Abstract—Total variation denoising is a nonlinear filtering method well suited for the estimation of piecewise-constant signals observed in additive white Gaussian noise. The method is defined by the minimization of a particular nondifferentiable convex cost function. This letter describes a generalization of this cost function that can yield more accurate estimation of piecewise-constant signals. The new cost function involves a nonconvex penalty (regularizer) designed to maintain the convexity of the cost function. The new penalty is based on the Moreau envelope. The proposed total variation denoising method can be implemented using forward–backward splitting.

Index Terms—Convex, denoising, sparse, total variation (TV).

I. INTRODUCTION

TOTAL variation (TV) denoising is a nonlinear filtering method based on the assumption that the underlying signal is piecewise constant (equivalently, the derivative of the underlying signal is sparse) [45]. Such signals arise in geoscience, biophysics, and other areas [31]. The TV denoising technique is also used in conjunction with other methods in order to process more general types of signals [20], [23], [24], [26].

TV denoising is prototypical of methods based on sparse signal models. It is defined by the minimization of a convex cost function comprising a quadratic data fidelity term and a nondifferentiable convex penalty term. The penalty term is the composition of a linear operator and the \( \ell_1 \) norm. Although the \( \ell_1 \) norm stands out as the convex penalty that most effectively induces sparsity [27], nonconvex penalties can lead to more accurate estimation of the underlying signal [37], [38], [40], [44], [50].

A few recent papers consider the prescription of nonconvex penalties that maintain the convexity of the TV denoising cost function [1], [30], [32], [48]. (The motivation for this is to leverage the benefits of both nonconvex penalization and convex optimization, e.g., to accurately estimate the amplitude of jump discontinuities while guaranteeing the uniqueness of the solution.) The penalties considered in these works are separable (additive). But nonseparable penalties can outperform separable penalties in this context. This is because preserving the convexity of the cost function is a severely limiting requirement. Nonseparable penalties can more successfully meet this requirement because they are more general than separable penalties [47].

This letter proposes a nonseparable nonconvex penalty for total variation denoising that generalizes the standard penalty and maintains the convexity of the cost function to be minimized. The new penalty, which is based on the Moreau envelope, can more accurately estimate the amplitudes of jump discontinuities in an underlying piecewise constant signal.

A. Relation to Prior Work

Numerous nonconvex penalties and algorithms have been proposed to outperform \( \ell_1 \)-norm regularization for the estimation of sparse signals e.g., [8], [10], [12], [13], [15], [33], [34], [39], [43], [51], [53]. However, a few of these methods maintain the convexity of the cost function. The prescription of nonconvex penalties maintaining cost function convexity was pioneered by Blake and Zisserman [6], and Nikolova [36], [39], [40], and further developed in [3], [4], [14], [20], [28], [30], [32], [42], [46], [48]. These works rely on the presence of both strongly and weakly convex terms, which is also exploited in [35].

The proposed penalty is expressed as a differentiable convex function subtracted from the standard penalty (i.e., \( \ell_1 \) norm). Previous works also use this idea [41], [42], [47]. But the differentiable convex functions used therein are either separable [41], [42] or sums of bivariate functions [47].

In parallel with the submission of this letter, Carlsson has also proposed using Moreau envelopes to prescribe nontrivial convex cost functions [9]. While the approach in [9] starts with a given nonconvex cost function (e.g., with the \( \ell_0 \) pseudo-norm penalty) and seeks the convex envelope, our approach starts with the \( \ell_1 \)-norm penalty and seeks a class of convexity-preserving penalties.

Some forms of generalized TV are based on infimal convex enve- lopment (related to the Moreau envelope) [5], [7], [11], [49]. But these works propose convex penalties suitable for nonpiecewise-constant signals, while we propose nonconvex penalties suitable for piecewise-constant signals.

II. TOTAL VARIATION DENOISING

Definition 1: Given \( y \in \mathbb{R}^N \) and \( \lambda > 0 \), total variation denoising is defined as

\[
\text{tvd}(y; \lambda) = \arg \min_{x \in \mathbb{R}^N} \left\{ \frac{1}{2} \| y - x \|^2_2 + \lambda \| Dx \|_1 \right\} \tag{1}
\]

\[
= \text{prox}_{\lambda D_1}(y) \tag{2}
\]

where \( D \) is the \((N - 1) \times N \) matrix

\[
D = \begin{bmatrix}
-1 & 1 & 0 & \cdots & 0 \\
-1 & -1 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & -1 & 1
\end{bmatrix} \tag{3}
\]
As indicated in (2), TV denoising is the proximity operator [16] of the function \( x \mapsto \lambda \| Dx \|_1 \). Notably, TV denoising can be calculated exactly in finite-time [17], [18], [22], [29].

### III. Moreau Envelope

Before we define the nondifferentiable nonconvex penalty in Section IV, we first define a differentiable convex function. We use the Moreau envelope from convex analysis [2].

**Definition 2:** Let \( \alpha \geq 0 \). We define \( S_\alpha : \mathbb{R}^N \rightarrow \mathbb{R} \) as

\[
S_\alpha(x) = \min_{v \in \mathbb{R}^N} \left\{ \| Dv \|_1 + \frac{\alpha}{2} \| x - v \|_2^2 \right\}
\]

(4)

where \( D \) is the first-order difference matrix (3).

If \( \alpha > 0 \), then \( S_\alpha \) is the Moreau envelope of index \( \alpha^{-1} \) of the function \( x \mapsto \| Dx \|_1 \).

**Proposition 1:** The function \( S_\alpha \) can be calculated by

\[
\begin{align*}
S_0(x) &= 0 \\
S_\alpha(x) &= \| D \text{tvd}(x; 1/\alpha) \|_1 \\
& \quad + \frac{\alpha}{2} \| x - \text{tvd}(x; 1/\alpha) \|_2^2, \quad \alpha > 0.
\end{align*}
\]

(5)

**Proof:** For \( \alpha = 0 \): Setting \( \alpha = 0 \) and \( v = 0 \) in (4) gives (5).

For \( \alpha > 0 \): By the definition of TV denoising, the \( v \in \mathbb{R}^N \) minimizing the function in (4) is the TV denoising of \( x \), i.e.,

\[ v_{\text{opt}} = \text{tvd}(x; 1/\alpha). \]

**Proposition 2:** Let \( \alpha \geq 0 \). The function \( S_\alpha \) satisfies

\[
0 \leq S_\alpha(x) \leq \| Dx \|_1 \quad \forall x \in \mathbb{R}^N.
\]

(7)

**Proof:** From (4), we have \( S_\alpha(x) \leq \| Dv \|_1 + (\alpha/2)\| x - v \|_2^2 \) for all \( v \in \mathbb{R}^N \). In particular, \( v = x \) leads to \( S_\alpha(x) \leq \| Dx \|_1 \). Also, \( S_\alpha(x) \geq 0 \) since \( S_\alpha(x) \) is defined as the minimum of a non-negative function.

**Proposition 3:** Let \( \alpha \geq 0 \). The function \( S_\alpha \) is convex and differentiable.

**Proposition 4:** Let \( \alpha \geq 0 \). The gradient of \( S_\alpha \) is given by

\[
\nabla S_0(x) = 0
\]

(8)

\[
\nabla S_\alpha(x) = \alpha \left( x - \text{tvd}(x; 1/\alpha) \right), \quad \alpha > 0
\]

(9)

where \( \text{tvd} \) denotes total variation denoising (1).

**Proof:** Since \( S_\alpha \) is the Moreau envelope of index \( \alpha^{-1} \) of the function \( x \mapsto \| Dx \|_1 \) when \( \alpha > 0 \), it follows by Proposition 12.29 in [2] that

\[
\nabla S_\alpha(x) = \alpha \left( x - \text{prox}_{(1/\alpha)\| \cdot \|_1}(x) \right).
\]

(10)

This proximity operator is TV denoising, giving (9).

### IV. Nonconvex Penalty

To strongly induce sparsity of \( Dx \), we define a nonconvex generalization of the standard TV penalty. The new penalty is defined by subtracting a differentiable convex function from the standard penalty.

**Definition 3:** Let \( \alpha \geq 0 \). We define the penalty \( \psi_\alpha : \mathbb{R}^N \rightarrow \mathbb{R} \) as

\[
\psi_\alpha(x) = \| Dx \|_1 - S_\alpha(x)
\]

(11)

where \( D \) is the matrix (3) and \( S_\alpha \) is defined by (4).

The proposed penalty is upper bounded by the standard TV penalty, which is recovered as a special case.

**Proposition 5:** Let \( \alpha \geq 0 \). The penalty \( \psi_\alpha \) satisfies

\[
\psi_\alpha(x) = \| Dx \|_1 \quad \forall x \in \mathbb{R}^N
\]

(12)

\[
0 \leq \psi_\alpha(x) \leq \| Dx \|_1 \quad \forall x \in \mathbb{R}^N.
\]

(13)

**Proof:** It follows from (5) and (7).

When a convex function is subtracted from another convex function [as in (11)], the resulting function may well be negative on part of its domain. Inequality (13) states that the proposed penalty \( \psi_\alpha \) avoids this fate. This is relevant because the penalty function should be non-negative.

Figures in the supplemental material show examples of the proposed penalty \( \psi_\alpha \) and the function \( S_\alpha \).

### V. Enhanced TV Denoising

We define “Moreau-enhanced” TV denoising. If \( \alpha > 0 \), then the proposed penalty penalizes large amplitude values of \( Dx \) less than the \( \ell_1 \) norm does (i.e., \( \psi_\alpha(x) \leq \| Dx \|_1 \)), hence it is less likely to underestimate jump discontinuities.

**Definition 4:** Given \( y \in \mathbb{R}^N \), \( \lambda > 0 \), and \( \alpha \geq 0 \), we define Moreau-enhanced total variation denoising as

\[
\text{mtvd}(y; \lambda, \alpha) = \arg \min_{x \in \mathbb{R}^N} \left\{ \frac{1}{2} \| y - x \|_2^2 + \lambda \psi_\alpha(x) \right\}
\]

(14)

where \( \psi_\alpha \) is given by (11).

The parameter \( \alpha \) controls the nonconvexity of the penalty. If \( \alpha = 0 \), then the penalty is convex and Moreau-enhanced TV denoising reduces to TV denoising. Greater values of \( \alpha \) make the penalty more nonconvex. What is the greatest value of \( \alpha \) that maintains convexity of the cost function? The critical value is given by Theorem 1.

**Theorem 1:** Let \( \lambda > 0 \) and \( \alpha \geq 0 \). Define \( F_\alpha : \mathbb{R}^N \rightarrow \mathbb{R} \) as

\[
F_\alpha(x) = \frac{1}{2} \| y - x \|_2^2 + \lambda \psi_\alpha(x)
\]

(15)

where \( \psi_\alpha \) is given by (11). If

\[
0 \leq \alpha \leq 1/\lambda
\]

(16)

then \( F_\alpha \) is convex. If \( 0 \leq \alpha < 1/\lambda \) then \( F_\alpha \) is strongly convex.

**Proof:** We write the cost function as

\[
F_\alpha(x) = \frac{1}{2} \| y - x \|_2^2 + \lambda \| Dx \|_1 - \lambda S_\alpha(x)
\]

(17)

\[
= \frac{1}{2} \| y - x \|_2^2 + \lambda \| Dx \|_1
\]

(18)

\[
- \lambda \min_{v \in \mathbb{R}^N} \left\{ \| Dv \|_1 + \frac{\alpha}{2} \| x - v \|_2^2 \right\}
\]

\[
= \max_{v \in \mathbb{R}^N} \left\{ \frac{1}{2} \| y - x \|_2^2 + \lambda \| Dx \|_1
\]

(19)

\[
- \lambda \| Dv \|_1 - \frac{\lambda \alpha}{2} \| x - v \|_2^2 \right\}
\]

\[
= \max_{v \in \mathbb{R}^N} \left\{ \frac{1}{2} (1 - \lambda \alpha) \| x \|_2^2 + \lambda \| Dx \|_1 + g(x, v) \right\}
\]

(20)

\[
= \frac{1}{2} (1 - \lambda \alpha) \| x \|_2^2 + \lambda \| Dx \|_1 + \max_{v \in \mathbb{R}^N} g(x, v)
\]

(21)

where \( g(x, v) \) is affine in \( x \). The last term is convex as it is the pointwise maximum of a set of convex functions. Hence, \( F_\alpha \) is a convex function if \( 1 - \lambda \alpha \geq 0 \). If \( 1 - \lambda \alpha > 0 \), then \( F_\alpha \) is strongly convex (and strictly convex).
VI. ALGORITHM

Proposition 6: Let \( y \in \mathbb{R}^N, \lambda > 0, \) and \( 0 < \alpha < 1/\lambda. \) Then, \( x^{(k)} \) produced by the iteration

\[
\begin{align*}
    z^{(k)} &= y + \lambda \alpha (x^{(k)} - \text{tvd}(x^{(k)}; 1/\alpha)) \\
    x^{(k+1)} &= \text{tvd}(z^{(k)}; \lambda)
\end{align*}
\]

(22a, 22b)

converges to the solution of the Moreau-enhanced TV denoising problem (14).

Proof: If the cost function (15) is strongly convex, then the minimizer can be calculated using the forward–backward splitting (FBS) algorithm [2], [16]. This algorithm minimizes a function of the form

\[
F(x) = f_1(x) + f_2(x)
\]

(23)

where both \( f_1 \) and \( f_2 \) are convex and \( \nabla f_1 \) is additionally Lipschitz continuous. The FBS algorithm is given by

\[
\begin{align*}
    z^{(k)} &= x^{(k)} - \mu \nabla f_1(x^{(k)}) \\
    x^{(k+1)} &= \text{arg min}_x \left\{ \frac{1}{2} \| z^{(k)} - x \|^2 + \mu f_2(x) \right\}
\end{align*}
\]

(24a, 24b)

where \( 0 < \mu < 2/\rho \) and \( \rho \) is the Lipschitz constant of \( \nabla f_1 \).

The iterates \( x^{(k)} \) converge to a minimizer of \( F \).

To apply the FBS algorithm to the proposed cost function (15), we write it as

\[
F_n(x) = \frac{1}{2} \| y - x \|^2 + \lambda \psi_n(x)
\]

(25)

\[
= \frac{1}{2} \| y - x \|^2 + \lambda \| Dx \|_1 - \lambda S_n(x)
\]

(26)

\[
= f_1(x) + f_2(x)
\]

(27)

where

\[
\begin{align*}
    f_1(x) &= \frac{1}{2} \| y - x \|^2 - \lambda S_n(x) \\
    f_2(x) &= \lambda \| Dx \|_1.
\end{align*}
\]

(28a, 28b)

The gradient of \( f_1 \) is given by

\[
\nabla f_1(x) = x - y - \lambda \nabla S_n(x)
\]

(29)

\[
= x - y - \lambda \alpha (x - \text{tvd}(x; 1/\alpha))
\]

(30)

using Proposition 4. Subtracting \( S_n \) from \( f_1 \) does not increase the Lipschitz constant of \( \nabla f_1 \), the value of which is 1. Hence, we may set \( 0 < \mu < 2 \).

Using (28), the FBS algorithm (24) becomes

\[
\begin{align*}
    z^{(k)} &= x^{(k)} - \mu [x^{(k)} - y] - \lambda \alpha (x^{(k)} - \text{tvd}(x^{(k)}; 1/\alpha)) \\
    x^{(k+1)} &= \text{arg min}_x \left\{ \frac{1}{2} \| z^{(k)} - x \|^2 + \mu \lambda \| Dx \|_1 \right\}
\end{align*}
\]

(31a, 31b)

Note that (31b) is TV denoising (1). Using the value \( \mu = 1 \) gives iteration (22). (Experimentally, we found this value yields fast convergence.)

Each iteration of (22) entails solving two standard TV denoising problems. In this work, we calculate TV denoising using the fast exact C language program by Condat [17]. Like the iterative shrinkage/thresholding algorithm [19], [25], algorithm (22) can be accelerated in various ways.

We suggest not setting \( \alpha \) too close to the critical value \( 1/\lambda \) because the FBS algorithm generally converges faster when the cost function is more strongly convex (\( \alpha < 1/\lambda \)).

In summary, the proposed Moreau-enhanced TV denoising method comprises the steps:

1) Set the regularization parameter \( \lambda (\lambda > 0) \).
2) Set the nonconvexity parameter \( \alpha (0 \leq \alpha < 1/\lambda) \).
3) Initialize \( x^{(0)} = 0 \).
4) Run iteration (22) until convergence.

VII. OPTIMALITY CONDITION

To avoid terminating the iterative algorithm too early, it is useful to verify convergence using an optimality condition.

Proposition 7: Let \( y \in \mathbb{R}^N, \lambda > 0, \) and \( 0 < \alpha < 1/\lambda. \) If \( x \) is a solution to (14), then

\[
[C((x - y)/\lambda + \alpha (\text{tvd}(x; 1/\alpha) - x))]_n \in \text{sign}([Dx]_n)
\]

(32)

for \( n = 0, \ldots, N - 1 \), where \( C \subseteq \mathbb{R}^{(N-1) \times N} \) is given by

\[
C_{m,n} = \begin{cases} 1, & m \geq n \\ 0, & m < n \end{cases}
\]

i.e., \( [Cx]_n = \sum_{m \leq n} x_m \)

(33)

and \( \text{sign} \) is the set-valued signum function

\[
\text{sign}(t) = \begin{cases} \{-1\} & t < 0 \\ [-1, 1] & t = 0 \\ \{1\} & t > 0 \end{cases}
\]

(34)

According to (32), if \( x \in \mathbb{R}^N \) is a minimizer, then the points \(([Dx]_n, u_n) \in \mathbb{R}^2\) must lie on the graph of the signum function, where \( u_n \) denotes the value on the left-hand side of (32). Hence, the optimality condition can be depicted as a scatter plot. Figures in the supplemental material show how the points in the scatter plot converge to the signum function as the algorithm (22) progresses.

Proof of Proposition 7: A vector \( x \) minimizes a convex function \( F \) if \( 0 \in \partial F(x) \) where \( \partial F(x) \) is the subdifferential of \( F \) at \( x \). The subdifferential of the cost function (15) is given by

\[
\partial F_n(x) = \{ x - y - \lambda \nabla S_n(x) + \partial (\lambda \| D \cdot |_1)(x) \}
\]

(35)

which can be written as

\[
\partial F_n(x) = \{ x - y - \lambda \nabla S_n(x) + \lambda D^T u \}
\]

\[
: u_n \in \text{sign}([Dx]_n), u \in \mathbb{R}^{N-1}.
\]

(36)

Hence, the condition \( 0 \in \partial F_n(x) \) can be written as

\[
(y - x)/\lambda + \nabla S_n(x) 
\]

\[
in \{ D^T u : u_n \in \text{sign}([Dx]_n), u \in \mathbb{R}^{N-1} \}.
\]

(37)

Let \( C \) be a matrix of size \((N - 1) \times N\) such that \( CD^T = -I \), e.g., (33). It follows that the condition \( 0 \in \partial F_n(x) \) implies that

\[
[C((x - y)/\lambda - \nabla S_n(x))]_n \in \text{sign}([Dx]_n)
\]

(38)

for \( n = 0, \ldots, N - 1 \). Using Proposition 4 gives (32).

VIII. EXAMPLE

This example applies TV denoising to the noisy piecewise constant signal shown in Fig. 1(a). This is the “blocks” signal (length \( N = 256 \)) generated by the Wavelab [21] function
SELESNICK: TOTAL VARIATION DENOISING VIA THE MOREAU ENVELOPE

Fig. 1. TV denoising using three different penalties. (The dashed line is the true noise-free signal.) (a) Noisy signal ($\sigma = 0.50$). (b) Standard convex penalty (L1 norm). (c) Scalar minimax-concave (MC) penalty. (d) Proposed penalty.

MakeSignal with additive white Gaussian noise ($\sigma = 0.5$).

We set the regularization parameter to $\lambda = \sqrt{N\sigma}/4$ following a discussion in [22]. For Moreau-enhanced TV denoising, we set the nonconvexity parameter to $\alpha = 0.7/\lambda$.

Fig. 1 shows the result of TV denoising with three different penalties. In each case, a convex cost function is minimized. Fig. 1(b) shows the result using standard TV denoising (i.e., using the $\ell_1$-norm). This denoised signal consistently underestimates the amplitudes of jump discontinuities, especially those occurring near other jump discontinuities of opposite sign. Fig. 1(c) shows the result using a separable nonconvex penalty [48]. This method can use any nonconvex scalar penalty satisfying a prescribed set of properties. Here we use the MC penalty [3], [52] with nonconvexity parameter set to maintain cost function convexity. This result significantly improves the root-mean-square error (RMSE) and mean-absolute-deviation (MAE), but still underestimates the amplitudes of jump discontinuities.

Moreau-enhanced TV denoising, shown in Fig. 1(d), further reduces the RMSE and MAE and more accurately estimates the amplitudes of jump discontinuities. The proposed nonseparable nonconvex penalty avoids the consistent underestimation of discontinuities seen in Fig. 1(b) and (c).

To further compare the denoising capability of the considered penalties, we calculate the average RMSE as a function of the noise level. We let the noise standard deviation span the interval $0.2 \leq \sigma \leq 1.0$. For each $\sigma$ value, we calculate the average RMSE of 100 noise realizations. Fig. 2 shows that the proposed penalty yields the lowest average RMSE for all $\sigma \geq 0.4$. However, at low noise levels, separable convexity-preserving penalties [48] perform better than the proposed nonseparable convexity-preserving penalty.

Fig. 2. TV denoising using four penalties: RMSE as a function of noise level.

IX. CONCLUSION

This letter demonstrates the use of the Moreau envelope to define a nonseparable nonconvex TV denoising penalty that maintains the convexity of the TV denoising cost function. The basic idea is to subtract from a convex penalty its Moreau envelope. This idea should also be useful for other problems, e.g., analysis tight-frame denoising [41].

Separable convexity-preserving penalties [48] outperformed the proposed one at low noise levels in the example. It is yet to be determined if a more general class of convexity-preserving penalties can outperform both across all noise levels.
