Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 20(4): 6932–6946.
DOI: 10.3934/mbe.2023299
Received: 05 December 2022
Revised: 19 January 2023
Accepted: 03 February 2023
Published: 08 February 2023

Research article

Global optimality analysis and solution of the ℓ_0 total variation signal denoising model

Shanshan Pan[1], **Qianqian Dai**[1] **and Huangyue Chen**[2,*]

[1] School of Science, Guangxi University of Science and Technology, Liuzhou 545006, China
[2] Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

* **Correspondence:** Email: hychen2022@amss.ac.cn.

Abstract: The total variation regularizer is diffusely emerged in statistics, image and signal processing to obtain piecewise constant estimator. The ℓ_0 total variation (L0TV) regularized signal denoising model is a nonconvex and discontinuous optimization problem, and it is very difficult to find its global optimal solution. In this paper, we present the global optimality analysis of L0TV signal denoising model, and design an efficient algorithm to pursuit its solution. Firstly, we equivalently rewrite the L0TV denoising model as a partial regularized (PL0R) minimization problem by aid of the structured difference operator. Subsequently, we define a P-stationary point of PL0R, and show that it is a global optimal solution. These theoretical results allow us to find the global optimal solution of the L0TV model. Therefore, an efficient Newton-type algorithm is proposed for the PL0R problem. The algorithm has a considerably low computational complexity in each iteration. Finally, experimental results demonstrate the excellent performance of our approach in comparison with several state-of-the-art methods.

Keywords: total variation; signal denoising; partial ℓ_0 regularized model; Global optimization; Newton-type algorithm

1. Introduction

The total variation (TV) regularization term has been used broadly in many areas, such as statistics [1–3], signal and image processing [4–7], in order to get piecewise constant estimator. The TV-based signal denoising methods are very effective for recovering piecewise constant (PWC) signals compared with conventional linear time-invariant filtering approach because the TV regularizer can capture the jump-sparsity in data. The classical TV (L1-TV) denoising model [4] based on ℓ_1 norm penalty function is formulated as a strongly convex optimization problem. For the observed signal

$y \in R^n$, the L1-TV denoising model is defined as

$$\min_{x \in R^n} \frac{1}{2}\|x - y\|_2^2 + \lambda \|Dx\|_1, \tag{1.1}$$

where $\lambda > 0$ is a regularization parameter, and $D \in R^{(n-1)\times n}$ is the first-order difference operator

$$D = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{pmatrix}.$$

In this case, the optimization problem (1.1) has unique global optimal solution, and it can be solved in finite time by using very fast direct (noniterative) algorithms [8, 9]. However, it has been shown in the literatures that nonconvex penalties can lead to more accurate estimation by comparison to ℓ_1 penalty, see, e.g., [10–13]. To enhance the performance of L1-TV denoising method, numerous nonconvex penalties function are used to substitute the ℓ_1 norm, see, e.g., [5, 14–19]. In view of the sparsity of the derivative of the underlying signal, the number of discontinuous points is a natural and cogent regularization term [17–19], namely,

$$\|Dx\|_0 = \#\{i : x_{i+1} - x_i \neq 0\}.$$

However, it is a nonconvex and discontinuous function. To escape the computational challenge arising from this regularizer, some continuous surrogates of $\|Dx\|_0$ are proposed in [5, 14–16]. In [5], the authors adopt the logarithmic penalty and arctangent penalty to substitute ℓ_1 norm in (1.1), and show that the corresponding objective functions are both convex if regularization parameter λ is less than a threshold. In [14], the authors propose the MCP-TV denoising method by using minimax concave penalty [12] to induce the sparsity of Dx. Based on the generalized Moreau envelope, Selesnick et al. [16] define the generalized Moreau envelope TV (GME-TV) penalty with matrix parameter B and use in denoising of signals. Note that the L1-TV, MCP-TV and Moreau enhanced TV [15] are the special cases of the GME-TV regularizer.

In this paper, we revisit the ℓ_0 TV regularized signal denoising model (L0TV), which is also called L2-Potts [17, 18]. The corresponding optimization model is

$$\min_{x \in R^n} \frac{1}{2}\|x - y\|_2^2 + \lambda \|Dx\|_0, \tag{1.2}$$

where $\|Dx\|_0$ is the ℓ_0 pseudo-norm of Dx, for counting the number of non-zero elements of Dx. The objective function of (1.2) is discontinuous and nonconvex. In [17, 18], the model (1.2) is solved by dynamic programming algorithm. To well resolve the challengeable problem, we first equivalently rewrite the L0TV as a partial ℓ_0 regularized (PL0R) optimization problem by resorting the interesting structure of D. For the PL0R problem, we define a P-stationary point and show that it is a global optimal solution. These fascinating theoretical results allow us to find a global optimal solution of the model (1.2). Although many existing numerical methods can be used to solve the ℓ_0 regularized (L0R) problem such as iterative hard-thresholding algorithm [20, 21], penalty decomposition [22], active set Barzilar-Borwein [23], these methods cannot be used to solve the model (1.2) and may suffer from low accuracy and slow convergence due to only make use of the first-order information of the involved functions. Very recently, an efficient Newton-type algorithm was proposed for L0R problem, and its global and quadratic convergence was established [24]. Inspired by this work, we design an effective algorithm based on the Newton's method to pursuit the global optimal solution of the L0TV model (NL0TV). The algorithm has a low computational complexity since a small-scale linear equation system is solved to update the Newton direction in per iteration. Comparing with several state-of-the-art methods, our experimental results show that the NL0TV achieves an excellent performance. The main contributions of this paper can be summarized as follows.

- An equivalent reformulation of the L0TV denoising model (1.2) is proposed in Theorem 1. After that, its necessary and sufficient conditions of global optimal solution are established in Theorem 2.
- We systematically analyze the relationships among the L0TV model (1.2), PL0R model (2.1) and the optimization problem (2.3) in terms of global optimal solution.
- Based on these theoretical results, an effective Newton algorithm is designed to pursuit the global optimal solution of the model (1.2).
- The excellent performance of our approach is illustrated by comparing with several state-of-the-art methods.

The remainder of this paper is organized as follows. Section 2 includes an equivalent reformulation of the L0TV denoising model, optimality conditions and Newton algorithm. Experiments and conclusions are presented in Sections 3 and 4, respectively.

2. Our method

The model (1.2) is a natural and cogent model for denoising of signals. But it is generally difficult to find the global optimal solution due to the intrinsic combinatorial property and inseparability of the regularization term. Next, we will give an equivalent reformulation of (1.2), which is helpful for designing effective optimization algorithm.

For convenience, we first definite several symbols. Let $z = Dx \in R^{n-1}$ and

$$H = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix} \in R^{n\times(n-1)}.$$

Based on these symbols, we give an equivalent reformulation of (1.2) in next theorem.

Theorem 1. *The L0TV signal denoising model (1.2) is equivalent to the following optimization problem*

$$\min_{w=(x_1;z)} \frac{1}{2}\|Gw - y\|_2^2 + \lambda\|z\|_0, \tag{2.1}$$

where $G = [e, H] \in R^{n\times n}$ and e is an n-dimensional column vector with each component is one.

Proof. It follows from $z = Dx \in R^{n-1}$ that $z_i = x_{i+1} - x_i$, for any $i = 1, 2, \cdots, n-1$, by means of the fascinating structure of D. Thus,

$$x_{i+1} = x_1 + \sum_{j=1}^{i} z_j, \text{ for any } i = 1, 2, \cdots, n-1.$$

Together with $w = (x_1; z)$, we have

$$x = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \ldots & 1 \end{pmatrix} \begin{bmatrix} x_1 \\ z \end{bmatrix} = Gw,$$

and $w = G^{-1}x$. Hence, the L0TV signal denoising model (1.2) can be equivalently written as (2.1) by substituting z and w into (1.2). □

The model (2.1) is named as partial ℓ_0 regularized (PL0R) optimization problem because the regularization term in (2.1) is with respect to z rather than w. Note that its the first term $f(w) = \frac{1}{2}\|Gw - y\|_2^2$ is a smooth and strongly convex function. Specifically, the gradient and Hessian of $f(w)$ are

$$\nabla f(w) = \begin{bmatrix} \nabla_{x_1} f(w) \\ \nabla_z f(w) \end{bmatrix} = \begin{bmatrix} e^\top (Gw - y) \\ H^\top (Gw - y) \end{bmatrix},$$

and $\nabla^2 f(w) = G^\top G$, respectively. Moreover, $\nabla^2 f(w)$ is a positive definite matrix due to the nonsingularity of G. Hence, there exist two positive constants L_f and l_f such that all eigenvalues of $\nabla^2 f(w)$ are greater than l_f and less than L_f.

2.1. Optimality conditions

Some first-order optimality conditions of ℓ_0 regularized problem have been established in Theorems 2.2 and 2.4 of [22]. But it is worth noting that they analyzed merely the relationship between Karush-Kuhn-Tucker (KKT) type stationary points and local minimizer. Inspired by the definition of the L-stationarity in [25, Definition 4.8], we introduce a P-stationary point of the PL0R optimization problem. For convenience, we first briefly review the definition of L-stationarity. Let $L > 0$. A vector x^* is called an L-stationary point of the ℓ_0 regularized optimization problem $\min_x g(x) + \lambda\|x\|_0$ if

$$x^* \in \mathrm{Prox}_{\frac{\lambda}{L}\|\cdot\|_0}\left(x^* - \frac{1}{L}\nabla g(x^*)\right).$$

Here $\mathrm{Prox}_{\frac{\lambda}{L}\|\cdot\|_0}(\cdot)$ is the proximal operator of $\frac{\lambda}{L}\|\cdot\|_0$, and defined as

$$\mathrm{Prox}_{\frac{\lambda}{L}\|\cdot\|_0}\left(x^* - \frac{1}{L}\nabla g(x^*)\right) := \arg\min_v \left\{\frac{\lambda}{L}\|v\|_0 + \frac{1}{2}\left\|v - \left(x^* - \frac{1}{L}\nabla g(x^*)\right)\right\|_2^2\right\}.$$

Next we present the definition of P-stationary point for the PL0R optimization problem (2.1).

Definition 1. *Let $\alpha > 0$. We say that $w^* = (x_1^*; z^*)$ is a P-stationary point with α of (2.1) if*

$$\begin{cases} 0 = e^\top(ex_1^* + Hz^* - y), \\ z^* \in Prox_{\alpha\lambda\|\cdot\|_0}(z^* - \alpha\nabla_z f(w^*)) . \end{cases} \tag{2.2}$$

It is worth mentioning that the proximal operator of $\alpha\lambda\|\cdot\|_0$ is multi-valued due to the non-convexity of $\alpha\lambda\|\cdot\|_0$. Fortunately, the analytic formula of $\text{Prox}_{\alpha\lambda\|\cdot\|_0}(\cdot)$ can be characterized by using the intrinsic discreteness and separability. Specifically, for any $i = 1, 2, \cdots, n-1$,

$$\text{Prox}_{\alpha\lambda\|\cdot\|_0}(u_i) = \arg\min_{v_i} \left\{ \mathcal{K}(v_i) := \alpha\lambda\mathcal{J}(v_i) + \frac{1}{2}(v_i - u_i)^2 \right\},$$

where $\mathcal{J} : R \to \{0, 1\}$ is the function defined by $\mathcal{J}(0) = 0$ and $\mathcal{J}(v_i) = 1$ for $v_i \neq 0$. It is easy to show that the minimum value of $\mathcal{K}(v_i)$ is the smaller one of $\alpha\lambda$ and $\frac{1}{2}u_i^2$. If $\alpha\lambda > \frac{1}{2}u_i^2$, then the minimizer of $\mathcal{K}(v_i)$ is 0. If $\alpha\lambda = \frac{1}{2}u_i^2$, then the minimizer of $\mathcal{K}(v_i)$ is 0 or u_i. If $\alpha\lambda < \frac{1}{2}u_i^2$, then the minimum value and minimizer of $\mathcal{K}(v_i)$ are $\alpha\lambda$ and u_i respectively. Hence, the proximal operator of $\alpha\lambda\|\cdot\|_0$ can be characterized as follows:

$$\text{Prox}_{\alpha\lambda\|\cdot\|_0}(u_i) = \begin{cases} u_i, & |u_i| > \sqrt{2\alpha\lambda}, \\ u_i \text{ or } 0, & |u_i| = \sqrt{2\alpha\lambda}, \\ 0, & |u_i| < \sqrt{2\alpha\lambda}. \end{cases} \quad \text{for any } i = 1, 2, \cdots, n-1.$$

Whereafter, the relationship between P-stationary point and global optimal solution of (2.1) is revealed.

Theorem 2. *For the PL0R optimization problem (2.1), the following results hold.*
(a) (Sufficiency) Let $(x_1^*; z^*)$ be a P-stationary point with $\alpha \geq 1/l_f$, then it is a global optimal solution.
(b) (Necessity) If $(x_1^*; z^*)$ is a global optimal solution, then it is a P-stationary point with $\alpha \in (0, 1/L_f)$.

Proof. (a) Since $(x_1^*; z^*)$ is a P-stationary point. It follows from Definition 1 that $0 = \nabla_{x_1} f(w^*)$ and

$$z^* \in \arg\min_z \left\{ \alpha\lambda\|z\|_0 + \frac{1}{2}\|z - (z^* - \alpha\nabla_z f(w^*))\|_2^2 \right\}.$$

According to the optimality of z^*, we have

$$\alpha\lambda\|z\|_0 + \frac{1}{2}\|z - (z^* - \alpha\nabla_z f(w^*))\|_2^2 \geq \alpha\lambda\|z^*\|_0 + \frac{1}{2}\|\alpha\nabla_z f(w^*)\|_2^2, \forall z.$$

After some simple manipulations, we can further obtain that

$$\langle z - z^*, \nabla_z f(w^*)\rangle + \frac{1}{2\alpha}\|z - z^*\|_2^2 + \lambda\|z\|_0 \geq \lambda\|z^*\|_0.$$

From the strong convexity of f, one can see, for any $w \in R^n$,

$$\begin{aligned} f(w) &\geq f(w^*) + \langle\nabla f(w^*), w - w^*\rangle + \frac{1}{2}l_f\|w - w^*\|_2^2 \\ &\geq f(w^*) + \langle\nabla_z f(w^*), z - z^*\rangle + \frac{1}{2}l_f\|z - z^*\|_2^2. \end{aligned}$$

Combining the above two aspects, we have

$$\begin{aligned} f(w) + \lambda\|z\|_0 &\geq f(w^*) + \frac{1}{2}(l_f - \frac{1}{\alpha})\|z - z^*\|_2^2 + \lambda\|z^*\|_0 \\ &\geq f(w^*) + \lambda\|z^*\|_0. \end{aligned}$$

Here, the last inequality of the above equation holds from $\alpha \geq 1/l_f$. Hence, $(x_1^*; z^*)$ is a global solution of (2.1).

(b) Suppose $w^* = (x_1^*; z^*)$ is a global optimal solution of (2.1). Then, from Fermat's rule [26, Theorem 10.1], we have

$$0 \in \nabla_z f(w^*) + \partial\lambda\|z^*\|_0 \text{ and } 0 = \nabla_{x_1} f(w^*),$$

which imply the first equation in (2.2) holds. On the other hand, from the strong smoothness of f, we obtain that

$$\begin{aligned} f(w) &\leq f(w^*) + \langle \nabla f(w^*), w - w^* \rangle + \frac{L_f}{2}\|w - w^*\|_2^2 \\ &= f(w^*) + \langle \nabla_z f(w^*), z - z^* \rangle + \frac{L_f}{2}\|w - w^*\|_2^2 \\ &\leq f(w^*) - \frac{1}{2\alpha}\|z - z^*\|_2^2 - \frac{\alpha}{2}\|\nabla_z f(w^*)\|_2^2 \\ &+ \frac{1}{2\alpha}\|z - z^* + \alpha\nabla_z f(w^*)\|_2^2 + \frac{L_f}{2}\|w - w^*\|_2^2. \end{aligned}$$

Let $x_1 = x_1^*$ and $z \in \Omega = \text{Prox}_{\alpha\lambda\|\cdot\|_0}(z^* - \alpha\nabla_z f(w^*))$. We know

$$\begin{aligned} f(w) + \lambda\|z\|_0 &\leq f(w^*) + \lambda\|z^*\|_0 + \frac{1}{2}\left(L_f - \frac{1}{\alpha}\right)\|z - z^*\|_2^2 \\ &\leq f(w) + \lambda\|z\|_0 + \frac{1}{2}\left(L_f - \frac{1}{\alpha}\right)\|z - z^*\|_2^2, \end{aligned}$$

where last inequality holds from the fact that $(x_1^*; z^*)$ is a global solution of (2.1). This together with $\alpha < 1/L_f$ leads to

$$0 \leq \left(L_f - \frac{1}{\alpha}\right)\|z - z^*\|_2^2 \leq 0,$$

yielding $z = z^*$. Therefore,

$$z^* \in \text{Prox}_{\alpha\lambda\|\cdot\|_0}(z^* - \alpha\nabla_z f(w^*)).$$

Namely, $(x_1^*; z^*)$ is a P-stationary point with $\alpha \in (0, 1/L_f)$. □

To solve the PL0R problem well, we first consider an optimization problem that is closely relevant to it, namely,

$$\min_{z \in R^{n-1}} h(z) + \lambda\|z\|_0. \tag{2.3}$$

Here, $h(z) = \frac{1}{2}(Hz - y)^\top M(Hz - y)$, $M = I_n - ee^\top/n$ and I_n is an identity matrix. Thus, the gradient and Hessian matrix of $h(z)$ are

$$\nabla h(z) = H^\top M(Hz - y) \text{ and } \nabla^2 h(z) = H^\top MH,$$

respectively. Moreover, $h(z)$ is a strongly convex function because its Hessian matrix is positive definite. Hence, there exists a constant $l_h > 0$ such that, for any z^1, z^2,

$$h(z^1) \geq h(z^2) + \langle \nabla h(z^2), z^1 - z^2 \rangle + \frac{1}{2} l_h \|z^1 - z^2\|_2^2.$$

Similarly, we say that z^* is a P-stationary point with $\alpha > 0$ of (2.3) if

$$\begin{aligned} z^* &\in \mathrm{Prox}_{\alpha\lambda\|\cdot\|_0}(z^* - \alpha \nabla h(z^*)) \\ &= \mathrm{Prox}_{\alpha\lambda\|\cdot\|_0}(z^* - \alpha H^\top M(Hz^* - y)). \end{aligned} \tag{2.4}$$

The following theoretical results present systematically the relationship between (2.1) and (2.3).

Theorem 3. *A point z^* is a P-stationary point with α of (2.3) if and only if*

$$v^* = (e^\top (y - Hz^*)/n; z^*)$$

is a P-stationary point with α of the PL0R model (2.1).

Proof. It follows that

$$\begin{aligned} \nabla_z f(v^*) &= H^\top (Gv^* - y) \\ &= H^\top (ee^\top (y - Hz^*)/n + Hz^* - y) \\ &= H^\top M(Hz^* - y) = \nabla h(z^*), \end{aligned}$$

and $\mathrm{Prox}_{\alpha\lambda\|\cdot\|_0}(z^* - \alpha \nabla_z f(v^*)) = \mathrm{Prox}_{\alpha\lambda\|\cdot\|_0}(z^* - \alpha \nabla h(z^*))$. Together with the equivalence of the first equation in (2.2) and $x_1^* = e^\top (y - Hz^*)/n$, the desired conclusions are proved. □

Theorem 4. *Let z^* be a global optimal solution of (2.3). Then it is a P-stationary point of (2.3) with $\alpha > 1/l_h$, and $v^* = (e^\top (y - Hz^*)/n, z^*)$ is also a global optimal solution of (2.1).*

Proof. Since the strong convexity of $h(z)$, we obtain that

$$\begin{aligned} h(z^*) &\geq h(z) + \langle \nabla h(z), z^* - z \rangle + \frac{1}{2} l_h \|z^* - z\|_2^2 \\ &= h(z) + \frac{1}{2\alpha} \|z^* - z + \alpha \nabla h(z)\|_2^2 + \frac{1}{2}\left(l_h - \frac{1}{\alpha}\right) \|z^* - z\|_2^2 - \frac{\alpha}{2} \|\nabla h(z)\|_2^2. \end{aligned}$$

Let z^* be a global optimal solution of (2.3) and z be a P-stationary point of (2.3). Then we have

$$\begin{aligned} & h(z^*) + \lambda \|z^*\|_0 \\ & \geq h(z) + \lambda \|z\|_0 + \frac{1}{2}\left(l_h - \frac{1}{\alpha}\right) \|z^* - z\|_2^2 \\ & \geq h(z^*) + \lambda \|z^*\|_0 + \frac{1}{2}\left(l_h - \frac{1}{\alpha}\right) \|z^* - z\|_2^2, \end{aligned}$$

where last inequality holds from the fact that z^* is a global optimal solution of (2.3). This together with $\alpha > 1/l_h$, we obtain that $z^* = z$. Therefore, z^* is a P-stationary point of (2.3). From Theorem 3 and Theorem 2 (a), we obtain that the desired conclusion immediately. □

Algorithm 1 NL0TV: Newton algorithm for the L0TV model (1.2)

Step 1: find z^* which is a P-stationary point of (2.3) by Newton algorithm;
Step 2: calculate $x_1^* = e^\top(y - Hz^*)/n$;
Step 3: calculate $x^* = ex_1^* + Hz^*$.

2.2. Newton algorithm

These optimality conditions establish a foundation for finding the global optimal solution of (1.2) effectively. Particularly, our algorithm can be divided into three steps, see Algorithm 1 for details.

After Step 1 and Step 2, we obtain $(x_1^*; z^*)$ which is a P-stationary point of (2.1) by Theorem 3. Further, it follows from Theorem 2 (a) that $(x_1^*; z^*)$ is a global optimal solution of the PL0R model (2.1). Therefore, we acquire the global optimal solution of the L0TV model (1.2) in Step 3. The main computational cost of NL0TV is to find a P-stationary point of (2.3). Inspired by [24], we adopt Newton-type algorithm to solve the model (2.3) and thereby obtaining its P-stationary point. To express the solution of (2.4) more explicitly, we introduce the following stationary equation

$$F(z, T) = \begin{bmatrix} \nabla_T h(z) \\ z_{\overline{T}} \end{bmatrix} = 0, \tag{2.5}$$

where $T = \left\{ i : |z_i - \alpha \nabla_i h(z)| \geq \sqrt{2\lambda\alpha} \right\}$ and $\overline{T}$ is the complementary set of T. The relationship between stationary equation (2.5) and a P-stationary point of (2.3) is established in next theorem.

Theorem 5. *Let z^* be a solution of (2.5). Then it is a P-stationary point of (2.3).*

Proof. Suppose that z^* is a solution of (2.5). Then $T^* = \left\{ i : |z_i^* - \alpha \nabla_i h(z^*)| \geq \sqrt{2\lambda\alpha} \right\}$ and

$$F(z^*, T^*) = \begin{bmatrix} \nabla_{T^*} h(z^*) \\ z^*_{\overline{T}^*} \end{bmatrix} = 0,$$

Moreover, $|z_i^*| \geq \sqrt{2\lambda\alpha}$ for any $i \in T^*$, and $|\nabla_i h(z^*)| < \sqrt{2\lambda/\alpha}$ for any $i \in \overline{T}^*$. Together with [20, Lemma 2], we obtain z^* is a P-stationary point of (2.3). □

To find a solution of (2.5), we first need to locate the index set T that is unknown in general and then solve the equation. Therefore, we employ an adaptively updating rule as follows. Let z^k be the k-th iteration point. We first calculate an approximation T_k, and then apply the Newton method to solve (2.5) with T_k. Namely, update d^k and z^{k+1} by, respectively,

$$\begin{cases} \nabla^2_{T_k, T_k} h(z^k) d^k_{T_k} = \nabla^2_{T_k, \overline{T}_k} h(z^k) z^k_{\overline{T}_k} - \nabla_{T_k} h(z^k) \\ \qquad\qquad d^k_{\overline{T}_k} = -z^k_{\overline{T}_k}, \end{cases} \tag{2.6}$$

and

$$z^{k+1} = z^k(t_k) = z^k + \begin{bmatrix} t_k d^k_{T_k} \\ d^k_{\overline{T}_k} \end{bmatrix} = \begin{bmatrix} z^k_{T_k} + t_k d^k_{T_k} \\ 0 \end{bmatrix}. \tag{2.7}$$

Here t_k is a step size generated by the Amijio line search as described in [24]. One can observe that the major computational cost of Newton algorithm arises from solving the equations (2.6). Its complexity

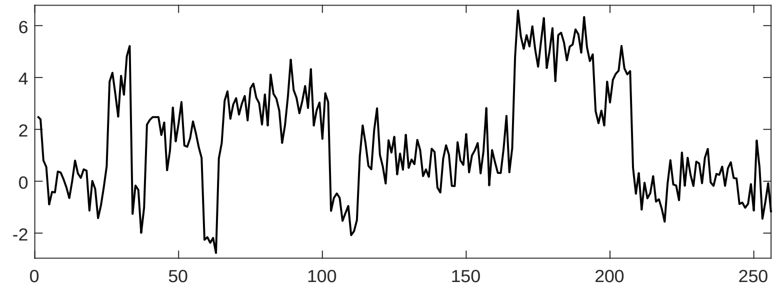


Figure 1. The blocks signal with noisy ($\sigma = 0.7$).

is approximately $O(|T_k|^3)$ in the k-th iteration. Note that, a sparse solution z^* is admitted, namely, $\|z^*\|_0 \ll n-1$, then $|T_k|$ can be quite small. Hence, NL0TV has a low computational complexity. As shown in [24], the generated sequence of Newton algorithm converges to a P-stationary point of (2.3) globally and quadratically. Thus, the global optimal solution of (2.1) can be obtained according to Theorem 4.

3. Experiments

This section conducts numerical experiments to illustrate the effectiveness of our proposed NL0TV method. All experiments are implemented by MATLAB (R2020a) on a personal laptop with 8 GB of RAM and Inter Core i7 2.3 GHz CPU.

To demonstrate the excellent performance of our approach, we compare NL0TV with five state-of-the-art TV-based denoising methods, including L1-TV [8], Atan-TV [5], MCP-TV [14], GME-TV [16] and L2-Potts [18]. For a fair comparison, the regularization parameters of all methods are traversed in $\{0.1, 0.2, \cdots, 5\}$ to output the best experimental results. The root-mean-square error (RMSE), mean-absolute-error (MAE) and signal-to-noise ratio (SNR) are used to quantify the performance of NL0TV and other compared methods. For comparison, we use the same test signal as described in the relevant literatures [5, 15, 16]. The true signal x^* is named 'blocks', which is generated by the function `MakeSignal` in the Wavelab software library (see `https://statweb.stanford.edu/~wavelab/` for details). The noisy signal y is obtained by adding Gaussian noise to the true signal, i.e., $y = x^* + \sigma N$, where N is the standard normal distribution, σ is the noise factor. Figure 1 illustrates an example of the 'blocks' signal with noise ($\sigma = 0.7$). Experimental results of six methods on the noisy signals with different σ are summarized in Figures 2-3 and Table 1.

Figure 2 reports the experimental results of six methods on the noisy signal with $\sigma = 0.7$. Some comments on Figure 2 can be made. (i) The RMSE of L1-TV, Atan-TV and MCP-TV methods are more than 0.3, while our approach achieves the minimum values 0.1460. (ii) Regarding MAE, NL0TV method obtains the minimum values compared with other five methods. Moreover, our approach also achieves the biggest SNR in comparison with other methods, which shows that the proposed method can remove more contaminants in the noisy signal. (iii) As presented in Figure 2, there are some small jumps in the signals estimated by L1-TV, Atan-TV, MCP-TV and GME-TV methods that are not present in the true signal, but this is rarely the case with NL0TV and L2-Potts methods. The reason may be that the ℓ_0 TV regularization term is more suitable for piecewise constant data than its surrogates. (iv) It can be seen from Figure 2 (e) and (f) that the NL0TV possesses superior denoising

Table 1. The experimental results of six methods on the noisy signal with $\sigma = 0.6$.

Method	RMSE Mean (SD)	MAE Mean(SD)	SNR Mean (SD)
L1-TV	0.2584 (0.0319)	0.1856 (0.0248)	19.6523 (**1.0803**)
Atan-TV	0.2227 (0.0330)	0.1588 (0.0247)	20.9736 (1.3033)
MCP-TV	0.2201 (0.0391)	0.1543 (0.0264)	21.1149 (1.5572)
GME-TV	0.1525 (0.0377)	0.1069 (0.0229)	24.4340 (2.1847)
L2-Potts	0.1632 (0.0457)	0.1007 (0.0258)	23.9373 (2.5716)
NL0TV	**0.1331** (**0.0248**)	**0.0967** (**0.0219**)	**25.4948** (1.6063)

performance comparing to the L2-Potts method, which indicates that our Newton algorithm can obtain more accurate solution than dynamic programming algorithm developed in [17, 18, 27]. Overall, the performance of NL0TV is better than other compared methods in terms of the RMSE, MAE and SNR.

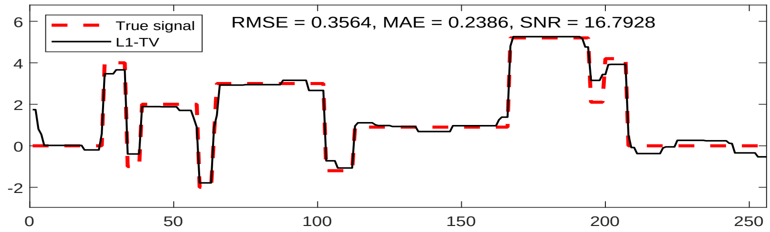
To further evaluate the stability of the denoising performance of our approach, we repeat 100 times noise realizations for each σ value and report the experimental results of six methods in Table 1 and Figure 3.

Table 1 presents the mean and standard deviation (SD) of RMSE, MAE and SNR for each method on the noisy signal with $\sigma = 0.6$. The values in bold represent the best performances. It is not difficult to see the following facts from this table. (i) The smallest values of average RMSE and average MAE are obtained by our approach. These experimental results imply that the proposed method possesses the best denoising performance. In particular, NL0TV method reduces the average RMSE by 48.49% and 18.44% compared with the L1-TV method and the L2-Potts method, respectively. (ii) The largest value of average SNR is achieved by the proposed approach. Specifically, NL0TV method increases the average SNR by 29.73% compared with the L1-TV method. Moreover, our approach increases the average SNR by 6.51% compared with the L2-Potts method, which indicates that the proposed Newton algorithm is effective. (iii) NL0TV method achieves the minimal standard deviation of RMSE and MAE. This phenomenon indicates that the proposed method is more stable compared with other several methods.

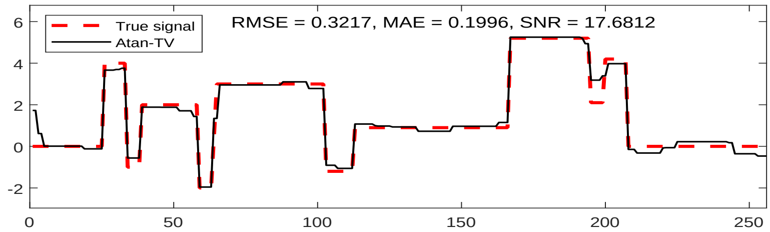
Figure 3 shows the experimental results of 100 replicates of each method on the noisy signal with $\sigma = 0.8$. This boxplot conveys several interesting phenomena. (i) In Figure 3 (a), the median (red line) of the RMSE of the NL0TV method is lower than that of the other five methods. (ii) In Figure 3 (a) and (b), the boxplot of our method does not have a red plus sign (+), which illustrates that our method does not produce abnormal experimental results. (iii) In Figure 3 (c), the median (red line) of the SNR of our method is higher than that of the other five methods. These phenomena show that the L0TV method is superior to other methods.

4. Conclusions

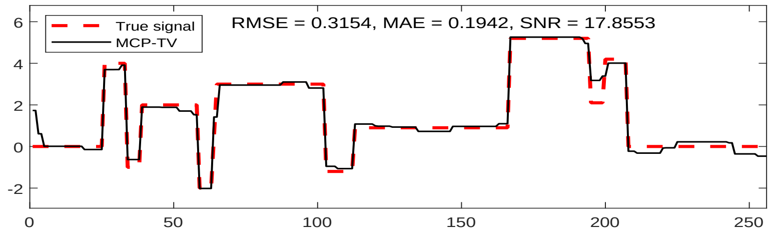
In this paper, we first give an equivalent reformulation of the L0TV model and analyze the relationship between its P-stationary point and global optimal solution. Based on these theoretical results, an efficient approach based on Newton's method is designed to find the global optimal solution of the L0TV model. The algorithm has a considerably low computational complexity and enjoys an excellent



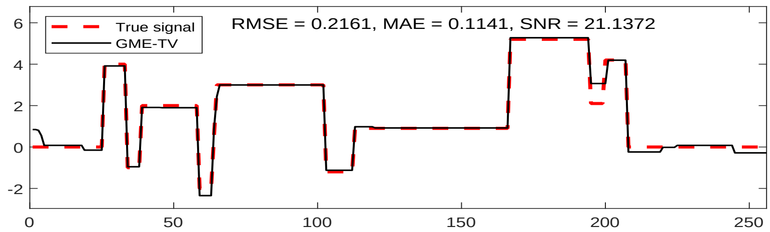
(a) L1-TV denoising



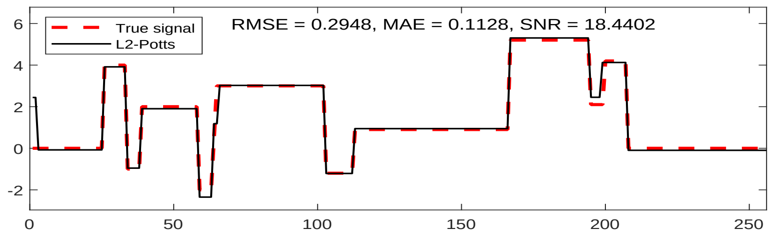
(b) Atan-TV denoising



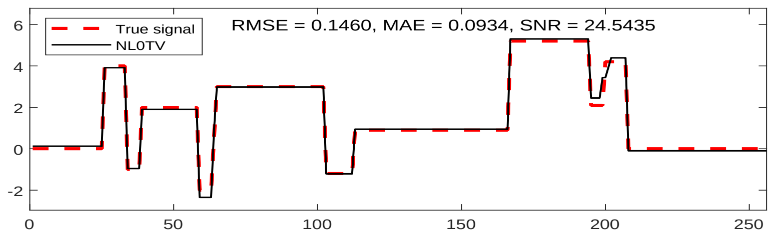
(c) MCP-TV denoising



(d) GME-TV denoising

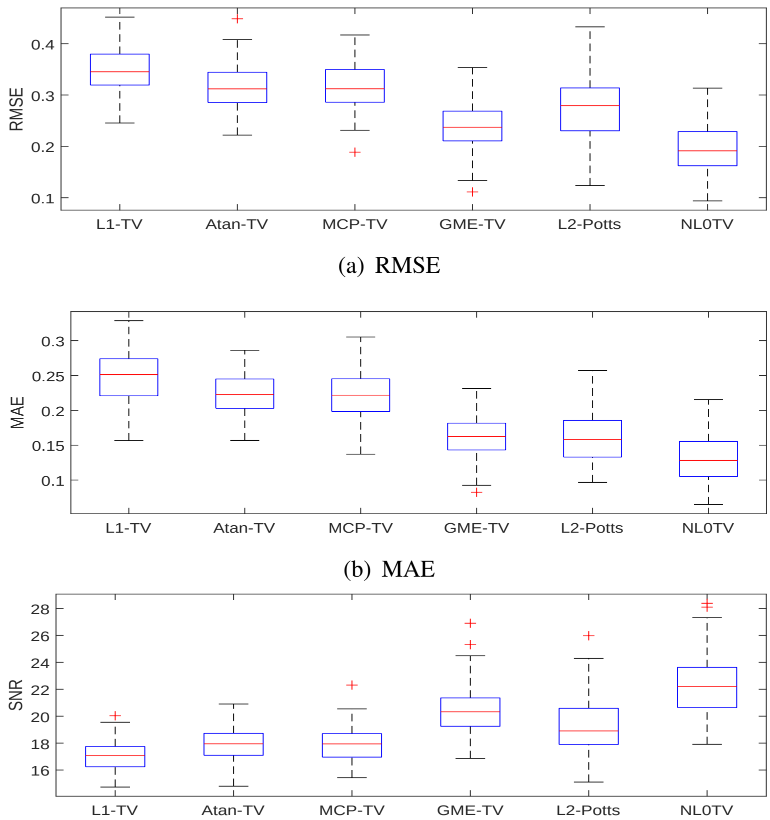


(e) L2-Potts denoising



(f) NL0TV denoising

Figure 2. The experimental results of six methods on the noisy signal with $\sigma = 0.7$.



(a) RMSE

(b) MAE

(c) SNR

Figure 3. The experimental results of six methods on the noisy signal with $\sigma = 0.8$.

performance when against five state-of-the-art methods. Furthermore, the techniques developed in this paper can be also used in conjunction with other TV-based methods in order to process more general signal and image processing problems.

Acknowledgment

We would like to thank the editors and anonymous referees for their insightful remarks and constructive comments that have helped to improve the quality of the paper. This work was supported in part by the Natural Science Foundation of Guangxi (2021GXNSFBA075012); in part by the Project funded by China Postdoctoral Science Foundation (2022M723327); in part by the Middle-aged and Young Teachers' Basic Ability Promotion Project of Guangxi(2020KY08005); and in part by the Doctoral Foundation of Guangxi University of Science and Technology (19Z40).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, K. Knight, Sparsity and smoothness via the fused lasso, *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, **67** (2005), 91–108. https://doi.org/10.1111/j.1467-9868.2005.00490.x
2. 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
3. B. Fang, A. Guntuboyina, B. Sen, Multivariate extensions of isotonic regression and total variation denoising via entire monotonicity and hardy-krause variation, *Ann. Statist.*, **49** (2021), 769–792. https://doi.org/10.1214/20-AOS1977
4. 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
5. 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
6. G. Yuan, B. Ghanem, ℓ_0 tv: A sparse optimization method for impulse noise image restoration, *IEEE Trans. Pattern Anal. Mach. Intell.*, **41** (2019), 352–364. https://doi.org/10.1109/TPAMI.2017.2783936
7. J. J. Liu, R. J. Ma, X. Y. Zeng, W. Q. Liu, An efficient non-convex total variation approach for image deblurring and denoising, *Appl. Math. Comput.*, **397** (2021), 125977. https://doi.org/10.1016/j.amc.2021.125977
8. 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

9. L. Dumbge, A. Kovac. Extensions of smoothing via taut strings, *Electron. J. Stat.*, **3** (2009), 41–75. https://doi.org/10.1214/08-EJS216

10. 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

11. Z. B. Xu, H. Zhang, Y. Wang, X. Y. Chang, Y. Liang, $\ell_{1/2}$ regularization, *Sci. China Inf. Sci.*, **53** (2010), 1159–1169. https://doi.org/10.1007/s11432-010-0090-0

12. C. H. Zhang, Nearly unbiased variable selection under minimax concave penalty. *Ann. Stat.*, **38** (2010), 894–945. https://doi.org/10.1214/09-AOS729

13. T. Zhang, Analysis of multi-stage convex relaxation for sparse regularization, *J. Mach. Learn. Res.*, **11** (20104), 1081–1107.

14. H. Q. Du, Y. L. Liu, Minmax-concave total variation denoising. *Signal, Image and Video Process.*, **12** (2018), 1027–1034. https://doi.org/10.1007/s11760-018-1248-2

15. I. W. Selesnick, Total variation denoising via the moreau envelope, *IEEE Signal Process. Lett.*, **24** (2017), 216–220. https://doi.org/10.1109/LSP.2017.2647948

16. I. W. Selesnick, A. Lanza, S. Morigi, F. Sgallari, Non-convex total variation regularization for convex denoising of signals, *J. Math. Imaging Vision*, **62** (2020), 825–841. https://doi.org/10.1007/s10851-019-00937-5

17. M. Storath, A. Weinmann, L. Demaret, Jump-sparse and sparse recovery using potts functionals. *IEEE Trans. Signal Process.*, **62** (2014), 3654–3666. https://doi.org/10.1109/TSP.2014.2329263

18. J. Frecon, N. Pustelnik, N. Dobigeon, H. Wendt, P. Abry, Bayesian selection for the l2-potts model regularization parameter: 1d piecewise constant signal denoising, *IEEE Trans. Signal Process.*, **65** (2017), 5215–5224. https://doi.org/10.1109/TSP.2017.2715000

19. R. B. Potts, Some generalized order-disorder transformations, *Math. Proc. Cambridge Philos. Soc.*, **48** (1952), 106–109. https://doi.org/10.1017/S0305004100027419

20. T. Blumensath, M. E. Davies, Iterative thresholding for sparse approximations, *J. Fourier Anal. Appl.*, **14** (2008), 629–654. https://doi.org/10.1007/s00041-008-9035-z

21. Z. S. Lu, Iterative hard thresholding methods for ℓ_0 regularized convex cone programming, *Math. Program.*, **147** (2014), 125–154. https://doi.org/10.1007/s10107-013-0714-4

22. Z. S. Lu, Y. Zhang, Sparse approximation via penalty decomposition methods, *SIAM J. Optim.*, **23** (2013), 2448–2478. https://doi.org/10.1137/100808071

23. W. Y. Cheng, Z. Chen, Q. J. Hu, An active set barzilar-borwein algorithm for ℓ_0 regularized optimization, *J. Global Optim.*, **76** (2020), 769–791. https://doi.org/10.1007/s10898-019-00830-w

24. S. L. Zhou, L. L. Pan, N. H. Xiu, Newton method for ℓ_0 regularized optimization, *Numer. Algor.*, **88** (2021), 1541–1570. https://doi.org/10.1007/s11075-021-01085-x

25. A. Beck, N. Hallak, Proximal mapping for symmetric penalty and sparsity, *SIAM J. Optim.*, **28** (2018), 496–527. https://doi.org/10.1137/17M1116544

26. R. T. Rockafellar, R. J-B. Wets, *Variational Analysis*, Springer, Berlin, 1997. https://doi.org/10.1007/978-3-642-02431-3

27. M. Storath, A. Weinmann, Fast partitioning of vector-valued images, *SIAM J. Imag. Sci.*, **7** (2014), 1826–1852. https://doi.org/10.1137/130950367