



Research article

Performance analysis of the convex non-convex total variation denoising model

Yating Zhu, Zixun Zeng, Zhong Chen, Deqiang Zhou and Jian Zou*

School of Information and Mathematics, Yangtze University, Jingzhou, Hubei, China

* **Correspondence:** Email: zoujian@yangtzeu.edu.cn.

Abstract: Total variation (TV) regularization is a powerful tool in image denoising, but it often exhibits limited performance in preserving edges. In contrast, non-convex TV regularization can more effectively preserve edges and contours, albeit posing challenges when solving. Recently, the convex non-convex (CNC) strategy has emerged as a potent approach that allows incorporating non-convex TV regularization terms while maintaining the overall convexity of the objective function. Its superior performance has been validated through various numerical experiments; however, theoretical analysis remains lacking. In this paper, we provided theoretical analysis of the performance of the CNC-TV denoising model. By utilizing the oracle inequality, we derived an improved upper bound on its performance compared to TV regularization. In addition, we devised an alternating direction method of multipliers (ADMM) algorithm to address the proposed model and verified its convergence properties. Our proposed model has been validated through numerical experiments in 1D and 2D denoising, demonstrating its exceptional performance.

Keywords: total variation regularization; convex non-convex strategy; performance analysis; ADMM; image denoising

Mathematics Subject Classification: 49J40, 65K10, 68U10

1. Introduction

Image denoising aims to reduce or eliminate the noise in the image degradation process while maintaining the basic features of the image [1]. Mathematically, the image degradation process can be expressed as

$$y = \theta^* + \varepsilon, \tag{1.1}$$

where $\theta^* \in \mathbb{R}^n$ is the original image, $y \in \mathbb{R}^n$ is the observed image, and ε is the additive white Gaussian noise (AWGN) with variance σ^2 , i.e., $\varepsilon \sim N(0, \sigma^2 I_n)$. The concept of total variation (TV) is

a mathematical measure that quantifies the level of variation or change present in an image, measuring how rapidly the pixel intensities change across neighboring pixels. In cases where an image contains noise, its pixel intensities exhibit random fluctuations, resulting in high TV values. By minimizing the TV of an image while preserving crucial features such as edges and textures, we can effectively reduce noise and enhance its quality [2–7]. The standard TV estimator $\hat{\theta}$ can be mathematically defined as

$$\hat{\theta} = \arg \min_{\theta} \frac{1}{2n} \|y - \theta\|_2^2 + \lambda \|\theta\|_{TV}, \quad (1.2)$$

where regularization parameter $\lambda > 0$ is used to control the trade-off between the fidelity term $\frac{1}{2n} \|y - \theta\|_2^2$ and the TV regularization $\|\theta\|_{TV}$. n is the pixel size of the image. By introducing a first-order difference matrix $D \in R^{n \times n}$, the TV regularization $\|\theta\|_{TV}$ can be represented by the ℓ_1 norm, i.e., $\|\theta\|_{TV} = \|D\theta\|_1$.

Due to the convexity of the ℓ_1 norm, the TV denoising model (1.2) possesses several advantages, such as having an objective function devoid of redundant local minima and ensuring a unique minimum. However, owing to the inherent property of biased estimation associated with the ℓ_1 norm, this model may also underestimate the magnitude of signal discontinuity [8–10].

The limitations of TV regularization can be overcome by employing non-convex TV regularization, which not only enhances image texture details but also sharpens edges more effectively [11–13]. This non-convex TV denoising model can be described as follows:

$$\hat{\theta} = \arg \min_{\theta} \frac{1}{2n} \|y - \theta\|_2^2 + \lambda \Psi^{NC}(D\theta), \quad (1.3)$$

where $\Psi^{NC}(\cdot)$ is a non-convex function such as the ℓ_q norm with $0 < q < 1$ [14, 15]. However, the cost function of (1.3) is usually non-convex and often has an extraneous local minimum, which brings challenges when solving.

Recently, Selesnick and Lanza et al. proposed a novel non-convex TV regularization for signal and image denoising [9]. This type of non-convex regularization enables the cost function to become convex by adjusting the non-convex control parameters, thereby addressing the limitations of TV regularization and avoiding unnecessary local minima in the cost function. The concept, later known as convex non-convex total variation (CNC-TV) [9, 16, 17], has been widely utilized in various fields such as medical image reconstruction [18–20], fault diagnosis [21, 22], and computer vision [23, 24]. The advantages of CNC-TV are twofold: First, the utilization of non-convex regularization not only effectively suppresses noise but also preserves image edge and texture information. Second, by adjusting the non-convexity control parameters, the global convexity of the objective function can be ensured, thereby enabling the attainment of the global optimal value through convexity optimization algorithms [25, 26]. However, to the best of our knowledge, most of the research on CNC-TV focuses on algorithm design and applications, with a lack of theoretical analysis of performance such as [27]. This motivates us to conduct a thorough analysis on the performance of the CNC-TV estimator to substantiate that CNC-TV regularization surpasses TV regularization.

In this paper, we consider the following CNC-TV denoising model:

$$\hat{\theta} = \arg \min_{\theta} \frac{1}{2n} \|y - \theta\|_2^2 + \lambda \Psi_B^{CNC}(D\theta). \quad (1.4)$$

The matrix parameter B in (1.4) not only controls the non-convexity of the non-convex regularization term $\Psi_B^{CNC}(D\theta)$, but also guarantees the global convexity of the cost function in (1.4).

Through rigorous theoretical analysis and comprehensive experimental evaluations, we show that the utilization of non-convex TV regularization significantly enhances the performance of the CNC-TV estimator, enabling accurate estimation of signal discontinuity. In summary, our contributions can be succinctly summarized as follows:

- Theoretically, for the CNC-TV denoising model, we first establish the conditions necessary to ensure global convexity of the cost function. Subsequently, by employing the oracle inequality, we derive an improved upper bound on performance compared to that of the standard TV denoising model.
- Algorithmically, we derive an ADMM algorithm that ensures convergence to a critical point for the proposed CNC-TV denoising model.
- Empirically, we demonstrate that the proposed CNC-TV denoising model outperforms the standard TV denoising model in both 1D and 2D denoising experiments, owing to its utilization of non-convex regularizations.

The paper is structured as follows. In Section 2, we present some preliminaries and related works related to non-convex TV regularization, performance analysis results, and convex optimization algorithms. In Section 3, we give a theoretical analysis of the performance of the CNC-TV denoising model and further propose an effective ADMM algorithm to solve it. The superiority of the proposed model is validated in Section 4 through both 1D and 2D denoising experiments. Finally, the conclusions are summarized in Section 5.

Notation: Throughout this paper, for a matrix A , A^T and A^\dagger represent its transpose and Moore-Penrose pseudo-inverse, respectively. σ_{\max} and σ_{\min} denote the maximum eigenvalue and minimum eigenvalue of matrix A . $\ker(A)$ and $\ker(A)^\perp$ denote the kernel space and orthogonal complement space of matrix A . Moreover, $0 < A$ indicates that matrix A is positive definite, while $0 \leq A$ implies that matrix A is positive semi-definite. Then we use the symbol $\mathcal{O}(n)$ to indicate the efficiency or complexity of comparing different algorithms. For a set $T \subset \{1, 2, \dots, n\}$, $|T|$ and T^c denote its cardinality and complement. For $\theta \in \mathbb{R}^n$, its ℓ_1 norm, ℓ_2 norm, and ℓ_∞ norm can be defined as $\|\theta\|_1 = \sum_{i=1}^n |\theta_i|$, $\|\theta\|_2 = \sqrt{\sum_{i=1}^n \theta_i^2}$, and $\|\theta\|_\infty = \max_{1 \leq i \leq n} |\theta_i|$. For a function $f(\theta)$, we use $\nabla f(\theta)$ to denote the gradient or subdifferential of $f(\theta)$.

2. Primary and related work

2.1. Non-convex TV regularization

Standard TV denoising can be typically formulated as a convex optimization problem that incorporates an ℓ_1 regularization term. However, it has a drawback in the sense that it tends to underestimate the magnitudes of abrupt changes in the signal. To address this limitation, non-convex TV_q ($0 < q < 1$) regularization is proposed, which can be expressed as $\|D\theta\|_q$ ($0 < q < 1$) [14, 15]. The effectiveness of edge preservation in TV_q is further enhanced by incorporating non-convex sparse regularization. However, due to its non-convex nature, the TV_q regularization introduces challenges in algorithm design as it may lead to convergence toward local optima [28].

By leveraging the advantages of non-convex regularization and convex optimization techniques, Selesnick and Lanza et al. [9] proposed a CNC design strategy for designing non-convex TV regularization, thereby ensuring the global convexity of the objective function. In essence, this is achieved by subtracting its smooth version from the TV regularization, with different smoothing

techniques yielding distinct forms of non-convex TV regularization [29]. The most notable among these is the generalized Moreau envelope TV (GME-TV) regularization [9], which can be mathematically defined as follows:

$$\Psi_B^{GME}(D\theta) = \|D\theta\|_1 - \arg \min_v \left\{ \|v\|_1 + \frac{1}{2} \|B(D\theta - v)\|_2^2 \right\}. \quad (2.1)$$

The last term in (2.1), denoted as

$$S_B(D\theta) = \arg \min_v \left\{ \|v\|_1 + \frac{1}{2} \|B(D\theta - v)\|_2^2 \right\}, \quad (2.2)$$

represents the generalized Moreau envelope of $\|D\theta\|_1$, which is a commonly used smoothing technique. The matrix parameter B is utilized to adjust the smoothness of $S_B(D\theta)$ and consequently affects the non-convexity of $\Psi_B^{GME}(D\theta)$. Specifically, when $B = 0$, then $S_B(D\theta) = 0$, and $\Psi_B^{GME}(D\theta)$ reduces to the TV regularization term $\|D\theta\|_1$. When $B^T B$ is a diagonal matrix, i.e. $B = \sqrt{b}I$, then $\Psi_B^{GME}(D\theta)$ simplifies to the generalized minimax-concave TV (GMC-TV) regularization introduced in [30].

2.2. Performance analysis results

In accordance with the conventions followed in the literature on image denoising [27, 31–33], the performance of a TV estimator $\hat{\theta}$ is evaluated by the high-probability upper bound on its mean squared error (MSE), which is defined by

$$\frac{1}{n} \|\hat{\theta} - \theta^*\|_2^2. \quad (2.3)$$

The presentation of upper bounds is accomplished through the utilization of oracle inequalities, which serve as theoretical assurances for the estimator $\hat{\theta}$ in terms of its ability to capture the underlying structure of θ^* and adapt accordingly.

Let n represent the total number of pixels in the image. Mammen and Van De Geer [31] first defined total variation by total derivatives and achieved an estimation rate of order $\mathcal{O}(n^{-3/5})$. Wang et al. [33] investigated the more comprehensive framework of trend filtering on graphs and obtained a rate of order $\mathcal{O}(n^{-4/5})$. Hütter and Rigollet [27] established sharp oracle inequalities with a rate of $\mathcal{O}(n^{-1/2})$. Sadhanala et al. [34] demonstrated that the penalized TVD estimator achieves an optimal near minimax rate of $\mathcal{O}(n^{-1/2})$. Chatterjee and Goswami [35] analyzed the recovery of piecewise rectangular images through the constrained version of 2D total variation denoising, providing a rate of $\mathcal{O}(n^{-3/4})$. Varma et al. [36] determined the statistical error rates associated with first-order stationary points in graph trend filtering, employing non-convex regularizations like smoothly clipped absolute deviation (SCAD) and minimax concave penalty (MCP). A more detailed summary of the oracles results on TV denoising can be found in [37–39].

The analysis of error rates in our study utilizes techniques from [27] that yield precise error rates through the use of oracle inequalities. Here, we present some basic definitions and conclusions from [27], which will also be used in the theoretical analysis of this paper.

We first give the definitions of the compatibility factor and inverse scaling factor as follows.

Definition 2.1. ([27], Definition 1) Let D be a first-order difference matrix, and the compatibility factor of D for a set $T \subseteq [m]$ is defined as

$$\kappa_\emptyset := 1, \quad \kappa_T = \kappa_T(D) := \inf_{\theta} \frac{\sqrt{|T|} \|\theta\|_2}{\|(D\theta)_T\|_1} \quad \text{for } T \neq \emptyset. \quad (2.4)$$

Moreover, the inverse scaling factor of D is defined as

$$\varrho = \varrho(D) := \max_{j \in [m]} \|s_j\|_2. \quad (2.5)$$

Then the sharp oracle inequality for TV denoising in [27] is expressed as Theorem 2.1.

Theorem 2.1. ([27], Theorem 2) Fix $\delta \in (0, 1)$, $T \subset [m]$, and let D be the first-order difference matrix. Define the regularization parameter

$$\lambda := \frac{1}{n} \sigma \varrho \sqrt{2 \log \left(\frac{em}{\delta} \right)}. \quad (2.6)$$

With this choice of λ , and for $\forall \bar{\theta} \in \mathbb{R}^n$, the TV estimator $\hat{\theta}$ satisfies

$$\frac{1}{n} \|\hat{\theta} - \theta^*\|_2^2 \leq \inf_{\bar{\theta}} \left\{ \frac{1}{n} \|\bar{\theta} - \theta^*\|^2 + 4\lambda \|(D\bar{\theta})_{T^c}\|_1 \right\} + \frac{8\sigma^2}{n} \left(\frac{|T|\varrho^2}{\kappa_T^2} \log \left(\frac{em}{\delta} \right) + \log \left(\frac{e}{\delta} \right) \right), \quad (2.7)$$

on the estimation error with a probability of at least $1 - 2\delta$.

2.3. Convex optimization algorithms

For the following optimization problem:

$$\min_{\theta} \{f_1(\theta) + f_2(\theta)\}, \quad (2.8)$$

where $f_1(\theta)$ is convex and differentiable, $f_2(\theta)$ is a proper closed, convex, but nonsmooth regularization term. The proximal gradient descent (PGD) [40] and alternating direction method of multipliers (ADMM) [41] algorithms exhibit excellent performance in solving such problems, with theoretical guarantees of convergence. The iterative formulas for these algorithms for solving (2.8) are represented as

$$\theta^{k+1} = \text{prox}_{\alpha f_2}(\theta^k - \alpha \nabla f_1(\theta^k)), \quad (\text{PGD})$$

and

$$\begin{cases} \theta^{k+1} = \text{prox}_{\alpha f_1}(z^k + u^k), \\ z^{k+1} = \text{prox}_{\alpha f_2}(\theta^{k+1} - u^k), \\ u^{k+1} = u^k + \theta^{k+1} - z^{k+1}. \end{cases} \quad (\text{ADMM})$$

Furthermore, $\text{prox}_f(\cdot)$ is the proximal operator of f , which is defined as

$$\text{prox}_f(\theta) = \arg \min_v \left\{ f(v) + \frac{1}{2} \|\theta - v\|_2^2 \right\}. \quad (2.9)$$

As a special case, if $f(\theta) = \lambda \|\theta\|_1$, then its proximal operator $\text{prox}_{\lambda \|\cdot\|_1}$ is the element-wise soft thresholding function

$$\text{prox}_{\lambda \|\cdot\|_1}(\theta_i) = \text{sgn}(\theta_i) \max\{|\theta_i| - \lambda, 0\}, \quad (2.10)$$

where θ_i is the i -th element of θ .

3. CNC-TV

In this section, we analyze the CNC-TV denoising model both theoretically and algorithmically. In particular, we consider the following GME-TV denoising model:

$$\hat{\theta} = \arg \min_{\theta} \frac{1}{2n} \|y - \theta\|_2^2 + \lambda \Psi_B^{GME}(D\theta), \quad (3.1)$$

where the GME-TV regularization $\Psi_B^{GME}(D\theta)$ is defined as (2.1), and the matrix parameter B can influence the non-convexity of $\Psi_B^{GME}(D\theta)$ and further the global convexity of the cost function in (3.1).

We first give the convexity condition of the cost function in (3.1), and then we analyze the performance of the GME-TV estimator. Finally, we solve the GME-TV denoising model (3.1) by the ADMM algorithm and prove the convergence.

3.1. Convexity condition

In this subsection, we verify that the cost function in (3.1) can be guaranteed to be convex by selecting the appropriate non-convex control parameter B .

Theorem 3.1. *Let $\theta \in \mathbb{R}^n$, $\lambda > 0$, $B \in R^{p \times n}$, define the loss function of (3.1) as*

$$f_B^{GME}(\theta) = \frac{1}{2n} \|y - \theta\|_2^2 + \lambda \|D\theta\|_1 - \lambda \min_v \left\{ \frac{1}{2} \|B(D\theta - v)\|_2^2 + \|v\|_1 \right\}. \quad (3.2)$$

Then, $f_B^{GME}(\theta)$ is convex if $D^T B^T B D \leq (1/\lambda)I$, and $f_B^{GME}(\theta)$ is strong convex if $D^T B^T B D < (1/\lambda)I$.

Proof. We first rewrite (3.2) as follows:

$$\begin{aligned} f_B^{GME}(\theta) &= \frac{1}{2n} \|\theta - y\|_2^2 + \lambda \|D\theta\|_1 - \lambda \min_v \left\{ \|v\|_1 + \frac{1}{2} \|B(D\theta - v)\|_2^2 \right\} \\ &= \frac{1}{2n} \|\theta - y\|_2^2 + \lambda \|D\theta\|_1 + \frac{\lambda}{2} \|BD\theta\|_2^2 - \lambda \min_v \left\{ \|v\|_1 + \frac{1}{2} \|Bv\|_2^2 - v^T B^T B(D\theta) \right\} \\ &= \frac{1}{2n} \theta^T (I - \lambda(BD)^T (BD)) \theta + \frac{1}{2n} \|y\|_2^2 - \frac{1}{n} \theta^T y + \lambda \|D\theta\|_1 \\ &\quad + \lambda \max_v \left\{ -\|v\|_1 + (Bv)^T B(D\theta) - \frac{1}{2} \|Bv\|_2^2 \right\}. \end{aligned} \quad (3.3)$$

Next, we analyze the convexity of each term of the last equation in (3.3). First, $-\frac{1}{n} \theta^T y$ is a linear function with respect to θ , thus it is inherently convex. Second, $\lambda \|D\theta\|_1$ represents the ℓ_1 norm form of $D\theta$, which is also known to be convex. Moving on to the last term, since the expression within the curly braces is affine and therefore convex with respect to θ , and considering that the point-wise maximum of a set of convex functions remains convex, it can be concluded that this particular term is also convex when considered in relation to θ . Consequently, the overall convexity of $f_B^{GME}(\theta)$ solely relies on the first term $\frac{1}{2n} \theta^T (I - \lambda(BD)^T (BD)) \theta$. If $D^T B^T B D \leq (1/\lambda)I$ holds true, then $I - \lambda(BD)^T (BD)$ becomes positive semidefinite and as a result, $f_B^{GME}(\theta)$ exhibits convexity. Conversely, if $D^T B^T B D < (1/\lambda)I$, then $I - \lambda(BD)^T (BD)$ turns out to be positive definite leading to strong convexity being displayed by $f_B^{GME}(\theta)$.

The convexity conditions mentioned above can be maintained by constructing the matrix parameter B . Following a similar approach as in [30], one simple way to determine the value of parameter B is

$$BD = \frac{\zeta}{\sqrt{\lambda}}I, \quad 0 \leq \zeta \leq 1. \quad (3.4)$$

When $0 < \zeta < 1$, then $D^T B^T B D = (\zeta/\lambda)I$ satisfies the convexity condition. When $\zeta = 0$, we have $B = 0$ and the GME-TV regularization $\Psi_B^{GME}(D\theta)$ reduce to the TV regularization $\|D\theta\|_1$. When $\zeta = 1$, (3.2) is satisfied with equality, resulting in a "maximally" non-convex regularization.

According to Theorem 3.1, we can deduce the following corollary, which is useful in our proof of algorithm convergence.

Corollary 3.1. *For all $\mu \geq \lambda\sigma_{\max}I$, $\lambda\Psi_B^{GME}(D\theta) + \frac{\mu}{2}\|D\theta\|_2^2$ is convex, where the σ_{\max} is the largest eigenvalue of $B^T B$.*

Corollary 3.1 can be derived by replacing $\frac{\mu}{2}\|\theta\|_2^2$ with $\frac{1}{2n}\|y - \theta\|_2^2$ in (3.3), following the proof process of Theorem 3.1.

3.2. Performance analysis

In this subsection, we analyze the performance of the GME-TV estimator (3.1), and use the oracle inequality to get an upper bound on the error rates, which is better than the upper bound of the standard TV estimator (1.2).

We consider a general GME-TV estimator (3.1), that is, the GME-TV regularization term is non-convex and does not necessarily satisfy the convexity condition given by Theorem 3.1. Due to the non-convexity, it is possible that the global minimum of the proposed GME-TV estimator cannot be achieved. Therefore, it is crucial to comprehensively understand the statistical performance of any first-order stationary points of the GME-TV estimator, making it more desirable in practical applications. We define $\hat{\theta} \in R^p$ as a stationary point of $F_B^{GME}(\theta)$ when it fulfills

$$0 \in \nabla_{\theta} f_B^{GME}(\theta)|_{\theta=\hat{\theta}}. \quad (3.5)$$

The compatibility factor, as defined by Definition 2.1, allows us to derive the subsequent oracle inequality that is relevant to the stationary points of the GME-TV estimator (3.1).

Theorem 3.2. *(Sharp oracle inequality for the GME-TV estimator) Let D be the first-order difference matrix, and σ_{\max} and σ_{\min} be the maximum and minimum eigenvalues of $B^T B$, respectively. The regularization parameter λ is set as $\lambda = \frac{1}{n}\sigma_{\min} \sqrt{2 \log\left(\frac{em}{\delta}\right)}$. For the fixed $\delta \in (0, 1)$, $0 < \mu < \frac{1}{n\|D\|_2^2}$ and $T \subset [m]$, the GME-TV estimator $\hat{\theta}$ defined in (3.1) satisfies*

$$\begin{aligned} \frac{1}{n}\|\hat{\theta} - \theta^*\|_2^2 \leq & \inf_{\theta} \left\{ \frac{1}{n}\|\bar{\theta} - \theta^*\|_2^2 + 4\lambda\Psi_{\sqrt{\sigma_{\max}I}}^{GME}((D\bar{\theta})_{T^c}) + 2\lambda\Psi_{\sqrt{\sigma_{\min}I}}^{GME}(D\bar{\theta}) - 2\lambda\Psi_{\sqrt{\sigma_{\max}I}}^{GME}(D\bar{\theta}) \right\} \\ & + \frac{16\sigma^2}{n(1 - n\mu\|D\|_2^2)} \left(\frac{|T|\varrho^2}{\kappa_T^2} \log\left(\frac{em}{\delta}\right) + \log\left(\frac{e}{\delta}\right) \right). \end{aligned} \quad (3.6)$$

Proof. Since $\hat{\theta}$ is a stationary point of $f(\theta)$, according to the definition of a subgradient, we have

$$0 \in \nabla_{\theta} f(\theta)|_{\theta=\hat{\theta}} = \frac{1}{n}(\hat{\theta} - y) + \nabla_{\theta} \Psi_B^{GME}(D\theta)|_{\theta=\hat{\theta}}. \quad (3.7)$$

Furthermore, by the chain rule, we have

$$\nabla_{\theta} \Psi_B^{GME}(D\theta)|_{\theta=\hat{\theta}} = D^T \text{sign}(D\theta) + D^T B^T B \left(D\theta - \min_v \left\{ \|v\|_1 + \frac{1}{2} \|B(D\theta - v)\|_2^2 \right\} \right). \quad (3.8)$$

Then by (3.8), there exist $z \in \left\{ \text{sign}(D\theta) + B^T B \left(D\theta - \min_v \left\{ \|v\|_1 + \frac{1}{2} \|B(D\theta - v)\|_2^2 \right\} \right) \right\}$ such that

$$\frac{1}{n}(y - \hat{\theta}) = \lambda D^T z. \quad (3.9)$$

In particular, for $\forall \bar{\theta} \in R^n$, we have

$$\frac{1}{n} \bar{\theta}^T (y - \hat{\theta}) = \lambda \bar{\theta} D^T z = \lambda (D\bar{\theta})^T z. \quad (3.10)$$

Moreover, for a specific $\hat{\theta}$, we have

$$\frac{1}{n} \hat{\theta}^T (y - \hat{\theta}) = \lambda (D\hat{\theta})^T z. \quad (3.11)$$

By subtracting (3.11) from (3.10) and in accordance with the definition of a subgradient, we obtain

$$\frac{1}{n} (y - \hat{\theta})(\bar{\theta} - \hat{\theta})^T = \lambda z (D\bar{\theta} - D\hat{\theta})^T \leq \Psi_B^{GME}(D\bar{\theta}) - \Psi_B^{GME}(D\hat{\theta}). \quad (3.12)$$

Considering the model $y = \theta^* + \varepsilon$ and utilizing the polarization equality, i.e., $2x^T y = \|x\|_2^2 + \|y\|_2^2 - \|x - y\|_2^2$, we can reformulate the left-hand side of (3.12) as

$$\begin{aligned} \frac{1}{n} (\theta^* - \hat{\theta} + \varepsilon)(\bar{\theta} - \hat{\theta})^T &= \frac{1}{n} [(\bar{\theta} - \hat{\theta})^T (\theta^* - \hat{\theta}) + \varepsilon^T (\bar{\theta} - \hat{\theta})] \\ &= \frac{1}{2n} \|\bar{\theta} - \hat{\theta}\|_2^2 + \frac{1}{2n} \|\theta^* - \hat{\theta}\|_2^2 - \frac{1}{2n} \|\bar{\theta} - \theta^*\|_2^2 + \frac{1}{n} \varepsilon^T (\bar{\theta} - \hat{\theta}). \end{aligned} \quad (3.13)$$

For $\forall \theta \in R^n$, the combination of (3.12) and (3.13) yields

$$\frac{1}{2n} \|\hat{\theta} - \bar{\theta}\|_2^2 + \frac{1}{2n} \|\hat{\theta}^* - \hat{\theta}\|_2^2 \leq \frac{1}{2n} \|\bar{\theta} - \theta^*\|_2^2 + \frac{2}{n} \varepsilon^T (\bar{\theta} - \hat{\theta}) + \lambda \Psi_B^{GME}(D\bar{\theta}) - \lambda \Psi_B^{GME}(D\hat{\theta}). \quad (3.14)$$

Next, we will analyze each term on the right-hand side of (3.14) separately.

We first address $\varepsilon^T (\bar{\theta} - \hat{\theta})$. Let Π and $I - \Pi$ denote the projection matrices onto $\ker(D)$ and $\ker(D)^\perp$, respectively. Moreover, we have $D^\dagger D = (I - \Pi)$. With the decomposition $\varepsilon = (\Pi\varepsilon)^T + ((I - \Pi)\varepsilon)^T$, we have

$$\begin{aligned} \varepsilon^T (\hat{\theta} - \bar{\theta}) &= (\Pi\varepsilon)^T (\hat{\theta} - \bar{\theta}) + ((I - \Pi)\varepsilon)^T (\hat{\theta} - \bar{\theta}) \\ &= (\Pi\varepsilon)^T (\hat{\theta} - \bar{\theta}) + \varepsilon^T D^\dagger D (\hat{\theta} - \bar{\theta}) \end{aligned}$$

$$\leq \|\Pi\varepsilon\|_2 \|\hat{\theta} - \bar{\theta}\|_2 + \|(D^\dagger)^T \varepsilon\|_\infty \|D(\hat{\theta} - \bar{\theta})\|_1. \quad (3.15)$$

The last inequality in (3.15) is derived from Hölder's inequality, i.e., $x^T y \leq \|x\|_p + \|y\|_q$ with $1/p + 1/q = 1$.

By using the maximal inequality for Gaussian random variables [42], we have

$$\|\Pi\varepsilon\|_2 \leq 2\sigma \sqrt{2 \log(e/\delta)}, \quad (3.16)$$

and

$$\|(D^\dagger)^T \varepsilon\|_\infty \leq \sigma \varrho \sqrt{2 \log(em/\delta)} = \lambda n, \quad (3.17)$$

with a probability of at least $1 - 2\delta$.

The substitution of (3.17) into (3.15) yields

$$\frac{2}{n} \varepsilon^T (\hat{\theta} - \bar{\theta}) \leq 2\lambda \|D(\hat{\theta} - \bar{\theta})\|_1 + \frac{2}{n} \|\Pi\varepsilon\|_2 \|\hat{\theta} - \bar{\theta}\|_2. \quad (3.18)$$

By using $\lambda \|D\theta\|_1 \leq \lambda \Psi_{\sqrt{\sigma_{\max} I}}^{GME}(D\theta) + \frac{\mu}{2} \|D\theta\|_2^2$ and $\|D(\hat{\theta} - \theta)\|_2^2 \leq \|D\|_2^2 \|\hat{\theta} - \theta\|_2^2$, we have

$$\frac{2}{n} \varepsilon^T (\hat{\theta} - \bar{\theta}) \leq 2\lambda \Psi_{\sqrt{\sigma_{\max} I}}^{GME}(D(\hat{\theta} - \bar{\theta})) + \mu \|D(\hat{\theta} - \bar{\theta})\|_2^2 + \frac{2}{n} \|\Pi\varepsilon\|_2 \|\hat{\theta} - \bar{\theta}\|_2. \quad (3.19)$$

Substituting (3.19) into (3.14), we obtain

$$\begin{aligned} \frac{1}{n} \|\hat{\theta} - \bar{\theta}\|_2^2 + \frac{1}{n} \|\hat{\theta}^* - \hat{\theta}\|_2^2 &\leq \frac{1}{n} \|\bar{\theta} - \theta^*\|_2^2 + \frac{2}{n} \|\Pi\varepsilon\|_2 \|\hat{\theta} - \bar{\theta}\|_2 + \mu \|D\|_2^2 \|\hat{\theta} - \bar{\theta}\|_2^2 \\ &\quad + 2\lambda \Psi_{\sqrt{\sigma_{\max} I}}^{GME}(D(\hat{\theta} - \bar{\theta})) + 2\lambda \Psi_B^{GME}(D\bar{\theta}) - 2\lambda \Psi_B^{GME}(D\hat{\theta}). \end{aligned} \quad (3.20)$$

For the GME-TV regularization $\Psi_B^{GME}(D\bar{\theta})$, we have

$$\Psi_B^{GME}(D\bar{\theta}) = \|D\bar{\theta}\|_1 - \min_v \left\{ \|v\|_1 + \frac{1}{2} \|B(D\bar{\theta} - v)\|_2^2 \right\} \leq \Psi_{\sqrt{\sigma_{\min} I}}^{GME}(D\bar{\theta}), \quad (3.21)$$

and

$$\Psi_B^{GME}(D\hat{\theta}) = \|D\hat{\theta}\|_1 - \min_v \left\{ \|v\|_1 + \frac{1}{2} \|B(D\hat{\theta} - v)\|_2^2 \right\} \geq \Psi_{\sqrt{\sigma_{\max} I}}^{GME}(D\hat{\theta}). \quad (3.22)$$

After substituting (3.21) and (3.22) into (3.20), we can rewrite (3.20) as

$$\begin{aligned} \frac{1}{n} \|\hat{\theta} - \bar{\theta}\|_2^2 + \frac{1}{n} \|\hat{\theta}^* - \hat{\theta}\|_2^2 &\leq \frac{1}{n} \|\bar{\theta} - \theta^*\|_2^2 + \frac{2}{n} \|\Pi\varepsilon\|_2 \|\hat{\theta} - \bar{\theta}\|_2 + \mu \|D\|_2^2 \|\hat{\theta} - \bar{\theta}\|_2^2 \\ &\quad + 2\lambda \Psi_{\sqrt{\sigma_{\max} I}}^{GME}(D(\hat{\theta} - \bar{\theta})) + 2\lambda \Psi_{\sqrt{\sigma_{\min} I}}^{GME}(D\bar{\theta}) - 2\lambda \Psi_{\sqrt{\sigma_{\max} I}}^{GME}(D\hat{\theta}) \\ &= \frac{1}{n} \|\bar{\theta} - \theta^*\|_2^2 + \frac{2}{n} \|\Pi\varepsilon\|_2 \|\hat{\theta} - \bar{\theta}\|_2 + \mu \|D\|_2^2 \|\hat{\theta} - \bar{\theta}\|_2^2 \\ &\quad + 2\lambda \Psi_{\sqrt{\sigma_{\max} I}}^{GME}(D(\hat{\theta} - \bar{\theta})) + 2\lambda \Psi_{\sqrt{\sigma_{\max} I}}^{GME}(D\bar{\theta}) - 2\lambda \Psi_{\sqrt{\sigma_{\max} I}}^{GME}(D\hat{\theta}) \\ &\quad + 2\lambda \Psi_{\sqrt{\sigma_{\min} I}}^{GME}(D\bar{\theta}) - 2\lambda \Psi_{\sqrt{\sigma_{\max} I}}^{GME}(D\bar{\theta}). \end{aligned} \quad (3.23)$$

Note that for any T , $\Psi_B(D\theta) = \Psi_B((D\theta)_T) + \Psi_B((D\theta)_{T^c})$. Consequently, by utilizing the triangular inequality and subadditivity and symmetry of ϱ , we obtain

$$\begin{aligned} & \Psi_{\sqrt{\sigma_{\max}I}}^{GME}(D(\hat{\theta} - \bar{\theta})) + \Psi_{\sqrt{\sigma_{\max}I}}^{GME}(D\bar{\theta}) - \Psi_{\sqrt{\sigma_{\max}I}}^{GME}(D\hat{\theta}) \\ & \leq \Psi_{\sqrt{\sigma_{\max}I}}^{GME}((D(\hat{\theta} - \bar{\theta}))_T) + \Psi_{\sqrt{\sigma_{\max}I}}^{GME}((D\bar{\theta})_{T^c}) + \Psi_{\sqrt{\sigma_{\max}I}}^{GME}((D\hat{\theta})_{T^c}) \\ & \quad + \Psi_{\sqrt{\sigma_{\max}I}}^{GME}(D\bar{\theta}) - \Psi_{\sqrt{\sigma_{\max}I}}^{GME}((D\hat{\theta})_T) - \Psi_{\sqrt{\sigma_{\max}I}}^{GME}((D\hat{\theta})_{T^c}) \\ & = \Psi_{\sqrt{\sigma_{\max}I}}^{GME}((D(\hat{\theta} - \bar{\theta}))_T) + 2\Psi_{\sqrt{\sigma_{\max}I}}^{GME}((D\theta)_{T^c}) + \Psi_{\sqrt{\sigma_{\max}I}}^{GME}((D\theta)_T) - \Psi_{\sqrt{\sigma_{\max}I}}^{GME}((D\hat{\theta})_T) \\ & \leq 2\Psi_{\sqrt{\sigma_{\max}I}}^{GME}((D(\hat{\theta} - \bar{\theta}))_T) + 2\Psi_{\sqrt{\sigma_{\max}I}}^{GME}((D\bar{\theta})_{T^c}). \end{aligned} \quad (3.24)$$

For $\Psi_{\sqrt{\sigma_{\max}I}}^{GME}((D(\hat{\theta} - \bar{\theta}))_T)$ in (3.24), we have

$$\Psi_{\sqrt{\sigma_{\max}I}}^{GME}((D(\hat{\theta} - \bar{\theta}))_T) \leq \|D(\hat{\theta} - \bar{\theta})T\|_1 \leq \kappa_T^{-1} \sqrt{|T|} \|\hat{\theta} - \bar{\theta}\|_2. \quad (3.25)$$

The last inequality is based on the definition of a compatibility factor in Definition 2.1.

Next, through the integration of (3.23)–(3.25), we can derive

$$\begin{aligned} \frac{1}{n} \|\hat{\theta} - \bar{\theta}\|_2^2 + \frac{1}{n} \|\hat{\theta}^* - \hat{\theta}\|_2^2 & \leq \frac{1}{n} \|\bar{\theta} - \theta^*\|_2^2 + \frac{2}{n} \|\Pi\varepsilon\|_2 \|\hat{\theta} - \bar{\theta}\|_2 + \mu \|D\|_2^2 \|\hat{\theta} - \bar{\theta}\|_2^2 + 4\lambda \kappa_T^{-1} \sqrt{|T|} \|\hat{\theta} \\ & \quad - \bar{\theta}\|_2 + 4\lambda \Psi_{\sqrt{\sigma_{\max}I}}^{GME}((D\bar{\theta})_{T^c}) + 2\lambda \Psi_{\sqrt{\sigma_{\min}I}}^{GME}(D\bar{\theta}) - 2\lambda \Psi_{\sqrt{\sigma_{\max}I}}^{GME}(D\bar{\theta}). \end{aligned} \quad (3.26)$$

By utilizing Young's inequality, i.e., $2xy \leq \frac{x^2}{m} + my^2$ for $m > 0$, with $x = \frac{1}{n} \|\Pi\varepsilon\|_2 + 2\kappa_T^{-1} \sqrt{|T|}$, $y = \|\hat{\theta} - \bar{\theta}\|_2$, and $m = \frac{1}{n} - \mu \|D\|_2^2$ (with the requirement of $\mu < \frac{1}{n\|D\|_2^2}$), we can rewrite (3.26) as

$$\begin{aligned} 2\left(\frac{1}{n} \|\Pi\varepsilon\|_2 + 2\kappa_T^{-1} \sqrt{|T|}\right) \|\hat{\theta} - \bar{\theta}\|_2 & \leq \frac{1}{m} \left(\frac{1}{n} \|\Pi\varepsilon\|_2 + 2\lambda \kappa_T^{-1} \sqrt{|T|}\right)^2 + m \|\hat{\theta} - \bar{\theta}\|_2^2 \\ & \leq \frac{2}{\frac{1}{n} - \mu \|D\|_2^2} \left(\frac{1}{n^2} \|\Pi\varepsilon\|_2^2 + 4\lambda^2 \kappa_T^{-2} |T|\right). \end{aligned} \quad (3.27)$$

The substitution of (3.27) into (3.26) results in the following equation:

$$\begin{aligned} \frac{1}{n} \|\bar{\theta} - \hat{\theta}\|_2^2 + \frac{1}{n} \|\theta^* - \hat{\theta}\|_2^2 & \leq \frac{1}{n} \|\bar{\theta} - \theta^*\|_2^2 + \frac{1}{n} \|\hat{\theta} - \bar{\theta}\|_2^2 + 4\lambda \Psi_{\sqrt{\sigma_{\max}I}}^{GME}((D\bar{\theta})_{T^c}) \\ & \quad + \frac{2n}{1 - n\mu \|D\|_2^2} \left(\frac{1}{n^2} \|\Pi\varepsilon\|_2^2 + 4\lambda^2 \kappa_T^{-2} |T|\right) \\ & \quad + 2\lambda \Psi_{\sqrt{\sigma_{\min}I}}^{GME}(D\bar{\theta}) - 2\lambda \Psi_{\sqrt{\sigma_{\max}I}}^{GME}(D\bar{\theta}). \end{aligned} \quad (3.28)$$

The cancellation of $\frac{1}{n} \|\hat{\theta} - \bar{\theta}\|_2^2$ on both sides of (3.28) yields

$$\begin{aligned} \frac{1}{n} \|\hat{\theta} - \theta^*\|_2^2 & \leq \frac{1}{n} \|\bar{\theta} - \theta^*\|_2^2 + 4\lambda \Psi_{\sqrt{\sigma_{\max}I}}^{GME}((D\bar{\theta})_{T^c}) + 2\lambda \Psi_{\sqrt{\sigma_{\min}I}}^{GME}(D\bar{\theta}) - 2\lambda \Psi_{\sqrt{\sigma_{\max}I}}^{GME}(D\bar{\theta}) \\ & \quad + \frac{2n}{1 - n\mu \|D\|_2^2} \left(\frac{1}{n^2} \|\Pi\varepsilon\|_2^2 + 4\lambda^2 \kappa_T^{-2} |T|\right). \end{aligned} \quad (3.29)$$

By substituting $\|\Pi\varepsilon\|_2 \leq 2\sigma\sqrt{2\log e/\delta}$ and $\sigma_Q\sqrt{2\log em/\delta} = \lambda n$ into (3.29), we obtain

$$\begin{aligned} \frac{1}{n}\|\hat{\theta} - \theta^*\|_2^2 \leq \inf_{\theta} \left\{ \frac{1}{n}\|\bar{\theta} - \theta^*\|_2^2 + 4\lambda\Psi_{\sqrt{\sigma_{\max}I}}^{GME}((D\bar{\theta})_{T^c}) + 2\lambda\Psi_{\sqrt{\sigma_{\min}I}}^{GME}(D\bar{\theta}) - 2\lambda\Psi_{\sqrt{\sigma_{\max}I}}^{GME}(D\bar{\theta}) \right\} \\ + \frac{16\sigma^2}{n(1 - n\mu\|D\|_2^2)} \left(\frac{|T|\varrho^2}{\kappa_T^2} \log\left(\frac{em}{\delta}\right) + \log\left(\frac{e}{\delta}\right) \right). \end{aligned} \quad (3.30)$$

This completes the proof of Theorem 3.2.

The oracle inequality remains valid for any $\hat{\theta}$ that satisfies the first-order optimality condition, even in the case of non-convex regularization. This mild requirement imposed on $\hat{\theta}$ represents a significant divergence from previous findings, such as Theorem 2 of [27], which is suitable for global minimization but challenging to ensure when employing non-convex regularizations.

The error bound to the right-hand side of (3.6) can be optimized through careful choice of $\bar{\theta}$ and T . When we set $\bar{\theta} = \theta^*$, then

$$\begin{aligned} \frac{1}{n}\|\hat{\theta} - \theta^*\|_2^2 \leq \inf_{\theta} \left\{ 4\lambda\Psi_{\sqrt{\sigma_{\max}I}}^{GME}((D\theta^*)_{T^c}) + 2\lambda\Psi_{\sqrt{\sigma_{\min}I}}^{GME}(D\theta^*) - 2\lambda\Psi_{\sqrt{\sigma_{\max}I}}^{GME}(D\theta^*) \right\} \\ + \frac{16\sigma^2}{n(1 - n\mu\|D\|_2^2)} \left(\frac{|T|\varrho^2}{\kappa_T^2} \log\left(\frac{em}{\delta}\right) + \log\left(\frac{e}{\delta}\right) \right). \end{aligned} \quad (3.31)$$

When we set T as the empty set, i.e., $T = \emptyset$, then

$$\begin{aligned} \frac{1}{n}\|\hat{\theta} - \theta^*\|_2^2 \leq \inf_{\theta} \left\{ 4\lambda\Psi_{\sqrt{\sigma_{\max}I}}^{GME}(D\theta^*) + 2\lambda\Psi_{\sqrt{\sigma_{\min}I}}^{GME}(D\theta^*) - 2\lambda\Psi_{\sqrt{\sigma_{\max}I}}^{GME}(D\theta^*) \right\} \\ + \frac{16\sigma^2}{n(1 - n\mu\|D\|_2^2)} \log\left(\frac{e}{\delta}\right). \end{aligned} \quad (3.32)$$

The error bound for the GME-TV estimator is now compared with Theorem 2.1 in the special case of setting $T = \emptyset$, i.e., (3.32). First, due to the non-convex nature of $\Psi_B^{GME}(D\theta)$, it imposes particularly strict constraints when $D\theta^*$ contains large coefficients. Therefore, we have $\Psi_{\sqrt{\sigma_{\max}I}}^{GME}(D\theta^*) \ll \|D\theta^*\|_1$ and $\Psi_{\sqrt{\sigma_{\min}I}}^{GME}(D\theta^*) - \Psi_{\sqrt{\sigma_{\max}I}}^{GME}(D\theta^*) \ll \|D\theta^*\|_1$. Consequently, the first term in (3.32) is upper bounded by that of Theorem 2.1. Moreover, the second term in (3.32) includes $n(1 - n\mu\|D\|_2^2)$ in the denominator, which establishes it as an upper bound for the second term mentioned in Theorem 2.1. In summary, the error bound of GME-TV is the upper bound from Theorem 2.1.

On the other hand, the above error gap can be reduced by adjusting the matrix parameter B . When we adjust B such that $\Psi_B^{GME}(D\theta)$ trends to $\|D\theta\|_1$, the improvement in the first term of the bound can be erased.

In conclusion, the proposed GME-TV estimator ensures strong statistical guarantees for any stationary point, despite its non-convex nature.

3.3. ADMM algorithm

In this subsection, we solve the GME-TV denoising model (3.1) by using the ADMM algorithm and prove the convergence.

By introducing an auxiliary variable z , problem (3.1) can be rewritten as

$$\begin{aligned} \arg \min_{\theta} \frac{1}{2n} \|\theta - y\|_2^2 + \lambda \Psi_B^{GME}(z) \\ \text{s.t. } z = D\theta. \end{aligned} \quad (3.33)$$

Furthermore, its augmented Lagrangian form can be expressed as

$$L(\theta, z, u) = \frac{1}{2n} \|\theta - y\|_2^2 + \lambda \Psi_B^{GME}(z) + u^T (D\theta - z) + \frac{\rho}{2} \|z - D\theta\|_2^2, \quad (3.34)$$

where u represents the Lagrange multiplier and $\rho > 0$ serves as a penalty parameter.

According to ADMM iteration, solving the minimization problem (3.34) involves addressing three subproblems.

Step 1. Update θ^{k+1} while keeping z^k and u^k fixed.

$$\begin{aligned} \theta^{k+1} &= \arg \min_{\theta} \left\{ \frac{1}{2n} \|\theta - y\|_2^2 + (u^k)^T D\theta + \frac{\rho}{2} \|z - D\theta\|_2^2 \right\} \\ &= \arg \min_{\theta} \left\{ \frac{1}{2n} \|\theta - y\|_2^2 + \frac{\rho}{2} \|D\theta - (z^k - \frac{\rho}{2} u^k)\|_2^2 \right\} \\ &= (\frac{1}{n} I + \rho D^T D)^{-1} (\frac{1}{n} y + \rho D^T z^k - D^T z^k u^k). \end{aligned} \quad (3.35)$$

Step 2. Update z^{k+1} while keeping θ^{k+1} and u^k fixed.

$$\begin{aligned} z^{k+1} &= \arg \min_z \left\{ \lambda \Psi_B^{GME}(z) + (u^k)^T z + \frac{\rho}{2} \|z - D\theta^{k+1}\|_2^2 \right\} \\ &= \arg \min_z \left\{ \lambda \Psi_B^{GME}(z) + \frac{\rho}{2} \|z - D\theta^{k+1} + \frac{1}{\rho} u^k\|_2^2 \right\}. \end{aligned} \quad (3.36)$$

Equation (3.36) can be considered as the proximal operator of $\Psi_B^{GME}(z)$. Although this operator does not have an explicit expression, we obtain its iterative solution through the PGD iteration.

By substituting $\Psi_B^{GME}(z)$ into (3.36), we can reformulate equation (3.36) as

$$z^{k+1} = \arg \min_z \left\{ \lambda \|z\|_1 - \lambda \min_v \left\{ \|v\|_1 + \frac{1}{2} \|B(z - v)\|_2^2 \right\} + \frac{\rho}{2} \|z - D\theta^{k+1} + \frac{1}{\rho} u^k\|_2^2 \right\}. \quad (3.37)$$

Let $f_1(z) = \frac{\rho}{2} \|z - D\theta^{k+1} + \frac{1}{\rho} u^k\|_2^2 - \lambda \min_v \left\{ \|v\|_1 + \frac{1}{2} \|B(z - v)\|_2^2 \right\}$ and $f_2(z) = \lambda \|z\|_1$. The update of z^{k+1} can be obtained through the PGD algorithm as follows:

$$\begin{aligned} z^{k+1} &= \text{prox}_{\alpha f_2}(z^k - \alpha \nabla f_1(z^k)) \\ &= \text{prox}_{\alpha \lambda \|\cdot\|_1}(z^k - \alpha f_1(z^k)), \end{aligned} \quad (3.38)$$

where α is the step size. The main challenge of (3.38) is how to solve the gradient $\nabla f_1(z)$. According to Lemma 3 in [16], the last term of $f_1(z)$ is differentiable with respect to z . Furthermore, the gradient of $f(z)$ can be expressed as

$$\nabla f_1(z) = \rho(z - \theta^{k+1} + \frac{u^k}{\rho}) - \lambda \left(B^T B \left(z - \arg \min_v \left\{ \frac{1}{2} \|B(z - v)\|_2^2 + \|v\|_1 \right\} \right) \right). \quad (3.39)$$

Note that (3.39) contains an ℓ_1 regularization problem:

$$v^{k+1} = \arg \min_v \left\{ \frac{1}{2} \|B(z - v)\|_2^2 + \|v\|_1 \right\}, \quad (3.40)$$

which can be regarded as a proximal operator of the ℓ_1 norm, i.e.,

$$v^{k+1} = \text{prox}_{\|\cdot\|_1}(v^k + B(z - v^k)), \quad (3.41)$$

where $\text{prox}_{\|\cdot\|_1}(\cdot)$ can be simply solved by the soft-thresholding operator (2.10).

By substituting (3.41) and (3.39) into (3.38), we can succinctly summarize the update of z^{k+1} as follows:

$$\begin{cases} v^{k+1} = \text{prox}_{\|\cdot\|_1}\{v^k + B(z^k - v^k)\}, \\ w^{k+1} = (1 - \alpha\rho + \alpha\lambda B^T B)z^k + \alpha\rho D\theta^{k+1} - \alpha u^k - \alpha\lambda B^T Bv^{k+1}, \\ z^{k+1} = \text{prox}_{\alpha\lambda\|\cdot\|_1}(w^{k+1}). \end{cases} \quad (3.42)$$

Step 3. Update u^{k+1} .

$$u^{k+1} = u^k + \rho(D\theta^{k+1} - z^{k+1}). \quad (3.43)$$

Finally, the ADMM algorithm for GME-TV regularization can be obtained as Algorithm 1 through the rearrangement of Eqs (3.35), (3.42), and (3.43).

Algorithm 1. ADMM for the GME-TV denoising model

Require: $y, \theta^0, z^0, u^0, B, \lambda, \rho, \alpha$

Ensure: θ

- 1: while 'stopping criterion is not met' do
 - 2: $\theta^{k+1} = (\frac{1}{n}I + \rho D^T D)^{-1}(\frac{1}{n}y + \rho D^T z^k - D^T z^k u^k)$;
 - 3: $v^{k+1} = \text{prox}_{\|\cdot\|_1}\{v^k + B(z^k - v^k)\}$;
 - 4: $w^{k+1} = (1 - \alpha\rho + \alpha\lambda B^T B)z^k + \alpha\rho D\theta^{k+1} - \alpha u^k - \alpha\lambda B^T Bv^{k+1}$;
 - 5: $z^{k+1} = \text{prox}_{\alpha\lambda\|\cdot\|_1}(w^{k+1})$;
 - 6: $u^{k+1} = u^k + \rho(D\theta^{k+1} - z^{k+1})$.
 - 7: end while
-

The total computational complexity of Algorithm 1 mainly depends on the θ -subproblem, which involves the matrix inversion $(\frac{1}{n}I + \rho D^T D)^{-1}$ with a computation complexity of $\mathcal{O}(n^3)$ for $D \in \mathbb{R}^{n \times n}$ and $\theta \in \mathbb{R}^n$. Fortunately, under the periodic boundary conditions for θ , $D^T D$ can be represented as a block circulant matrix with circulant blocks. In this scenario, solving the θ -subproblem becomes feasible through one forward fast Fourier transform (FFT) and one inverse FFT, each requiring a cost of $\mathcal{O}(n \log n)$ [43]. Consequently, the total computational complexity of Algorithm 1 is significantly reduced to $\mathcal{O}(n \log n)$, representing a substantial improvement compared to its original complexity of $\mathcal{O}(n^3)$.

Algorithm 1 showcases the utilization of an ADMM update with the Lagrangian introduced in Eq (3.34). Additionally, we offer convergence assurance for Algorithm 1.

Theorem 3.3. *By selecting the appropriate parameter ρ , then the primal residual $r^k = \theta^k - z^k$ and the dual residual $s^{k+1} = \rho(z^{k+1} - z^k)$ of Algorithm 1 satisfy that $\lim_{k \rightarrow \infty} \|r^k\|_2^2 = 0$ and $\lim_{k \rightarrow \infty} \|s^k\|_2^2 = 0$.*

Proof. The proof follows the idea of proving Proposition 1 in [44] and Theorem 3 in [36].

Considering the augmented Lagrangian $L(\theta, z, u)$ with respect to z , while assuming the other variables are fixed, we can derive

$$\begin{aligned} L(\theta, z, u) &= \frac{1}{2n} \|\theta - y\|_2^2 + \lambda \Psi_B^{GME}(z) + u^T(z - D\theta) + \frac{\rho}{2} \|z - D\theta\|_2^2 \\ &= \frac{1}{2n} \|\theta - y\|_2^2 + \lambda \Psi_B^{GME}(z) - u^T z + \frac{\rho}{2} \|z\|_2^2 - \frac{\rho}{2} \|D\theta\|_2^2 - \rho u^T(D\theta). \end{aligned} \quad (3.44)$$

As stated in Corollary 3.1, there exists $\mu > 0$ such that $\lambda \Psi_B^{GME}(D\theta) + \frac{\mu}{2} \|D\theta\|_2^2$ is convex. Additionally, it should be noted that the term $\frac{1}{2n} \|\theta - y\|_2^2 - \frac{\rho}{2} \|D\theta\|_2^2 - \rho u^T(D\theta)$ is independent of z . Therefore, by choosing $\rho \geq \mu$, we can observe the convexity of $L(\theta, z, u)$ with respect to each variable θ , u , and z . The property enables us to effectively apply Theorem 5.1 in [45]. Consequently, Algorithm 1 converges toward the limit points θ^* , z^* , and u^* . This implies that for the dual residual, we have

$$\lim_{k \rightarrow \infty} \|s^k\|_2^2 = \|\rho D^T(z^* - z^*)\|_2^2 = 0. \quad (3.45)$$

For the primal residual, it is worth noting that the update step for u in Algorithm 1 implies that $\forall k, t \geq 0$,

$$u^{k+t} = u^k + \sum_{i=1}^t (D\theta^{k+i} - z^{k+i}). \quad (3.46)$$

For fixed t and as $k \rightarrow \infty$, we have

$$u^* = u^* + t(D\theta^* - z^*), \quad (3.47)$$

where $\forall t \geq 0$, and thus, $D\theta^* - z^* = 0$. Therefore, it follows that

$$\lim_{k \rightarrow \infty} \|r^k\|_2^2 = \|D\theta^* - z^*\|_2^2 = 0. \quad (3.48)$$

Theorem 3.3 establishes the convergence of Algorithm 1 toward achieving both primal and dual feasibility in the limit. It also demonstrates the equivalence between the augmented Lagrangian form (3.34) with fixed z and u , and the original form (3.33). Each iteration of Algorithm 1 yields a stationary point θ for (3.34) with fixed z and u , thereby implying that θ^* is also a stationary point for (3.33).

4. Numerical experiments

In this section, we will compare the denoising performance of the proposed GME-TV regularization with the standard TV regularization [46] and MC-TV regularization [9, 17] through 1D and 2D denoising experiments.

4.1. 1D denoising

In this experiment, we evaluate the performance of different regularizations for 1D signal denoising. Initially, a piecewise constant 1D signal with a length of $n = 180$ is generated, degraded by the AWGN with a noise level of $\sigma = 0.9$. Subsequently, GME-TV, MC-TV, and standard TV regularization are applied for the purpose of denoising. The denoising accuracy is quantified using the root mean squared error (RMSE), which is defined as

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{\theta}_i - \theta_i^*)^2}. \quad (4.1)$$

The regularization parameter λ and the non-convex control matrix parameter B play a crucial role in determining the denoising performance of GME-TV. We establish these two hyperparameters as follows: First, we plot the curve of RMSE with respect to λ , and subsequently select the optimal value of λ that exhibits the lowest RMSE value. Second, based on (3.4), the matrix parameter B is determined by ζ . Therefore, once λ is determined, we proceed to plot the curve of RMSE with respect to ζ and then choose the optimal value of ζ that exhibits the lowest RMSE value. The remaining hyperparameters α and ρ in Algorithm 1 are simply set as $\alpha = 1$, and $\rho = 1$, following the strategy proposed in [9]. The hyperparameters λ and ζ for MC-TV are set up in the same manner as for GME-TV. For TV, since there is no non-convex control parameter, we set λ in the same way as for GME-TV. The remaining hyperparameters of MC-TV and TV are set by default according to [9] and [46], respectively.

The changes in RMSE values of different denoising models with λ and ζ are illustrated in Figure 1. It can be observed that the denoising performance of the models varies with different hyperparameter values. Moreover, GME-TV demonstrates superior denoising performance compared to other models.

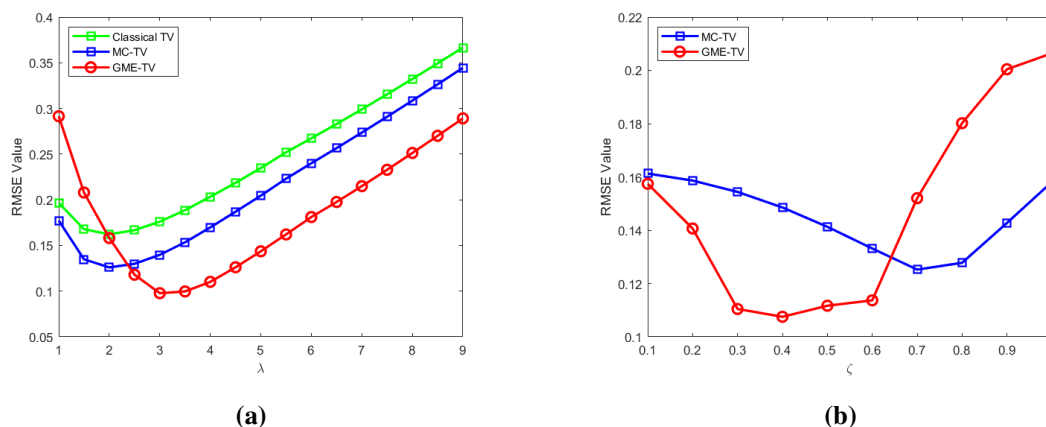


Figure 1. (a) RMSE as a function of the regularization parameter λ ; (b) RMSE as a function of the non-convex control parameter ζ .

We set the hyperparameters of different denoising models according to the results of Figure 1, and show the denoising results of different models in Figure 2. As depicted in Figure 2, TV regularization effectively mitigates noise. However, it tends to underestimate the magnitude of certain jump-type discontinuities and lacks the piecewise constant characteristics of the original noiseless signal. In contrast, GME-TV and MC-TV regularization provide a more precise estimation of segment

points in piecewise constant signals and demonstrate superior denoising performance compared to TV regularization. The RMSE values also substantiate the inherent superiority of GME-TV regularization.

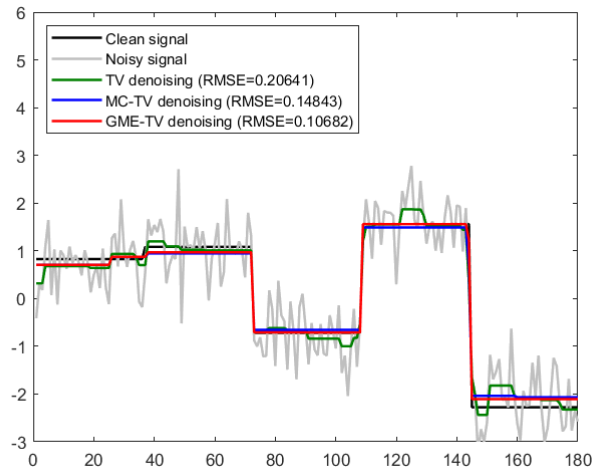


Figure 2. 1D signal denoising using different regularizations.

4.2. 2D denoising

In this experiment, we test the performance of different regularizations on image denoising. Four commonly used denoising test images, namely “Lena”, “Man”, “QRcode”, and “Geometric”, were utilized as shown in Figures 3(a), 4(a), 5(a), and 6(a). All images have a size of 512×512 with grayscale values ranging from 0 to 255. Subsequently, all images are degraded by additive white Gaussian noise with noise level $\sigma = 0.9$. The noisy images are shown in Figures 3(e), 4(e), 5(e), and 6(e).

The image denoising accuracy is quantified by the peak signal-to-noise ratio (PSNR), which is defined as

$$\text{PSNR} = 10 \times \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right), \quad (4.2)$$

where $\text{MAX} = 255$ is the maximum value of the images, and

$$\text{MSE} = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (\theta_{ij}^* - \hat{\theta}_{ij})^2, \quad (4.3)$$

where θ_{ij}^* and $\hat{\theta}_{ij}$ are the pixel values of the original image θ^* and the denoised image $\hat{\theta}$, respectively, with all images having a size of $M = N = 512$. The higher the PSNR value, the smaller the difference between the denoised image and the original image.

We employ the same strategy as the 1D denoising experiment to determine the hyperparameter values of different denoising models, and show the denoising results of each model in Figures 3–6. To enhance the visual contrast, in addition to enlarging a small patch of the local image, we also highlight the difference between the denoised image and the original one. As shown in Figures 3–6, all three

denoising models can remove noise well. Compared with TV denoising results, MC-TV and GME-TV not only enhance image texture and other intricate details but also sharpen edges more efficiently for a superior visual outcome. Moreover, the PSNR values further demonstrate the superior performance of GME-TV regularization, as evidenced by the improved PSNR values for all four images compared to TV regularization and MC-TV regularization.

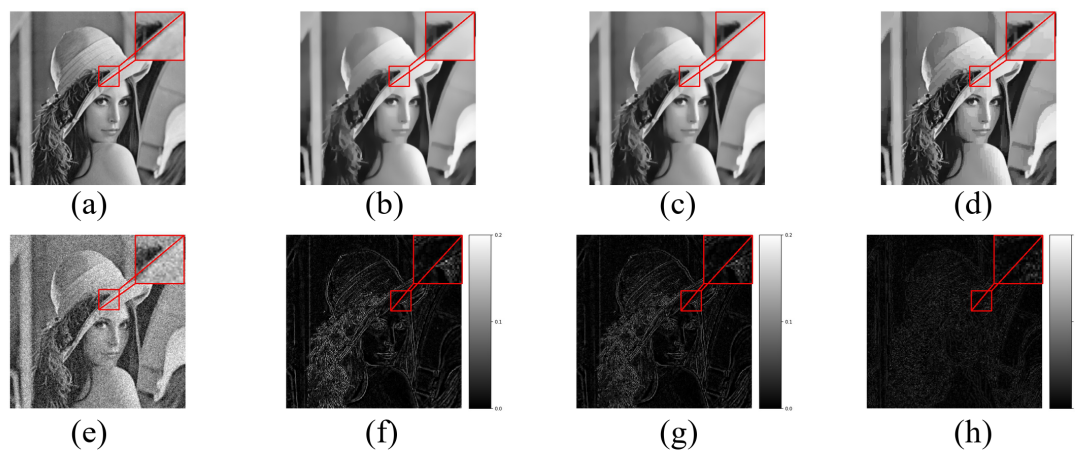


Figure 3. Image denoising using different regularizations. (a) Original image; (b) TV denoising result, PSNR = 32.0488 dB; (c) MC-TV denoising result, PSNR = 33.5046 dB; (d) GME-TV denoising result, PSNR = 36.4250 dB; (e) Noisy image; (f) Difference between (a) and (b); (g) Difference between (a) and (c); (h) Difference between (a) and (d).

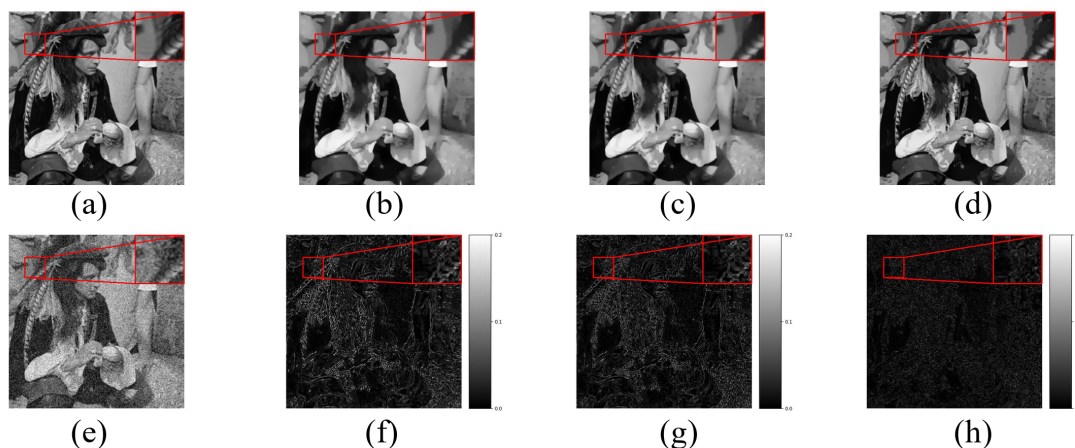


Figure 4. Image denoising using different regularizations. (a) Original image; (b) TV denoising result, PSNR = 32.0488 dB; (c) MC-TV denoising result, PSNR = 33.5046 dB; (d) GME-TV denoising result, PSNR = 36.4250 dB; (e) Noisy image; (f) Difference between (a) and (b); (g) Difference between (a) and (c); (h) Difference between (a) and (d).

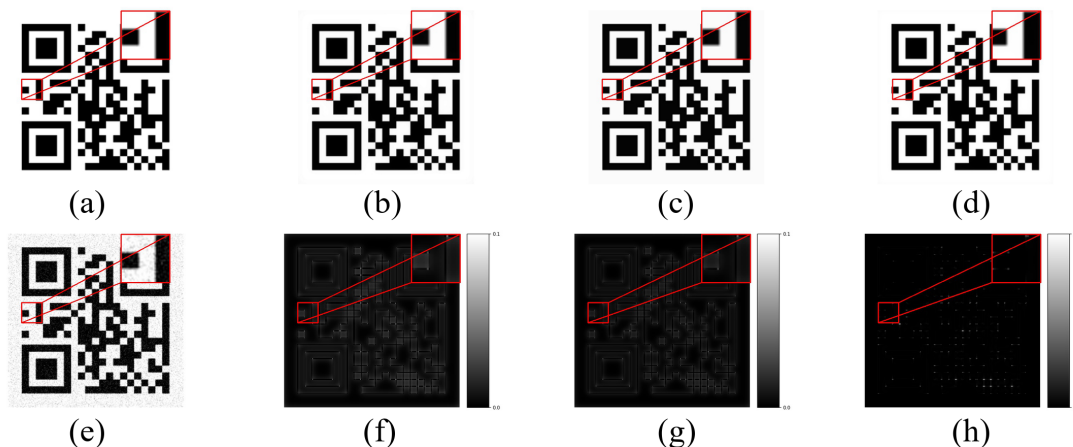


Figure 5. Image denoising using different regularizations. (a) Original image; (b) TV denoising result, PSNR = 41.5842 dB; (c) MC-TV denoising result, PSNR = 44.0425 dB; (d) GME-TV denoising result, PSNR = 54.2998 dB; (e) Noisy image; (f) Difference between (a) and (b); (g) Difference between (a) and (c); (h) Difference between (a) and (d).

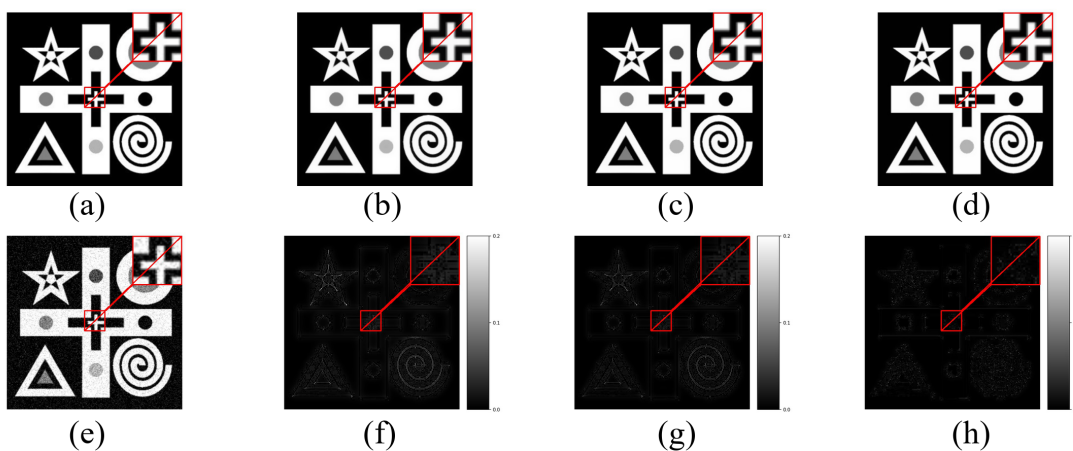


Figure 6. Image denoising using different regularizations. (a) Original image; (b) TV denoising result, PSNR = 41.1239 dB; (c) MC-TV denoising result, PSNR = 41.3220 dB; (d) GME-TV denoising result, PSNR = 43.1082 dB; (e) Noisy image; (f) Difference between (a) and (b); (g) Difference between (a) and (c); (h) Difference between (a) and (d).

5. Conclusions

In this paper, we propose CNC-TV regularization as a valuable alternative to TV regularization. This approach effectively addresses the issue of underestimation of signal discontinuity while ensuring the global convexity of the objective function. Empirical evidence is provided to demonstrate the

superior performance of CNC-TV regularization compared to TV regularization. Furthermore, our theoretical analysis establishes a foundation for a deeper understanding of the estimation properties associated with CNC-TV regularization. Additionally, we demonstrate that the CNC-TV regularization exhibits superior performance on both 1D and 2D denoising experiments.

In this work, we mainly focus on the theoretical analysis of CNC-TV. In fact, these findings suggest its potential significance in future applications such as image denoising, image reconstruction, seismic data processing, etc. Furthermore, there are also some challenges in the implementation of the CNC-TV model, such as the setting strategy of the non-convex control matrix parameter B , which will be further studied in our following work.

Author contributions

Yating Zhu: Conceptualization, Methodology, Software, Writing—original draft; Zixun Zeng: Software, Data curation; Zhong Chen: Funding acquisition, Supervision; Deqiang Zhou: Data curation, Formal analysis; Jian Zou: Project administration, Writing—review & editing, Supervision. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (grant number 62273060), and the Undergraduate Training Program of Yangtze University for Innovation and Entrepreneurship (Yz2023302).

Conflict of interest

All authors declare no conflicts of interest.

References

1. M. Elad, B. Kowar, G. Vaksman, Image denoising: the deep learning revolution and beyond—a survey paper, *SIAM J. Imaging Sci.*, **16** (2023), 1594–1654. <https://doi.org/10.1137/23M1545859>
2. L. I. 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](https://doi.org/10.1016/0167-2789(92)90242-F)
3. A. Chambolle, V. Caselles, D. Cremers, M. Novaga, T. Pock, An introduction to total variation for image analysis, In: *Theoretical foundations and numerical methods for sparse recovery*, Berlin, New York: De Gruyter, 2010, 263–340. <https://doi.org/10.1515/9783110226157.263>
4. M. Burger, S. Osher, A guide to the TV zoo, In: *Level set and PDE based reconstruction methods in imaging*, Cham: Springer, 2013, 1–70. https://doi.org/10.1007/978-3-319-01712-9_1
5. M. Pragliola, L. Calatroni, A. Lanza, F. Sgallari, On and beyond total variation regularization in imaging: the role of space variance, *SIAM Rev.*, **65** (2023), 601–685. <https://doi.org/10.48550/arXiv.2104.03650>
6. D. H. Zhao, R. Y. Huang, L. Feng, Proximity algorithms for the L^1L^2/TV^α image denoising model, *AIMS Math.*, **9** (2024), 16643–16665. <https://doi.org/10.3934/math.2024807>

7. A. Ben-loughfyry, A. Hakim, A bilevel optimization problem with deep learning based on fractional total variation for image denoising, *Multimed. Tools Appl.*, **83** (2024), 28595–28614. <https://doi.org/10.1007/s11042-023-16583-4>
8. A. Lanza, S. Morigi, F. Sgallari, Convex image denoising via non-convex regularization with parameter selection, *J. Math. Imaging Vis.*, **56** (2016), 195–220. <https://doi.org/10.1007/s10851-016-0655-7>
9. I. Selesnick, A. Lanza, S. Morigi, F. Sgallari, Non-convex total variation regularization for convex denoising of signals, *J. Math. Imaging Vis.*, **62** (2020), 825–841. <https://doi.org/10.1007/s10851-019-00937-5>
10. M. Kang, M. Jung, Nonconvex fractional order total variation based image denoising model under mixed stripe and Gaussian noise, *AIMS Math.*, **9** (2024), 21094–21124. <https://doi.org/10.3934/math.20241025>
11. I. W. Selesnick, A. Parekh, I. Bayram, Convex 1-D total variation denoising with non-convex regularization, *IEEE Signal Process. Lett.*, **22** (2015), 141–144. <https://doi.org/10.1109/LSP.2014.2349356>
12. J. Darbon, M. Sigelle, Image restoration with discrete constrained total variation part II: Levelable functions, convex priors and non-convex cases, *J. Math. Imaging Vis.*, **26** (2006), 277–291. <https://doi.org/10.1007/s10851-006-0644-3>
13. H. L. Zhang, L. M. Tang, Z. Fang, C. C. Xiang, C. Y. Li, Nonconvex and nonsmooth total generalized variation model for image restoration, *Signal Process.*, **143** (2018), 69–85. <https://doi.org/10.1016/j.sigpro.2017.08.021>
14. M. Hintermüller, T. Wu, Nonconvex TV^q -models in image restoration: analysis and a trust-region regularization–based superlinearly convergent solver, *SIAM J. Imaging Sci.*, **6** (2013), 1385–1415. <https://doi.org/10.1137/110854746>
15. Z. Fang, L. M. Tang, L. Wu, H. X. Liu, A nonconvex $TV_q - l_1$ regularization model and the ADMM based algorithm, *Sci. Rep.*, **12** (2022), 7942. <https://doi.org/10.1038/s41598-022-11938-7>
16. A. Lanza, S. Morigi, I. W. Selesnick, F. Sgallari, Sparsity-inducing nonconvex nonseparable regularization for convex image processing, *SIAM J. Imaging Sci.*, **12** (2019), 1099–1134. <https://doi.org/10.1137/18M1199149>
17. A. Lanza, S. Morigi, I. W. Selesnick, F. Sgallari, Convex non-convex variational models, In: *Handbook of mathematical models and algorithms in computer vision and imaging: mathematical imaging and vision*, Cham: Springer, 2023, 3–59. https://doi.org/10.1007/978-3-030-98661-2_61
18. Y. L. Liu, H. Q. Du, Z. X. Wang, W. B. Mei, Convex MR brain image reconstruction via non-convex total variation minimization, *Int. J. Imaging Syst. Technol.*, **28** (2018), 246–253. <https://doi.org/10.1002/ima.22275>
19. M. R. Shen, J. C. 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>
20. J. L. Li, Z. Y. Xie, G. Q. Liu, L. Yang, J. Zou, Diffusion optical tomography reconstruction based on convex-nonconvex graph total variation regularization, *Math. Methods Appl. Sci.*, **46** (2023), 4534–4545. <https://doi.org/10.1002/mma.8777>

21. Q. Li, A comprehensive survey of sparse regularization: fundamental, state-of-the-art methodologies and applications on fault diagnosis, *Expert Syst. Appl.*, **229** (2023), 120517.
22. H. B. Lin, F. T. Wu, G. L. He, Rolling bearing fault diagnosis using impulse feature enhancement and nonconvex regularization, *Mech. Syst. Signal Process.*, **142** (2020), 106790. <https://doi.org/10.1016/j.ymssp.2020.106790>
23. J. P. Wang, G. L. Xu, C. L. Li, Z. S. Wang, F. J. Yan, Surface defects detection using non-convex total variation regularized RPCA with kernelization, *IEEE Trans. Instrum. Meas.*, **70** (2021), 1–13. <https://doi.org/10.1109/TIM.2021.3056738>
24. T. H. Wen, Z. Chen, T. Zhang, J. Zou, Graph-based semi-supervised learning with non-convex graph total variation regularization, *Expert Syst. Appl.*, **255** (2024), 124709. <https://doi.org/10.1016/j.eswa.2024.124709>
25. G. Scrivanti, É. Chouzenoux, J. C. Pesquet, A CNC approach for directional total variation, In: *2022 30th European Signal Processing Conference (EUSIPCO)*, 488–492. <https://doi.org/10.23919/EUSIPCO55093.2022.9909763>
26. H. Q. Du, Y. L. Liu, Minmax-concave total variation denoising, *Signal Image Video Process.*, **12** (2018), 1027–1034. <https://doi.org/10.1007/s11760-018-1248-2>
27. J. C. Hütter, P. Rigollet, Optimal rates for total variation denoising, In: *Conference on Learning Theory*, **49** (2016), 1115–1146.
28. J. Q. Fan, R. Z. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, *J. Amer. Statist. Assoc.*, **96** (2001), 1348–1360. <https://doi.org/10.1198/016214501753382273>
29. A. H. Al-Shabli, Y. Feng, I. Selesnick, Sharpening sparse regularizers via smoothing, *IEEE Open J. Signal Process.*, **2** (2021), 396–409. <https://doi.org/10.1109/OJSP.2021.3104497>
30. I. Selesnick, Sparse regularization via convex analysis, *IEEE Trans. Signal Process.*, **65** (2017), 4481–4494. <https://doi.org/10.1109/TSP.2017.2711501>
31. E. Mammen, S. van de Geer, Locally adaptive regression splines, *Ann. Statist.*, **25** (1997), 387–413. <https://doi.org/10.1214/aos/1034276635>
32. D. Needell, R. Ward, Near-optimal compressed sensing guarantees for total variation minimization, *IEEE Trans. Image Process.*, **22** (2013), 3941–3949. <https://doi.org/10.1109/TIP.2013.2264681>
33. Y. X. Wang, J. Sharpnack, A. J. Smola, R. J. Tibshirani, Trend filtering on graphs, *J. Mach. Learn. Res.*, **17** (2016), 1–41.
34. V. Sadhanala, Y. X. Wang, R. J. Tibshirani, Total variation classes beyond 1d: minimax rates, and the limitations of linear smoothers, In: *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, 2016.
35. S. Chatterjee, S. Goswami, New risk bounds for 2D total variation denoising, *IEEE Trans. Inform. Theory*, **67** (2021), 4060–4091. <https://doi.org/10.1109/TIT.2021.3059657>
36. R. Varma, H. Lee, J. Kovačević, Y. Chi, Vector-valued graph trend filtering with non-convex penalties, *IEEE Trans. Signal Inform. Process. Netw.*, **6** (2020), 48–62. <https://doi.org/10.1109/TSIPN.2019.2957717>
37. A. Guntuboyina, D. Lieu, S. Chatterjee, B. Sen, Adaptive risk bounds in univariate total variation denoising and trend filtering, *Ann. Statist.*, **48** (2020), 205–229. <https://doi.org/10.1214/18-AOS1799>

38. F. Ortelli, S. van de Geer, Adaptive rates for total variation image denoising, *J. Mach. Learn. Res.*, **21** (2020), 1–38.
39. F. Ortelli, S. van de Geer, Prediction bounds for higher order total variation regularized least squares, *Ann. Statist.*, **49** (2021), 2755–2773. <https://doi.org/10.1214/21-AOS2054>
40. N. Parikh, S. Boyd, Proximal algorithms, *Found. Trends Optim.*, **1** (2014), 127–239. <https://doi.org/10.1561/24000000003>
41. 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. <http://dx.doi.org/10.1561/22000000016>
42. S. Boucheron, G. Lugosi, P. Massart, *Concentration inequalities: a nonasymptotic theory of independence*, Oxford University Press, 2013. <https://doi.org/10.1093/acprof:oso/9780199535255.001.0001>
43. N. Kumar, M. Sonkar, G. Bhatnagar, Efficient image restoration via non-convex total variation regularization and ADMM optimization, *Appl. Math. Model.*, **132** (2024), 428–453. <https://doi.org/10.1016/j.apm.2024.04.055>
44. S. J. Ma, J. Huang, A concave pairwise fusion approach to subgroup analysis, *J. Amer. Statist. Assoc.*, **112** (2017), 410–423. <https://doi.org/10.1080/01621459.2016.1148039>
45. P. Tseng, Convergence of a block coordinate descent method for nondifferentiable minimization, *J. Optim. Theory Appl.*, **109** (2001), 475–494. <https://doi.org/10.1023/A:1017501703105>
46. L. Condat, A direct algorithm for 1-D total variation denoising, *IEEE Signal Process. Lett.*, **20** (2013), 1054–1057. <https://doi.org/10.1109/LSP.2013.2278339>



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)