

Research Article

A General Proximal Alternating Minimization Method with Application to Nonconvex Nonsmooth 1D Total Variation Denoising

Xiaoya Zhang,¹ Tao Sun,¹ and Lizhi Cheng^{1,2}

¹College of Science, National University of Defense Technology, Changsha, Hunan 410073, China

²The State Key Laboratory for High Performance Computation, National University of Defense Technology, Changsha, Hunan 410073, China

Correspondence should be addressed to Xiaoya Zhang; suezhang246@163.com

Received 15 June 2016; Accepted 10 October 2016

Academic Editor: Maria L. Gandarias

Copyright © 2016 Xiaoya Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We deal with a class of problems whose objective functions are compositions of nonconvex nonsmooth functions, which has a wide range of applications in signal/image processing. We introduce a new auxiliary variable, and an efficient general proximal alternating minimization algorithm is proposed. This method solves a class of nonconvex nonsmooth problems through alternating minimization. We give a brilliant systematic analysis to guarantee the convergence of the algorithm. Simulation results and the comparison with two other existing algorithms for 1D total variation denoising validate the efficiency of the proposed approach. The algorithm does contribute to the analysis and applications of a wide class of nonconvex nonsmooth problems.

1. Introduction

In the past few years, increasing attentions have been paid to convex optimization problems, which consist of minimizing a sum of convex or smooth functions [1–3]. Each of the objective functions enjoys several appreciated properties, like strong convexity, Lipschitz continuity, or other convex conditions. These properties can usually lead to great advantage in computing. Meanwhile works on such convex problems have provided a sound theoretic foundation. Both of the theoretic and computing advantages created many benefits to practical use. It is particularly noticeable in signal/image processing, machine learning, computer vision, and so forth. However, what deserves the special attention is the fact that the convex or smooth models are always approximations of nonconvex nonsmooth problems. For example, nonconvex ℓ_0 -norm in sparse recovery problems is routinely relaxed as convex ℓ_1 -norm, and many related works were developed [4, 5]. Although the difference between the nonconvex nonsmooth problem and its approximations vanishes in certain case, it is nonnegligible sometimes, like the problem in paper [6]. On the other hand, excellent numerical performances of various

nonconvex nonsmooth algorithms inspire researchers to continue on their directions to the nonconvex methodology.

Nonconvex and nonsmooth convex optimization problems are ubiquitous in different disciplines, including signal denoising [7], image deconvolution [8, 9], or other ill-posed inverse problems [10], to name a few. In this paper, we aim at solving the following generic nonconvex nonsmooth optimization problem, formulated in real Hilbert spaces \mathcal{X} and $\{\mathcal{U}_m\}_{m=1,2,\dots,M}$, for some $M \in \mathbb{N}^+$:

$$\hat{x} \in \arg \min_x \left\{ P(x) := f(x) + \sum_{m=1}^M h_m(L_m x) \right\}, \quad (1)$$

where (i) $f(\cdot) : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ and $h_m(\cdot) : \mathcal{U}_m \rightarrow \mathbb{R} \cup \{+\infty\}$ are proper lower semicontinuous functions; (ii) the operators $L_m : \mathcal{X} \rightarrow \mathcal{U}_m$ are linear; and (iii) the set of minimizers is supposed to be nonempty.

It is quite meaningful to find a common convergent point in the optimal set of sums of simple functions [2, 11]. Insightful studies for nonconvex problems are presented in [12, 13]: if nonconvex structured functions of the type $L = f(x) + Q(x, y) + g(y)$ has the Kurdyka-Lojasiewicz

(K-L) property, then each bounded sequence generated by a proximal alternating minimization algorithm converges to a critical point of L . Our main work is mainly based on this convergence result and introduces a generic proximal minimization method for the general model (1). Note that if some of the functions in (1) are zeros and linear operator L_m is the identity I_d , the model would reduce to common one in compressed sensing [14]. The model we consider is the generalization of many application problems, such as the common “lasso” problem [15], and composition in papers [2, 16]. Here, we provide a few of the typical examples.

Example 1 (1D total variation minimization [7]). In this application, we need to solve the following denoising problem:

$$\hat{x} \in \arg \min_{x \in \mathcal{X}} \{ \|x - y\|_2^2 + \lambda \|Dx\|_0 \}, \quad (2)$$

where $y \in \mathcal{X}$ is the input signal, $\mathcal{X} = \mathbb{R}^n$, $\|z\|_0$ counts the number of nonzero elements of z , and $Dx := (x_2 - x_1, x_3 - x_2, \dots, x_n - x_{n-1})$ is defined as derivation of original signal x .

Noise removal is the basis and requisite of other sub-sequential applications and algorithm dealing with total variation (TV) regularizer; this regularizer is of great importance since it can efficiently deal with noisy signals which have sparse derivatives (or gradients), for instance, piecewise constant (PWC) signal that has flat sections with a number of abrupt jumps. The 1D total variation minimization can be extended to related 2-dimension image restoration.

Example 2 (group lasso [17]). In this application, one needs to solve

$$\hat{x} \in \arg \min_{x \in \mathbb{R}^N} \left\{ \frac{1}{2} \|Ax - y\|_2^2 + \lambda \sum_{i=1}^n \|x_i\|_2 \right\}, \quad (3)$$

where $x^T = (x_1^T, x_2^T, \dots, x_n^T)$, $x_i \in \mathbb{R}^{p_i}$ are the decision block variables and $\sum_{i=1}^n p_i = N$ with p_i is the corresponding block size.

In the past few years, researches on structural sparse signal recovery have been very popular and group lasso is typical one of those important problems. It attracts many attentions in face recognition, multiband signal processing, and other machine learning problems. The general case is also applied to many other kinds of structural sparse recovery problems, like l_{12} -minimization [18] and block sparse recovery [19].

Example 3 (image deconvolution [20]). In this application, one needs to solve

$$\hat{X} \in \arg \min_{X \in \mathcal{X}} \{ \|AX - Y\|_F^2 + \lambda \text{TV}(X) \}, \quad (4)$$

where $\mathcal{X} = \mathbb{R}^{m \times n}$ and $\|X\|_F = [\sum_i \sum_j (X_{i,j}^2)]^{1/2}$. The discrete total variation, denoted by TV in (4), is defined as follows. We define $r \times r$ matrix:

$$D_r := \begin{pmatrix} 1 & & & -1 \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix}, \quad (5)$$

and the discrete gradient operator $D : \mathcal{X} \rightarrow \mathcal{X}^2$ is defined as

$$DX_{i,j} = (dx_{i,j}, dy_{i,j}), \quad dx = D_m X, \quad dy = XD_n^T. \quad (6)$$

Then we have $\text{TV}(X) = \|DX\|_{1,2} = \sum_i \sum_j \sqrt{|dx_{i,j}|^2 + |dy_{i,j}|^2}$.

The concept of deconvolution finds lots of applications in signal processing and image processing [21–23]. In this paper, it would just be considered as a specific case to problem (1), although paying attention to this problem (4) with other details is equally important.

The main difficulty in solving (1) lies in that x is coupled by L_m . In order to surmount the computation barrier, we introduce a new auxiliary variable and split the problem into two sequences of subproblems, minutely described in the next section. Then our problem is an extension of problem given in paper [12]. The paper aims at giving a generic proximal alternating minimization method for a class of nonconvex nonsmooth problem (1), to be applied in many fields. The motivation is introduction of auxiliary variable and splitting the original problem into two kinds of easier nonconvex nonsmooth subproblems. Recent studies often give the regularization h_m a reasonable assumption; namely, the proximal map of h_m is easy to calculate. Then convergence analysis can be extended by the context of the present work [12]. In the last section, we show application to nonconvex nonsmooth 1D total variation signal denoising.

2. Algorithm

In order to reduce computation complexity caused by composite of nonconvex function $h_m(\cdot)$ and operator L_m , we introduce a sequence of auxiliary variables $\{\theta_m\}_{m=1}^M$. Then, the problem in (1) can be represented equivalently as follows: for each $m \in \{1, \dots, M\}$

$$(\hat{x}, \hat{\theta}_m) \in \arg \min_{x, \theta_m} \{ P_m(x, \theta_m) := f(x) + Q_m(x, \theta_m) + H(\theta) \}, \quad (7)$$

where $\{\theta_m\}_{m=1}^M$ are represented as a concise whole θ and

$$\begin{aligned} Q(x, \theta) &:= \sum_{m=1}^M \frac{\tau_m}{2} \|\theta_m - L_m x\|_2^2; \\ Q_m(x, \theta_m) &:= \frac{\tau_m}{2} \|\theta_m - L_m x\|_2^2; \\ H(\theta) &:= \sum_{m=1}^M h_m(\theta_m). \end{aligned} \quad (8)$$

The last term $\sum_{m=1}^M (\tau_m/2) \|\theta_m - L_m x\|_2^2$ ensures the high similarity between θ_m and $L_m x$ ($m = 1, \dots, M$). And this quadratic function minimization can be easily solved. Hence,

the original complex composite is split into two simpler objectives.

Apparently, now this form could be solved by a series of alternating proximal minimization methods [12]: for each $m \in \{1, \dots, M\}$,

$$x^{k+1} \in \arg \min_u \left\{ f(u) + Q(u, \theta^k) + \frac{1}{2\zeta_k} \|u - x^k\|_2^2 \right\}; \quad (9a)$$

$$\theta_m^{k+1} \in \arg \min_{v_m} \left\{ Q_m(x^{k+1}, v_m) + h_m(v_m) + \frac{1}{2\mu_k} \|v_m - \theta_m^k\|_2^2 \right\}. \quad (9b)$$

We then make the following assumptions about (9a) and (9b):

$$\inf_{x \in \prod_{m=1}^M \mathcal{U}_m} f(x) + H(\theta) + Q(x, \theta) > -\infty; \\ f(x) + H(\theta^0) + Q(x, \theta^0) \text{ is proper}; \quad (\mathcal{H})$$

for some positive $0 < r_- < r_+$, the stepsize sequences satisfy $r_- < \zeta_k$, $\mu_k < r_+$ $\forall k \geq 0$.

The assumption is weakly required and can be easily meet. In the first section, the given three examples all satisfy the first two conditions, and the third condition holds when proper parameters are set in practical tests. Besides, each of the proximal terms $(1/2\zeta_k)\|u - x^k\|_2^2$ and $(1/2\mu_k)\|v_m - \theta_m^k\|_2^2$ in (9a) and (9b) is used to meet the decrease condition of objective function. If we omit them, the algorithm can also perform well but can not build direct general convergence theories. At that time, it reduces to alternating projection minimization method.

3. Convergence Results

In fact, now the problem is a proximal alternating minimization case, whose global convergence has been detailedly analyzed in paper [12]. The paper mainly concentrates on theory analysis of problem $L(x, y) = f(x) + Q(x, y) + g(y)$ with the following form:

$$x_{k+1} \in \arg \min_x \left\{ L(x, y_k) + \frac{1}{2\zeta_k} \|x - x_k\|_2^2, x \in \mathbb{R}^m \right\}, \quad (10)$$

$$y_{k+1} \in \arg \min_y \left\{ L(x_{k+1}, y) + \frac{1}{2\mu_k} \|y - y_k\|_2^2, y \in \mathbb{R}^n \right\}.$$

And if L has the Kurdyka-Lojasiewicz (K-L) property, then each bounded sequence generated by the above algorithm converges to a critical point of L . Even convergence rate of the algorithm can be computed, which depends on the geometrical properties of the function L around its critical points. It is remarkable that assumption with K-L property can be verified in many common functions.

The convergence difference between algorithm of paper [12] and ours is that our minimization objective is not two variables (x, y) but x and a sequence of variables $\{\theta_m\}_{m=1}^M$. In this section, our work is to give similar consequence of algorithms (9a) and (9b).

3.1. Preliminary

Definition 4 (subdifferentials [24, 25]). Let $g : \mathbb{R}^N \rightarrow (-\infty, +\infty]$ be a proper and lower semicontinuous function.

- (1) For a given $x \in \text{dom}(g)$, the Fréchet subdifferential of g at x , written as $\widehat{\partial}g(x)$, is the set of all vectors $u \in \mathbb{R}^N$ which satisfy

$$\liminf_{y \neq x, y \rightarrow x} \frac{J(y) - J(x) - \langle u, y - x \rangle}{\|y - x\|_2} \geq 0. \quad (11)$$

When $x \notin \text{dom}(g)$, we set $\widehat{\partial}J(x) = \emptyset$.

- (2) The “limiting” subdifferential, or simply the subdifferential, of J at $x \in \mathbb{R}^N$, written as $\partial g(x)$, is defined through the following closure process:

$$\partial g(x) := \left\{ u \in \mathbb{R}^N : \exists x^k \rightarrow x, g(x^k) \rightarrow g(x), u^k \in \widehat{\partial}g(x^k) \rightarrow u \text{ as } k \rightarrow \infty \right\}. \quad (12)$$

A necessary condition for $z \in \mathcal{X}$ to be a minimizer of g is

$$0 \in \partial g(x). \quad (13)$$

A point that satisfies (13) is called limiting-critical or simply critical point. The set of critical points of g is denoted by $\text{crit } g$.

Being given $(x^0, \theta^0) \in \mathcal{X} \times \prod_{m=1}^M \mathcal{U}_m$, recall that sequence generated by (9a) and (9b) is of the form $(x^k, \theta^k) \rightarrow (x^{k+1}, \theta^k) \rightarrow (x^{k+1}, \theta^{k+1})$. According to the basic properties of P , we can deduce a few important points.

Corollary 5. Assume sequences (x^k, θ^k) are generated by (9a) and (9b) under assumption (\mathcal{H}) , and then they are well defined. Moreover, consider the following:

- (i) The following estimate holds:

$$P(x^k, \theta^k) + \frac{1}{2\zeta_k} \|x^k - x^{k-1}\|_2^2 + \sum_{m=1}^M \frac{1}{2\mu_k} \|\theta_m^k - \theta_m^{k-1}\|_2^2 \\ \leq P(x^{k-1}, \theta^{k-1}), \quad \forall k \geq 1; \quad (14)$$

hence $L(x^k, \theta^k)$ dose not increase.

(ii)

$$\sum_{k=1}^{\infty} \|x^k - x^{k-1}\|_2^2 + \sum_{m=1}^M \|\theta_m^k - \theta_m^{k-1}\|_2^2 < \infty; \quad (15)$$

$$\text{hence } \lim_{k \rightarrow \infty} \|x^k - x^{k-1}\|_2 + \sum_{m=1}^M \|\theta_m^k - \theta_m^{k-1}\|_2 = 0.$$

(iii) For $k \geq 1$, we have

$$\begin{aligned} & \left(\sum_{m=1}^M (\nabla_x Q_m(x^k, \theta_m^k) - \nabla_x Q_m(x^k, \theta_m^{k-1})), \mathbf{0} \right) \\ & - \left(\frac{1}{\zeta_{k-1}} (x^k - x^{k-1}), \prod_{m=1}^M \frac{1}{\mu_{k-1}} (\theta_m^k - \theta_m^{k-1}) \right) \quad (16) \\ & \in \partial P(x^k, \theta^k), \end{aligned}$$

where above $\mathbf{0}$ is a multivector with the same dimension of $\prod_{m=1}^M \theta_m$.

Besides, for all bounded subsequence (x^{k_j}, θ^{k_j}) of (x^k, θ^k) , $\text{dist}(\mathbf{0}, \partial P(x^{k_j}, \theta^{k_j})) \rightarrow 0$, $k_j \rightarrow +\infty$.

Corollary 6. Assume that (\mathcal{H}) hold. Let (x^k, θ^k) be a sequence complying with (9a) and (9b). Let $\omega(x^0, \theta^0)$ denote the set (possibly empty) of its limit points. Then

- (i) if (x^k, θ^k) is bounded, then $\omega(x^0, \theta^0)$ is a nonempty compact connected set and $\text{dist}((x^k, \theta^k), \omega(x^0, \theta^0)) \rightarrow 0$ as $k \rightarrow +\infty$,
- (ii) $\omega(x^0, \theta^0) \subset \text{crit } P$,
- (iii) P is finite and constant on $\omega(x^0, y^0)$, equal to $\inf_{k \in \mathbb{N}} P(x^k, y^k) = \lim_{k \rightarrow +\infty} P(x^k, \theta^k)$.

The above proposition gives some convergence results about sequences generated by (7), (9a), and (9b). Point (ii) guarantees that all limiting points produced by (9a) and (9b) must be limiting-critical points. And (iii) gives the point that objective P converges to the finite and constant.

3.2. Convergence to a Critical Point. This part gives more precise convergence analysis about the proximal algorithms (9a) and (9b).

Let $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. For $-\infty < \eta_1 < \eta_2 \leq +\infty$, let us set $[\eta_1 < f < \eta_2] = \{x \in \mathcal{X} : 1 < f(x) < 2\}$. Then we give an important definition in the optimization theory.

Definition 7 (K-L property [12]). The function f is said to have the K-L property at $\bar{x} \in \text{dom } \partial f$ if there exist $\eta \in (0, +\infty]$, a neighborhood U of \bar{x} , and a continuous concave function $\phi : [0, \eta] \rightarrow \mathbb{R}^+$ such that

- (i) $\phi(0) = 0$;
- (ii) ϕ is C^1 on $(0, \eta)$;
- (iii) for all $s \in (0, \eta)$, $\phi'(s) > 0$;
- (iv) for all $x \in U \cap [f(\bar{x}) < f < f(\bar{x}) + \eta]$, the Kurdyka-Lojasiewicz inequality holds:

$$\phi(f(x) - f(\bar{x})) \text{dist}(\mathbf{0}, \partial f(x)) \geq 1. \quad (17)$$

If we justify that a function has K-L property, we should estimate η, U, ϕ . Many convex functions, for instance, satisfy the above property with $U = \mathbb{R}^n$ and $\eta = +\infty$. Besides, a lot of nonconvex examples are also given in paper [12].

Below, we will give convergence analysis to critical point.

Theorem 8 (convergence). Assume that P satisfies (\mathcal{H}) and has the Kurdyka-Lojasiewicz property at each point of the domain of f . Then

- (i) either $\|(x^k, \theta^k)\|_2$ tends to infinity,
- (ii) or $(x^k - x^{k-1}, \theta^k - \theta^{k-1})$ is l^1 , that is,

$$\sum_{k=1}^{+\infty} \left(\|x^k - x^{k-1}\|_2 + \sum_{m=1}^M \|\theta_m^k - \theta_m^{k-1}\|_2 \right) < +\infty, \quad (18)$$

and, as a consequence, (x^k, θ^k) converges to a critical point of P .

The above theorem's proof is based on the same analysis process in paper [12], so here we just present the convergence results but omit their proofs.

4. Application to 1D TV Denoising

In practical scientific and engineering contexts, noise removal is the basis and requisite of other subsequential applications. It has received extensive attentions. A range of computational algorithms have been proposed to solve the denoising problem [26–28]. Among these solvers, total variation (TV) regularizer is of great importance since it can efficiently deal with noisy signals that have sparse derivatives (or gradients). For instance, piecewise constant (PWC) [29] signal with noise, whose derivative is sparse relative to signal dimension, could be denoised by powerful TV denoising method.

In 1D TV denoising problem [10], one needs to solve model (2). TV denoising minimizes a composite of two parts. The first part is to keep the error, between the observed data and the original, as small as possible. The second is devoted to minimizing the sparsity of the gradients. Usually, denoising model is defined as one combination of a quadratic data fidelity term and a convex regularization term or a differential regularization, for example, convex but nonsmooth problem [30]

$$\arg \min_{x \in \mathbb{R}^n} \{ \|y - x\|_2^2 + \lambda \|Dx\|_1 : x \in \mathbb{R}^n \} \quad (19)$$

or differential but nonconvex problem [7]

$$\begin{aligned} & \arg \min_{x \in \mathbb{R}^n} \left\{ \|y - x\|_2^2 + \lambda \sum_i \frac{1}{a} \log(1 + a |[Dx]_i|) : x \right. \\ & \left. \in \mathbb{R}^n \right\}, \quad a > 0, \end{aligned} \quad (20)$$

where $\|x\|_1 = \sum_{i=1}^n |x_i|$. Exact solution of the above two types can be obtained by very fast direct algorithms [7, 30]. In fact,

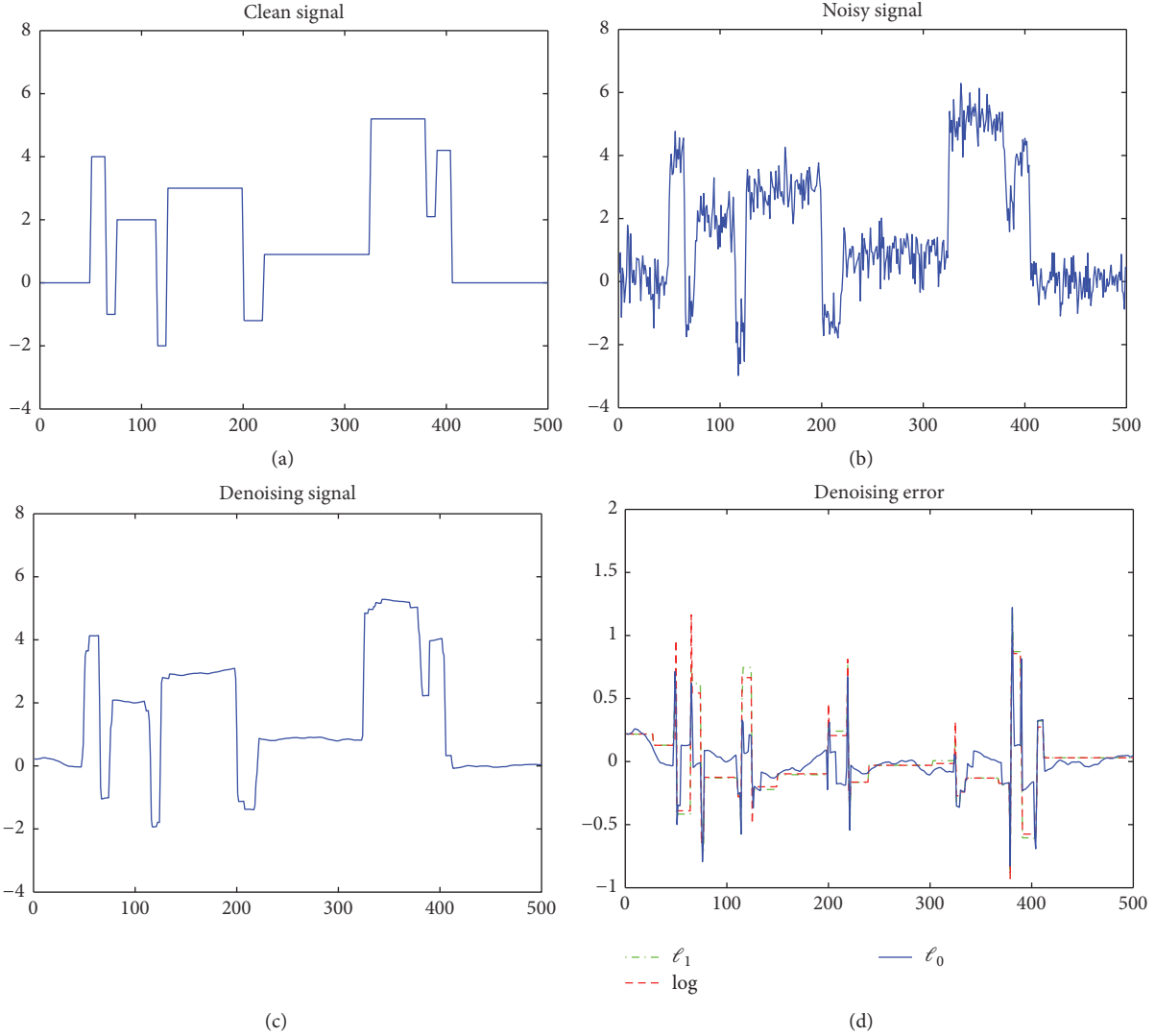


FIGURE 1: Total variation denoising with nonconvex nonsmooth ℓ_0 penalty (RMSE = 0.1709), compared with convex ℓ_1 penalty [21] (RMSE = 0.2720) and smooth log penalty [7] (RMSE = 0.2611).

convex ℓ_1 -norm is the replacement of nonconvex ℓ_0 -norm in (19) since convex optimization techniques have been deeply studied. The latter, like logarithmic penalty and arctangent penalty, can be solved by MM update iteration, in which total objective function (including data fidelity and regularization terms) should meet strictly convex condition.

In this test, we apply our algorithms (9a) and (9b) to this example. Auxiliary variable θ is introduced to reduce complexity of the composition $\|Dx\|_0$. Then (2) is represented as

$$\arg \min_{x \in \mathbb{R}^n, \theta \in \mathbb{R}^{n-1}} \left\{ \|x - y\|_2^2 + \frac{\tau}{2} \|Dx - \theta\|_2^2 + \lambda \|\theta\|_0 \right\}. \quad (21)$$

Apparently this problem satisfies the convergence conditions [12]. Concrete steps by algorithms (9a) and (9b) are shown in

$$\begin{aligned} x^{k+1} \in \arg \min_{u \in \mathbb{R}^n} & \left\{ \|u - y\|_2^2 + \frac{\tau}{2} \|Du - \theta^k\|_2^2 \right. \\ & \left. + \frac{1}{2\zeta_k} \|u - x^k\|_2^2 \right\}; \end{aligned} \quad (22a)$$

$$\begin{aligned} \theta^{k+1} \in \arg \min_{v \in \mathbb{R}^{n-1}} & \left\{ \lambda \|v\|_0 + \frac{\tau}{2} \|Dx^{k+1} - v\|_2^2 \right. \\ & \left. + \frac{1}{2\mu_k} \|v - \theta^k\|_2^2 \right\}. \end{aligned} \quad (22b)$$

In fact, when tests ζ_k and μ_k can be set as very large constants, the last proximal terms $(1/2\zeta_k)\|u - x^k\|_2^2$ and $(1/2\mu_k)\|v - \theta^k\|_2^2$ in each computing step could be negligible. Hence, the computation of (22a) and (22b) is as follows.

Computation of (22a). The former step (22a) is a quadratic function and could be computed through many techniques, like gradient descent.

Computation of (22b). Apparently, the latter (22b) could be rewritten as a proximal operator of function $g(\theta)$ [12]; that is, $\text{prox}_{\lambda/\tau \|\theta\|_0}(Dx^{k+1})$. Consider the proximal operator

$\text{prox}_{\lambda/\tau\|\cdot\|_0}(u)$. When $n = 1$, ℓ_0 norm is reduced to $|\cdot|_0$, where one easily establishes that

$$\text{prox}_{\lambda/\tau\|\cdot\|_0}(u) = \begin{cases} u, & \text{if } |u| > \sqrt{\frac{2\lambda}{\tau}}; \\ \{0, u\}, & \text{if } |u| = \sqrt{\frac{2\lambda}{\tau}}; \\ 0, & \text{otherwise.} \end{cases} \quad (23)$$

When n is arbitrary, trivial algebraic manipulations are given, with $u = (u_1, u_2, \dots, u_n) \in \mathbb{R}^n$:

$$[\text{prox}_{\lambda/\tau\|\cdot\|_0}(u)]_i = \text{prox}_{\lambda/\tau\|\cdot\|_0}(u_i), \quad (24)$$

and thus $\text{prox}_{\lambda/\tau\|\cdot\|_0}(Dx^{k+1})$ is a perfectly known object.

Total variation denoising examples with three convex and nonconvex regularization instances (the two others are convex and nonconvex but smooth algorithms in [7, 30], resp.) are figured in Figure 1. Original piece signal data $x^0 \in \mathbb{R}^n$ with length $n = 256$ is obtained with MakeSignal in paper [7]. The noisy data is obtained using additive white Gaussian noise (AWGN) ($\sigma = 0.5$). For both convex and nonconvex cases, we set $\lambda = \sqrt{n}\sigma/4$, consistent with the range suggested in [30] for standard (convex) TV denoising and nonconvexity parameter is set to its maximal value, $a = 1/(4\lambda)$ default in [7]. These settings could lead to the best denoising result in their papers. All the other settings are consistent with paper [7]. The maximum iteration numbers are all 500. All the codes are tested in the same computer.

According to the comparison between our algorithm for TV- ℓ_0 norm and the proposed algorithms in papers [7, 30], our algorithm has better result with smaller Root-Mean-Square-Error (RMSE), where $\text{RMSE}(x) := \sqrt{(1/n) \sum_{i=1}^n \|x_i - x_i^0\|_2^2}$. Referring to Figure 1, the best RMSE results for 1D TV denoising with convex ℓ_1 penalty [21] and smooth log penalty [7] are 0.2720 and 0.2611, respectively. And ours is 0.1709, much better than the convex and smooth cases.

5. Conclusion

Nonconvex nonsmooth algorithm finds many interesting applications in many fields. In this paper, we give a general proximal alternating minimization method for a kind of nonconvex nonsmooth problems with complex composition. It has concise form, good theory results, and promising numerical result. For specific 1D standard TV denoising problem, the improvement is more dramatic compared to the existing algorithms [7, 30]. Besides, our algorithm works on other nonconvex nonsmooth problems, such as block sparse recovery, group lasso, and image deconvolution, of which the examples are just too numerous to mention.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

The work is supported in part by National Natural Science Foundation of China, no. 61571008.

References

- [1] P. L. Combettes and J.-C. Pesquet, "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators," *Set-Valued and Variational Analysis*, vol. 20, no. 2, pp. 307–330, 2012.
- [2] L. Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *Journal of Optimization Theory and Applications*, vol. 158, no. 2, pp. 460–479, 2013.
- [3] J. E. Esser, *Primal dual algorithms for convex models and applications to image restoration, registration and nonlocal inpainting [Ph.D. thesis]*, University of California, Los Angeles, Calif, USA, 2010.
- [4] Y. B. Zhao and D. Li, "Reweighted ℓ_1 -minimization for sparse solutions to underdetermined linear systems," *SIAM Journal on Optimization*, vol. 22, no. 3, pp. 1065–1088, 2012.
- [5] J. Wright and Y. Ma, "Dense error correction via-minimization," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3540–3560, 2010.
- [6] T. Sun, H. Zhang, and L. Cheng, "Subgradient projection for sparse signal recovery with sparse noise," *Electronics Letters*, vol. 50, no. 17, pp. 1200–1202, 2014.
- [7] I. W. Selesnick, A. Parekh, and I. Bayram, "Convex 1-D total variation denoising with non-convex regularization," *IEEE Signal Processing Letters*, vol. 22, no. 2, pp. 141–144, 2015.
- [8] L. He and S. Schaefer, "Mesh denoising via L_0 minimization," *ACM Transactions on Graphics*, vol. 32, no. 4, article 64, 2013.
- [9] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via L_0 gradient minimization," *ACM Transactions on Graphics*, vol. 30, no. 6, article 174, 2011.
- [10] C. Brandt, H.-P. Seidel, and K. Hildebrandt, "Optimal spline approximation via ℓ_0 -minimization," *Computer Graphics Forum*, vol. 34, no. 2, pp. 617–626, 2015.
- [11] H. Zhang, L. Cheng, and W. Yin, "A dual algorithm for a class of augmented convex models," <https://arxiv.org/abs/1308.6337>.
- [12] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Lojasiewicz inequality," *Mathematics of Operations Research*, vol. 35, no. 2, pp. 438–457, 2010.
- [13] H. Attouch, J. Bolte, and B. F. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods," *Mathematical Programming*, vol. 137, no. 1-2, pp. 91–129, 2013.
- [14] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, vol. 146, no. 1-2, pp. 459–494, 2014.
- [15] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 73, no. 3, pp. 267–288, 1996.
- [16] G. Chen and M. Teboulle, "A proximal-based decomposition method for convex minimization problems," *Mathematical Programming*, vol. 64, no. 1-3, pp. 81–101, 1994.

- [17] L. Meier, S. Van De Geer, and P. Bhlmann, "The group Lasso for logistic regression," *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, vol. 70, no. 1, pp. 53–71, 2008.
- [18] X. Liao, H. Li, and L. Carin, "Generalized alternating projection for weighted-2, 1 minimization with applications to model-based compressive sensing," *SIAM Journal on Imaging Sciences*, vol. 7, no. 2, pp. 797–823, 2014.
- [19] E. Elhamifar and R. Vidal, "Block-sparse recovery via convex optimization," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4094–4107, 2012.
- [20] L. Condat, "A generic proximal algorithm for convex optimization—application to total variation minimization," *IEEE Signal Processing Letters*, vol. 21, no. 8, pp. 985–989, 2014.
- [21] T. F. Chan and C.-K. Wong, "Total variation blind deconvolution," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 370–375, 1998.
- [22] L. He, A. Marquina, and S. J. Osher, "Blind deconvolution using TV regularization and Bregman iteration," *International Journal of Imaging Systems and Technology*, vol. 15, no. 1, pp. 74–83, 2005.
- [23] N. Dey, L. Blanc-Feraud, C. Zimmer et al., "Richardson-Lucy algorithm with total variation regularization for 3D confocal microscope deconvolution," *Microscopy Research and Technique*, vol. 69, no. 4, pp. 260–266, 2006.
- [24] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation I: Basic Theory*, Springer Science & Business Media, Berlin, Germany, 2006.
- [25] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*, Springer Science & Business Media, Berlin, Germany, 2009.
- [26] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.
- [27] M. Lysaker, A. Lundervold, and X.-C. Tai, "Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time," *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1579–1590, 2003.
- [28] M. Lysaker, S. Osher, and X.-C. Tai, "Noise removal using smoothed normals and surface fitting," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1345–1357, 2004.
- [29] M. A. Little and N. S. Jones, "Generalized methods and solvers for noise removal from piecewise constant signals. I. Background theory," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 467, no. 2135, pp. 3088–3114, 2011.
- [30] L. Dumbgen and A. Kovac, "Extensions of smoothing via taut strings," *Electronic Journal of Statistics*, vol. 3, pp. 41–75, 2009.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

