	**43.4180**	**0.8824**	2.081267
	Random sampling 30%	TV	0.76	54.5952	0.9805	**1.644969**
		MCTV	0.59	56.7759	0.9858	9.789923
		TTV	0.09	72.6855	0.9996	2.157613
		MTL1TV	**0.05**	**78.7386**	**0.9999**	2.144944
	Cartesian sampling 34%	TV	0.70	55.2138	0.9849	**1.711586**
		MCTV	0.53	57.6397	0.9904	9.709789
		TTV	0.12	70.5970	**0.9999**	2.839062
		MTL1TV	**0.04**	**79.7220**	**0.9999**	2.107535
Brain	Radial sampling 3%	TV	22.28	23.4557	0.4981	**1.648866**
		MCTV	21.36	23.8205	0.6275	9.888602
		TTV	20.88	24.0185	0.7004	2.198074
		MTL1TV	**19.78**	**24.4883**	**0.7877**	2.186878
	Random sampling 30%	TV	3.69	38.5007	0.8673	**1.645874**
		MCTV	2.84	41.3783	0.9171	10.657604
		TTV	1.23	48.6261	0.9924	3.549396
		MTL1TV	**1.08**	**49.8008**	**0.9956**	2.450209
	Cartesian sampling 34%	TV	6.98	33.5339	0.8921	**1.655883**
		MCTV	5.64	35.3929	0.9243	9.734209
		TTV	5.64	35.3861	0.9374	2.348402
		MTL1TV	**5.40**	**35.7689**	**0.9478**	2.347148
Brain2	Radial sampling 3%	TV	18.88	22.7062	0.5848	**1.651263**
		MCTV	17.91	23.1698	0.6415	9.983066
		TTV	16.36	23.9466	0.7137	2.173014
		MTL1TV	**16.28**	**23.9891**	**0.7244**	2.149936
	Random sampling 30%	TV	3.61	37.0681	0.9468	**1.682748**
		MCTV	2.99	38.7059	0.9646	9.940363
		TTV	2.58	39.9759	0.9723	2.426169
		MTL1TV	**2.37**	**40.7350**	**0.9771**	2.414485
	Cartesian sampling 34%	TV	6.74	31.6495	0.9096	**1.643168**
		MCTV	6.17	32.4159	0.9240	9.893311
		TTV	6.29	32.2556	0.9208	2.343147
		MTL1TV	**6.10**	**32.5116**	**0.9241**	2.389379

5.2. The Parameters selection

In the proposed model, there are two important parameters: λ and a. The λ is the regularization parameter, which trades the sparsity and data consistency. The parameter a has an important impact on

the properties of regularization term and it also affects the quality of reconstructed images. Therefore, it is crucial to choose appropriate values of a and λ, which can help us obtain better reconstruction results. The authors in [22] claimed that $a = 1$ is the best among the tests on signals, and the author in [33] empirically selects $a = 5$, which achieves better results in image reconstruction tasks. To demonstrate how to select parameters a and λ, we present the reconstruction results of the Shepp-Logan phantom (256×256) data using a Cartesian mask with a sampling rate of 34% as an example.

To begin, we conduct the experiments with the fixed parameter $a = 1$ and $a = 5$,while allowing λ to vary. Subsequently, the parameter λ in TV, MCTV, and TTV can also be selected accordingly. The quantitative comparison results are listed in Table 2. From Table 2, we can see that when $\lambda = 0.005$, the performance of the proposed model is the "best". Not only that, when $\lambda = 0.0001$ in TV, $\lambda = 0.005$ in MCTV, $\lambda_1 = 0.001$ in TTV (when $a = 1$), and $\lambda_2 = 0.0005$ in TTV (when $a = 5$), and the reconstructed metric PSNR achieves the highest value. Therefore, we set the parameter $\lambda = 0.0001$ in TV, and $\lambda = 0.005$ in MCTV.

Next, we conduct the experiment with the fixed parameter λ_1 and λ_2, respectively, to asses the impact of different values of parameter a on reconstruction performance. From Table 3, it can be clearly seen that when the $a = 1$, TTV exhibits excellent performance. Similarly, when $a = 0.05$, MTL1TV achieves the best results. Furthermore, according to Table 3, it can be observed that when $a = 0.05 > 2\lambda_2 = 0.01$, the convexity of the MTL1TV model is clearly indicated. In Figure 9, it can be seen that as the parameter a assumes different values, the differences between the reconstructed MR image and the original image also vary. Notably,when $a = 0.05$, the error images are the smallest.

Table 2. Reconstruction results of four models under different λ.

λ	0.00005	0.0001	0.0005	0.001	0.005	0.01
TV	55.1473	**55.2138**	54.9330	54.3397	47.3366	39.9643
MCTV	56.6046	56.6744	57.0797	57.1284	**57.6397**	44.8390
TTV($a = 1$)	28.6023	29.6150	43.5340	**70.5970**	55.9932	49.4821
MTL1TV($a = 1$)	28.1268	28.6025	33.3470	43.5340	**61.9251**	55.9932
TTV($a = 5$)	28.2477	28.8494	35.0278	47.6527	**55.8338**	49.9436
MTL1TV($a = 5$)	28.1495	28.6462	33.4636	43.1281	**57.4406**	51.5726

Table 3. Reconstruction results of models TTV and MTL1TV under different a.

Method(λ)	$a = 0.005$	$a = 0.01$	$a = 0.05$	$a = 0.1$	$a = 0.5$	$a = 1$	$a = 5$
TTV(λ_1)	50.5093	53.2998	68.8556	67.2232	69.3398	**70.5970**	47.6527
TTV(λ_2)	42.9035	42.9824	47.4019	47.6102	54.6126	55.9932	55.8338
MTL1TV(λ_2)	28.3835	29.8307	**79.7220**	74.5064	64.5339	61.9251	57.4406

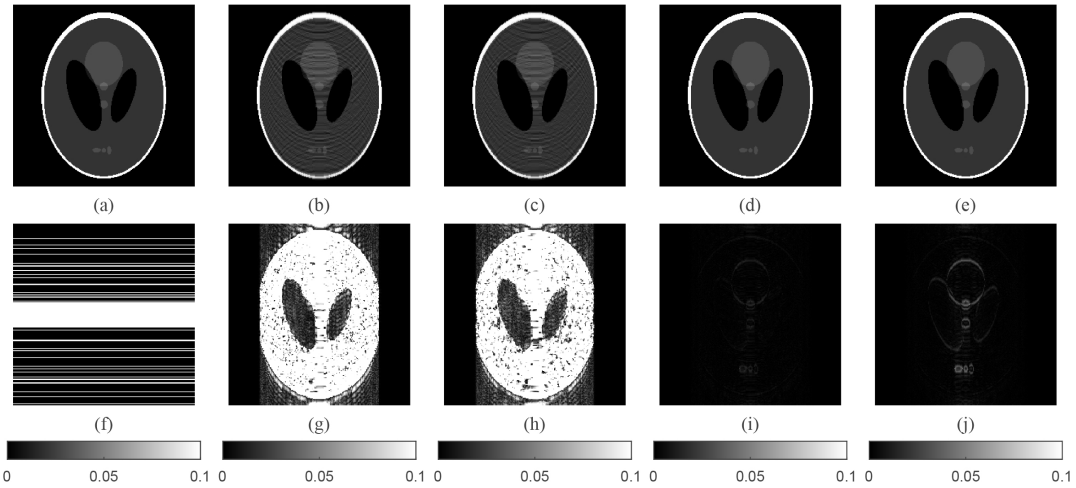


Figure 9. Results of Shepp-Logan reconstruction using a Cartesian mask with a sampling rate of 34%. (a) Original image; (f) Cartesian sampling; (b)–(e) reconstructed images using $a = 0.005$, $a = 0.01$, $a = 0.05$, and $a = 0.1$, respectively; (g)–(i) difference images between (a) and (b)–(e).

5.3. Experiments on noisy data

To demonstrate the performance of the proposed model in noisy condition, we conduct experiments on sampled data contaminated with additive Gaussian noise. In these experiments, Gaussian noise with standard deviation ($\sigma = 0.02$) was added to the real and imaginary parts of sampled data, respectively. Figure 10 presents that the reconstructed results, along with the corresponding errors, for Shepp-Logan data under radial sampling (10 trajectory lines and a 3% sampling ratio). When compared to the reconstruction errors, it is evident that our proposed method effectively suppresses noise and artifacts. Table 4 presents the quantitative results of the three images (in Figure 4) with additive noise under radial sampling (10 trajectory lines and a 3% sampling ratio). Even though all models were given identical parameter settings in both noisy and noise-free environments, it was observed that the values of PSNR, SSIM, and RE for three images decreased under noisy conditions, as shown in Table 4. Nevertheless, they still produced results that were comparable to their noise-free counterparts. When compared to other methods, our method consistently delivered superior results, highlighting its robustness.

6. Conclusions

In this paper, we introduced a non-convex TV regularization model for undersampled MRI reconstruction. In the new model, based on the MTL1 function, the traditional TV is replaced with the MTL1TV norm, thus improving the fitting performance of the model. We used the ADMM algorithm to compute the minimization problem. The experimental results demonstrated the effectiveness and efficiency of the MTL1TV method. In our future research, we seek to delve into the utilization of the proposed methodology in dynamic MRI, tomographic imaging, and electrical impedance tomography, by amalgamating deep learning frameworks with variational models.

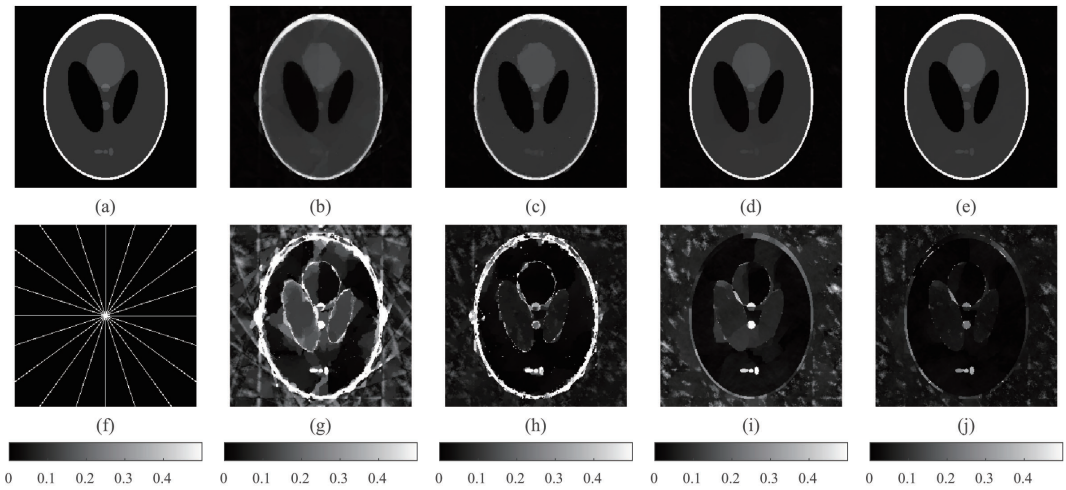


Figure 10. Reconstruction of Shepp-Logan with additive noise ($\sigma = 0.02$) using a radial mask with a sampling rate of 3%. (a) Original image; (f) radial sampling; (b)–(e) reconstructed images using TV, MCTV, TTV, and MTL1TV, respectively; (g)–(i) difference images between (a) and (b)–(e).

Table 4. Numerical results for MR image reconstruction with additive noise from Radial mask.

Test image	Method	RE (%)	PSNR(dB)	SSIM
Shepp-Logan	TV	21.48	25.5308	0.6446
	MCTV	17.00	27.5653	0.8167
	TTV	4.74	38.6586	0.7574
	MTL1TV	**3.11**	**42.8507**	**0.8710**
Brain	TV	22.31	23.4433	0.4966
	MCTV	21.46	23.7823	0.6257
	TTV	20.91	24.0075	0.6948
	MTL1TV	**19.89**	**24.4386**	**0.7808**
Brain2	TV	6.80	31.5799	0.9069
	MCTV	6.27	32.2734	**0.9214**
	TTV	6.41	32.0904	0.9171
	MTL1TV	**6.25**	**32.3101**	0.9209

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by Natural Science Foundation of Hubei Province (2023AFD013), Philosophy and Social Sciences of Educational Commission of Hubei Province of China (22Y109), and Foundation of Hubei Normal University (2022055).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. I. Pykett, J. Newhouse, F. Buonanno, T. Brady, M. Goldman, J. Kistler, et al., Principles of nuclear magnetic resonance imaging, *Radiology*, **143** (1982), 157–168. https://doi.org/10.1148/radiology.143.1.7038763
2. X. Gu, W. Xue, Y. Sun, X. Qi, X. Luo, Y. He, Magnetic resonance image restoration via least absolute deviations measure with isotropic total variation constraint, *Math. Biosci. Eng.*, **20** (2023), 10590–10609. https://doi.org/10.3934/mbe.2023468
3. Y. Beauferris, J. Teuwen, D. Karkalousos, N. Moriakov, M. Caan, G. Yiasemis, et al., Multi-coil MRI reconstruction challenge assessing brain MRI reconstruction models and their generalizability to varying coil configurations, *Front. Neurosci.*, **16** (2022), 1–16. https://doi.org/10.3389/fnins.2022.919186
4. M. Lustig, D. Donoho, J. M. Pauly, Sparse MRI: The application of compressed sensing for rapid MR imaging, *Magn. Reson. Med.*, **58** (2007), 1182–1195. https://doi.org/10.1002/mrm.21391
5. J. He, Q. Liu, A. Christodoulou, C. Ma, F. Lam, Z. Liang, Accelerated high-dimensional MR imaging with sparse sampling using low-rank tensors, *IEEE Trans. Med. Imaging*, **35** (2016), 2119–2129. https://doi.org/10.1109/TMI.2016.2550204
6. A. Tran, T. Nguyen, P. Doan, D. Tran, D. Tran, Parallel magnetic resonance imaging acceleration with a hybrid sensing approach, *Math. Biosci. Eng.*, **18** (2021), 2288–2302. https://doi.org/10.3934/mbe.2021116
7. F. Knoll, K. Hammernik, C. Zhang, S. Moeller, T. Sodickson, D. Sodickson, et al., Deep-Learning Methods for Parallel Magnetic Resonance Imaging Reconstruction: A Survey of the Current Approaches, Trends, and Issues, *IEEE Signal Process. Mag.*, **37** (2020), 128–140. https://doi.org/10.1109/MSP.2019.2950640
8. D. Donoho, Compressed sensing, *IEEE Trans. Inf. Theory*, **52** (2006), 1289–1306. https://doi.org/10.1109/TIT.2006.871582
9. J. Ye, Compressed sensing MRI: a review from signal processing perspective, *BMC Biomed. Eng.*, **1** (2019), 1–17. https://doi.org/10.1186/s42490-019-0006-z
10. X. Li, R. Feng, F. Xiao, Y. Yin, D. Cao, X. Wu, et al., Sparse reconstruction of magnetic resonance image combined with two-step iteration and adaptive shrinkage factor, *Math. Biosci. Eng.*, **19** (2022), 13214–13226. https://doi.org/10.3934/mbe.2022618
11. L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, *Phys. D*, **60** (1992), 259–268. https://doi.org/10.1016/0167-2789(92)90242-F
12. I. Selesnick, Sparse regularization via convex analysis, *IEEE Trans. Signal Process.*, **65** (2017), 4481–4494. https://doi.org/10.1109/TSP.2017.2711501
13. D. Peleg, R. Meir, A bilinear formulation for vector sparsity optimization, *Signal Process.*, **88** (2008), 375–389. https://doi.org/10.1016/j.sigpro.2007.08.015

14. J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, *J. Am. Stat. Assoc.*, **96** (2001), 1348–1360. https://doi.org/10.1198/016214501753382273

15. C. Zhang, Nearly unbiased variable selection under minimax concave penalty, *Ann. Stat.*, **38** (2010), 894–942. https://doi.org/10.1214/09-AOS729

16. F. Zhang, H. Wang, W. Qin, X. Zhao, J. Wang, Generalized nonconvex regularization for tensor RPCA and its applications in visual inpainting, *Appl. Intell.*, **53** (2023), 23124–23146. https://doi.org/10.1007/s10489-023-04744-9

17. J. Zou, M. Shen, Y. Zhang, H. Li, G. Liu, S. Ding, Total variation denoising with non-convex regularizers, *IEEE Access*, **7** (2018), 4422–4431. https://doi.org/10.1109/ACCESS.2018.2888944

18. M. Shen, J. Li, T. Zhang, J. Zou, Magnetic resonance imaging reconstruction via non-convex total variation regularization, *Int. J. Imaging Syst. Technol.*, **31** (2021), 412–424. https://doi.org/10.1002/ima.22463

19. I. Selesnick, I. Bayram, Sparse signal estimation by maximally sparse convex optimization, *IEEE Trans. Signal Process.*, **62** (2014), 1078–1092. https://doi.org/10.1109/TSP.2014.2298839

20. S. Zhang, J. Xin, Minimization of transformed L_1 penalty: Closed form representation and iterative thresholding algorithms, *Commun. Math. Sci.*, **15** (2017), 511–537. https://doi.org/10.4310/CMS.2017.v15.n2.a9

21. S. Zhang, J. Xin, Minimization of transformed L_1 penalty: theory, difference of convex function algorithm, and robust application in compressed sensing, *Math. Program.*, **169** (2018), 307–336. https://doi.org/10.1007/s10107-018-1236-x

22. H. Li, Q. Zhang, A. Cui, J. Peng, Minimization of fraction function penalty in compressed sensing, *IEEE Trans. Neural Networks Learn. Syst.*, **31** (2020), 1626–1637. https://doi.org/10.1109/TNNLS.2019.2921404

23. J. Li, Z. Xie, G. Liu, L. Yang, J. Zou, Diffusion optical tomography reconstruction based on convex-nonconvex graph total variation regularization, *Math. Meth. Appl. Sci.*, **23** (2023), 4534–4545. https://orcid.org/0000-0001-7897-7151

24. Y. Liu, H. Du, Z. Wang, W. Mei, Convex MR brain image reconstruction via no-convex total variation minimization, *Int. J. Imaging Syst. Technol.*, **28** (2018), 246–253. https://doi.org/10.1002/ima.22275

25. Z. Luo, Z. Zhu, B. Zhang, An AtanTV nonconvex regularization model for MRI reconstruction, *J. Sens.*, **2022** (2022), 1–15. https://doi.org/10.1155/2022/1758996

26. Z. Luo, Z. Zhu, B. Zhang, An SCADTV nonconvex regularization approach for magnetic resonance imaging, *IAENG Int. J. Comput. Sci.*, **48** (2021), 1005–1012. https://api.semanticscholar.org/CorpusID:248818745

27. Y. Lu, B. Zhang, Z. Zhu, Y. Liu, A CauchyTV non-convex regularization model for MRI reconstruction, *Signal, Image Video Process.*, **17** (2023), 3275–3282. https://doi.org/10.1007/s11760-023-02542-x

28. S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Found. Trends Mach. Learn.*, **3** (2011), 1–122. https://doi.org/10.1561/2200000016

29. D. G. Luenberger, Y. Ye, *Linear and Nonlinear Programming*, 3^{nd} edition, Springer, USA, 2008. https://doi.org/10.1007/978-3-319-18842-3

30. J. Yang, Y. Zhang, W. Yin, A fast alternating direction method for TVL1-L2 signal reconstruction from partial fourier data, *IEEE J. Sel. Top. Signal Process.*, **4** (2010), 288–297. https://doi.org/10.1109/JSTSP.2010.2042333

31. B. Zhang, G. Zhu, Z. Zhu, S. Kwong, Alternating direction method of multipliers for nonconvex log total variation image restoration, *Appl. Math. Modell.*, **114** (2023), 338–359. https://doi.org/10.1016/j.apm.2022.09.018

32. J. You, Y. Jiao, X. Lu, T. Zeng, A nonconvex model with minimax concave penalty for image restoration, *J. Sci. Comput.*, **78** (2019), 1063–1086. https://doi.org/10.1007/s10915-018-0801-z

33. L. Huo, W. Chen, H. Ge, M. K. Ng, Stable image reconstruction using transformed total variation minimization, *SIAM J. Imaging Sci.*, **15** (2022), 1104–1139. https://doi.org/10.1137/21M1438566

Appendix A. Proof of Lemma 2

Proof. 1) Define the new variable $\eta = a + x$ and substitute it into polynomial (2.6), then it becomes

$$\eta^3 - (a+t)\eta^2 + \lambda a^2 = 0, \tag{A.1}$$

whose discriminant is:

$$\Delta = \lambda a^2[4(a+t)^3 - 27\lambda a^2].$$

Due to $t > t_1^*$ and $\Delta > 0$, the cubic equation has three distinct real roots. We change variables as $\eta = y + \frac{a}{3} + \frac{t}{3} = x + a$. The relation between x and y is: $x = y - \frac{2a}{3} + \frac{t}{3}$. In terms of t, the cubic polynomial is turned into a depressed cubic as:

$$y^3 - \frac{(a+t)^2}{3}y + \lambda a^2 - \frac{2(a+t)^3}{27} = 0.$$

The three roots in trigonometric form are:

$$\begin{aligned} y_0 &= \tfrac{a+t}{3}\cos\left(\tfrac{\varphi}{3}\right) \\ y_1 &= \tfrac{2(a+t)}{3}\cos\left(\tfrac{\varphi}{3} + \tfrac{\pi}{3}\right) \\ y_2 &= -\tfrac{2(a+t)}{3}\cos\left(\tfrac{\pi}{3} - \tfrac{\varphi}{3}\right) \end{aligned} \tag{A.2}$$

where $\varphi = \arccos\left(1 - \frac{27\lambda a^2}{2(a+t)^3}\right)$. It is obvious that $y_0 > y_1 > y_2$ and $y_2 < 0$. By the relation $x = y - \frac{2a}{3} + \frac{t}{3}$, the three roots in variable x are $x_i = y_i - \frac{2a}{3} + \frac{t}{3}$. From these formulas, we can prove $x_0 > x_1 > x_2$ and $x_0 \le t$. Then, the largest root is x_0, i.e., $x_0 = y_0 - \frac{2a}{3} + \frac{t}{3} = g_\lambda(t)$.

2) We set: $\eta = a - x$ and $y = \eta - \frac{a}{3} + \frac{t}{3}$. So, $x = -y + \frac{2a}{3} + \frac{t}{3}$. By a similar analysis as in part (1), there are three distinct roots for the polynomial equation: $x_0 < x_1 < x_2$ with the smallest solution

$$x_0 = -\frac{2(a-t)}{3}\cos\left(\frac{\varphi}{3}\right) + \frac{x}{3} + \frac{2a}{3},$$

where $\varphi = \arccos\left(1 - \frac{27\lambda a^2}{2(a-t)^3}\right)$. Therefore, $x_0 = g_\lambda(t)$, when $t < -t_1^*$.

Appendix B. Proof of Theorom 1

Proof. We discuss $t = 0$, $t > 0$, and $t < 0$ in the following:

1) $t = 0$: In this case, optimization objective function is $\theta_{\lambda,t}(x) = \lambda\phi_a(x) + \frac{1}{2}x^2$. It is true that $\lambda\phi_a(x)$ and $\frac{1}{2}x^2$ are both increasing for $x > 0$, and decreasing with $x < 0$. Thus, $\theta(0)$ is the unique minimizer for function $\theta_{\lambda,t}(x)$. So $x^* = 0$, when $t = 0$.
2) $t > 0$: Since $\frac{1}{2}(x - t)^2$ and $\lambda\phi_a(x)$ are both decreasing with $x < 0$, our optimal solution will only be obtained at nonnegative values. Thus, we just need to consider $x \geq 0$. When $x \geq 0$, we obtain

$$\theta'_{\lambda,t}(x) = \frac{\lambda a^2}{(a + x)^2} + x - t,$$

and

$$\theta''_{\lambda,t}(x) = 1 - \frac{2\lambda a^2}{(a + x)^3}.$$

It is clear that $\theta''_{\lambda,t}(x)$ is increasing, and $\theta''_{\lambda,t}(0) = 1 - \frac{2\lambda}{a}$ determines the convexity for the function $\theta_{\lambda,t}(x)$. In the following proof, we further discuss the value of x^* by two conditions: $\lambda < \frac{a}{2}$ and $\lambda > \frac{a}{2}$.

a) $\lambda < \frac{a}{2}$: So, we have $\inf_{x>0} \theta''_{\lambda,t}(x) = \theta''_{\lambda,t}(0+) = 1 - \frac{2\lambda}{a} \geq 0$, which means function $\theta'_{\lambda,t}(x)$ is increasing for $x \geq 0$, with minimum value $\theta'_{\lambda,t}(0) = \lambda - t = t_2^* - t$.
   i) When $0 \leq x \leq t_2^*$, $\theta'_{\lambda,t}(x)$ is always positive, thus the optimal value $x^* = 0$.
   ii) When $x > t_2^*$, $\theta'_{\lambda,t}(x)$ is first negative, then positive, and $x > t_2^* > t_1^*$. The unique positive stationary point x^* of $\theta'_{\lambda,t}(x)$ satisfies equation: $\theta'_{\lambda,t}(x^*) = 0$, which implies

$$x(a + x)^2 - t(a + x)^2 + \lambda a^2 = 0. \tag{B.1}$$

   According to Lemma 2, the optimal value is $x^* = x_0 = g_\lambda(t)$.
   Above all, the value of x^* is

$$x^* = \begin{cases} 0, & 0 \leq t \leq t_2^*; \\ g_\lambda(t), & x > t_2^* \end{cases}$$

   under the condition $\lambda \leq \frac{a}{2}$.

b) $\lambda > \frac{a}{2}$: In this case, due to the sign of $\theta''_{\lambda,t}(x)$, we know that function $\theta'_{\lambda,t}(x)$ is decreasing at first then switches to be increasing. Its minimum point is $\overline{x} = \sqrt[3]{2\lambda a^2} - a$ and the least value

$$\theta'_{\lambda,t}(\overline{x}) = \frac{\lambda a^2}{(a + \overline{x})^2} + \overline{x} - t = t_1^* - t.$$

   Then, $\theta'_{\lambda,t}(x) \geq t_1^* - t$ with $x \geq 0$.
   i) When $0 \leq t \leq t_1^*$, function $\theta_{\lambda,t}(x)$ is always increasing. Thus, optimal value is $x^* = 0$.
   ii) When $t \geq t_2^*$, $\theta_{\lambda,t}(0+) \leq 0$. Thus, the function $\theta_{\lambda,t}(x)$ is first decreasing and then increasing. There is only one positive optimal stationary point, which is also the optimal solution. Using the Lemma 2, we know that $x^* = \theta_{\lambda,t}(x)$.

iii) When $t_2^* < t < t_1^*$, $\theta_{\lambda,t}(0+) > 0$. Thus, the function $\theta_{\lambda,t}(x)$ is first increasing, then decreasing and finally increasing, which implies that there are two positive stationary points and the larger one is a local minima. Using Lemma 2, the local minimized point will be $x_0 = \theta_{\lambda,t}(x)$, the largest root of (1.1). Since $x_0 - t + \frac{\lambda a^2}{(a+x_0)^2} = 0$, which implies $\frac{\lambda a}{a+x_0} = \frac{(t-x_0)(a+x_0)}{a}$, we have

$$\begin{aligned}\theta_{\lambda,t}(x_0) - \theta_{\lambda,t}(0) &= \frac{1}{2}x_0^2 - x_0 t + \frac{\lambda a x_0}{a+x_0} \\ &= x_0(\frac{1}{2}x_0 - t+)\frac{\lambda a}{a+x_0} \\ &= x_0(\frac{1}{2}x_0 - t + \frac{(t-x_0)(a+x_0)}{a}) \\ &= x_0^2(\frac{t-x_0}{a} - \frac{1}{2}).\end{aligned}$$

Define a function $h(t) = t - g_\lambda(t) - \frac{a}{2}$.
First, we prove that $t = t_3^*$ is a solution to $h(t) = 0$. Since $\lambda > \frac{a}{2}$, $t_3^* = \sqrt{2\lambda a} - \frac{a}{2} > 0$. Thus:

$$\cos(\varphi(t_3^*)) = 1 - \frac{27\lambda a^2}{2(a+t_3^*)^3} = 1 - \frac{27\lambda a^2}{2(\frac{a}{2} + \sqrt{2\lambda a})^3}.$$

Further, by using the relation $\cos(\varphi) = 4\cos^3(\frac{\varphi}{3}) - 3\cos(\frac{\varphi}{3})$ and $0 \le \frac{\varphi}{3} \le \frac{\pi}{3}$, we have

$$\cos(\frac{\varphi(t_3^*)}{3}) = \frac{\sqrt{2\lambda a} - \frac{a}{4}}{\frac{a}{2} + \sqrt{2\lambda a}}.$$

Plugging this formula into $g_\lambda(t_3^*)$ shows that $g_\lambda(t_3^*) = \sqrt{2\lambda a} - a = t_3^* - \frac{a}{2}$. Therefore, t_3^* is the root for $h(t)$ and $t_3^* \in (t_1^*, t_2^*)$.
Second, we prove that the function $h(t)$ changes sign at point $t = t_3^*$. We prefer to discuss it in two cases.
Case 1: $x \in (t_3^*, t_2^*)$. According to Lemma 2, we know that $g_\lambda(t)$ is the largest root of the cubic polynomial $H(y) = y(a+y)^2 - t(a+y)^2 + \lambda a^2$ under the condition of $t > t_1^*$.
For function $H(y)$, we have $H(t) = \lambda a^2$ and

$$H(t - \frac{a}{2}) = \lambda a^2 - \frac{a}{2}(\frac{a}{2} + t)^2.$$

Due to $t > t_3^* = \sqrt{2\lambda a} - \frac{a}{2}$ and $H(t - \frac{a}{2}) < 0$, there is a root $y = g_\lambda(t)$, such that $y = g_\lambda(t) \in (t - \frac{a}{2}, t)$ for the equation $H(y) = 0$. That is, $t - g_\lambda(t) < \frac{a}{2}$, and, thus, $h(t) < 0$.
Case 2: $x \in (t_1^*, t_3^*)$. We have $H(t - \frac{a}{2}) > 0$ and $H(t) > 0$. Due to the proof in Lemma 2, one possible situation is that there are two roots x_0 and x_1 within interval $(t - \frac{a}{2}, t)$. However, we can exclude this case. This is because

$$\begin{aligned}x_0 - x_1 &= \frac{2(a+t)}{3}\{\cos(\frac{\varphi}{3}) - \cos(\frac{\varphi}{3} + \frac{\pi}{3})\} \\ &= \frac{2(a+t)}{3}\{2\sin(\frac{\varphi}{3}) + \frac{\pi}{6})\sin(\frac{\pi}{6})\}\end{aligned}$$

$$= \frac{2(a+t)}{3}\sin(\frac{\varphi}{3}+\frac{\pi}{6}).$$

Furthermore, $x_0 - x_1 > \frac{a}{2}$ holds for $t > t_1^* > \frac{a}{2}$ and $\lambda > \frac{a}{2}$. This is a contradiction of the assumption that x_0 and x_1 are in $(t-\frac{a}{2}, t)$. Thus, $H(y) = 0$ has no root in $(t-\frac{a}{2}, t)$. Therefore, $x_0 = g_\lambda(t) < t - \frac{a}{2}$. That is to say, $h(t) > 0$.

From the discussion earlier, it is true that the optimal solution $x^* = 0$, if $0 < t \le t_3^*$, and $x^* = x_0 = g_\lambda(t)$, if $t > t_3^*$.

To sum up, we know that under the condition $\lambda > \frac{a}{2}$,

$$x^* = \begin{cases} 0, & 0 \le t \le t_3^*; \\ g_\lambda(t), & x > t_3^* \end{cases}$$

3) $t < 0$: Because

$$\inf_x \theta_{\lambda,t}(x) = \inf_x \theta_{\lambda,t}(-x) = \inf_x \frac{1}{2}(x-|t|)^2 + \phi_a(t),$$

$x^*(t) = -x^*(-t)$, which implies that the formula obtained when $t > 0$ above can extend to the case: $t < 0$ by odd symmetry.

Summarizing results from all cases, the proof is completed.

Appendix C. Proof of Corollary 1

Proof. It is clear to see that the function

$$\theta_{\lambda,t}(x) = \lambda\phi_a(x) + \frac{1}{2}(x-t)^2$$

is differentiable on $\mathbb{R}\backslash\{0\}$. For $x \neq 0$, the derivative of $\theta_{\lambda,t}(x)$ is given by

$$\theta'_{\lambda,t}(x) = \frac{\lambda a^2}{(a+|x|)^2}\mathrm{sgn}(x) + (x-t), x \neq 0.$$

Let us find the range of a for which $\theta_{\lambda,t}(x)$ is convex. Notice that

$$\phi_a(0^+) = 1 > -1 = \phi_a(0^-)$$

and

$$\theta''_{\lambda,t}(x) = 1 - \frac{2\lambda a^2}{(a+x)^3}, \forall x > 0.$$

When we set $a \ge 2\lambda$, the function $\theta'_{\lambda,t}$ is increasing, then the function $\theta_{\lambda,t}(x)$ is strictly convex.