

# The Suppression of Transient Artifacts in Time Series via Convex Analysis

Yining Feng<sup>1</sup>, Harry Graber<sup>2</sup>, and Ivan Selesnick<sup>1</sup>

(1) Tandon School of Engineering, New York University, Brooklyn, New York

(2) Photon Migration Technologies Corp., New York

Email: yf889@nyu.edu, harry@photonmigration.com, selesi@nyu.edu

**Abstract**—For the suppression of transient artifacts in time series data, we propose a non-convex generalized fused lasso penalty for the estimation of signals comprising a low-pass signal, a sparse piecewise constant signal, and additive white Gaussian noise. The proposed non-convex penalty is designed so as to preserve the convexity of the total cost function to be minimized, thereby realizing the benefits of a convex optimization framework (reliable, robust algorithms, etc.). Compared to the conventional use of L1 norm penalty, the proposed non-convex penalty does not underestimate the true amplitude of signal values. We derive a fast proximal algorithm to implement the method. The proposed method is demonstrated on the suppression of artifacts in near infrared spectroscopic (NIRS) measures.

## I. INTRODUCTION

Transient artifacts sometimes plague acquired biomedical signals, e.g., EEG/ECG [10], near infrared spectroscopic (NIRS) [9], infrared oculography (IROG) [4]. Thus, the suppression of transient artifacts without distorting the underlying signal of interest is important. Traditional linear time-invariant (LTI) filtering fails at this task because it requires that the frequency bands of the artifacts and of the underlying signal of interest do not overlap. That is usually not the case.

This paper considers the suppression of transient artifacts in noisy signals, where the artifacts are spikes or brief waves. We model the observed time series as the superposition of two morphologically distinct components [24] in additive noise:

$$y = f + x + w, \quad y, f, x, w \in \mathbb{R}^N. \quad (1)$$

Here  $f$  is a low-pass signal,  $x$  comprises transient artifacts, and  $w$  is the additive white Gaussian noise. We model  $x$  as a sparse signal piecewise constant signal. As an example, Fig. 1 illustrates the two components  $f$  and  $x$ , the clean signal  $f+x$ , and noisy observation  $y$ .

Given the time series  $y$ , we formulate the estimation of  $f$  and  $x$  as the solution to an optimization problem where the time series  $f$  and  $x$  are the optimization variables. The quality of the estimation of the underlying signal  $f$  depends on the quality of the estimation of the transient artifacts  $x$ ; thus the better  $x$  is estimated, the better  $f$  is estimated. We employ a sparse optimization approach which defines a type of nonlinear filter. Such optimization problems usually involve an  $\ell_1$  norm regularizer to induce sparsity.

The  $\ell_1$  norm is classically adopted to induce sparsity; however, it has a tendency to underestimate the true values of sparse signals. Hence, various non-convex regularizers are

often considered as alternatives to the  $\ell_1$  norm. However, then the objective function is generally non-convex and has extraneous suboptimal local minimizers [17]. To avoid such complications, it is advantageous to maintain the convexity of the objective function. This is possible, even when the regularizer is not convex, provided the non-convex regularizer is carefully defined [2], [3], [18], [21].

In this paper, we propose a new non-convex regularizer that improves upon the  $\ell_1$  norm and maintains the convexity of the objective function to be minimized. The proposed non-convex regularizer is specifically for the estimation of sparse piecewise constant signals. The proposed regularizer is defined using a generalized Moreau envelope (GME), a generalization of the well-known Moreau envelope defined in convex analysis [21]. While the generalized Moreau envelope of a convex function is always convex, in this paper, it is used to construct a non-convex function. We can prescribe the proposed regularizer in a way such that the objective function is convex. We provide a simple forward backward splitting (FBS) algorithm to reliably obtain the global minimum of the proposed convex optimization problem.

This paper is organized as follows. In Section II, we describe the fused lasso penalty for modeling sparse piecewise constant signals, we define the generalized Moreau envelope of a function, and we provide a formula for its gradient. In Section III, we formulate the transient suppression problem as a convex optimization problem using the conventional (convex) fused lasso penalty. In Section IV, we introduce a new non-convex generalized fused lasso penalty, which we define using the generalized Moreau envelope. In Section V, we formulate the transient suppression problem as a convex optimization problem using the new non-convex generalized fused lasso penalty. We also provide an iterative algorithm, based on forward-backward splitting (FBS) to solve this optimization problem. In Section VI, we demonstrate the proposed approach with simulated data and real NIRS data.

### A. Related Work

Numerous methods have been proposed to suppress transient artifacts in biomedical time series, e.g., NIRS data. A thorough review and comparison of five classical methods (both linear and nonlinear) was presented in Ref. [6]: band-pass filtering, principle component analysis, Kalman filtering, spline interpolation, and wavelet thresholding. Linear filtering

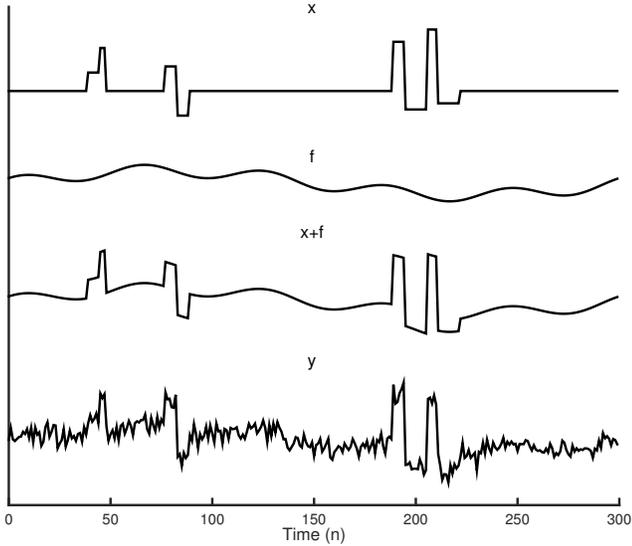


Fig. 1. Illustration of transient artifacts  $x$ , low pass signal  $f$ , clean signal  $x + f$  and noisy observation  $y$ .

techniques fail, as expected. Spline and wavelets were the best performing methods. Wavelet thresholding is a non-linear filtering technique [10], [15], [16]. Wavelet transforms decompose signals into multiple subbands: the transient artifacts manifest as large wavelet coefficients in the subbands, while the low-pass signal  $f$  is contained in only the low frequency subbands. By thresholding the wavelet coefficients and then reconstructing the data from the thresholded wavelet coefficients, the signal of transients can be estimated. For the spline interpolation approach, after identification of the segments containing the transient artifacts in the time domain, each segment is modeled separately by cubic spline interpolation [20]. The artifacts are extracted accordingly.

Recently, other methods has been explored, including non-negative matrix factorization [7] (non-convex optimization), Wiener filtering [11] (linear filtering), method with aid of acceleration data [14] (aided by extra information), and hybrid methods [12] (complex architecture).

The formulation of transient artifact suppression as an optimization was considered in [22], [23]. However, only separable non-convex regularizers were considered in those works, which rules out the possibility of maintaining the convexity of the objective function to be minimized. In contrast, in this paper, we develop a more sophisticated type of non-convex regularizer that maintains the convexity of the objective function.

## II. PRELIMINARIES

Let  $D_k$  be the order- $k$  difference matrix. For example, the matrices  $D_1$  and  $D_2$  are given by

$$D_1 = \begin{bmatrix} -1 & 1 & & & \\ & & \ddots & \ddots & \\ & & & & -1 & 1 \end{bmatrix}, \quad (2)$$

and

$$D_2 = \begin{bmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & & \ddots & \ddots & \\ & & & & 1 & -2 & 1 \end{bmatrix}, \quad (3)$$

etc. The linear operator  $D_k$  is a discrete approximation of the  $k$ -order derivative. The frequency response of  $D_k$  is given by

$$D_k^f(\omega) = (1 - e^{-j\omega})^k = (2j e^{-j\frac{\omega}{2}} \sin(\frac{\omega}{2}))^k. \quad (4)$$

### A. The Fused Lasso Penalty

The estimation of a sparse piecewise constant signal  $x$  from its noisy observation  $y$ , known as the fused lasso signal estimation problem [26], can be formulated as

$$x^* = \arg \min_x \left\{ \frac{1}{2} \|y - x\|_2^2 + \lambda_1 \|D_1 x\|_1 + \lambda_0 \|x\|_1 \right\} \quad (5)$$

where  $\lambda_0 > 0$ ,  $\lambda_1 > 0$ , and  $D_1$  is the matrix (2). We define the fused lasso penalty as the function  $\varphi: \mathbb{R}^N \rightarrow \mathbb{R}$

$$\varphi(x) := \lambda_1 \|D_1 x\|_1 + \lambda_0 \|x\|_1 = \|Ax\|_1 \quad (6)$$

where  $A$  is a matrix of size  $(2N - 1) \times N$ , given by

$$A := \begin{bmatrix} \lambda_1 D_1 \\ \lambda_0 I \end{bmatrix}. \quad (7)$$

The matrix  $A$  absorbs the parameters  $\lambda_0, \lambda_1$ .

The solution to problem (5) is given by [8]

$$x^* = \text{soft}(\text{tvd}(y, \lambda_1), \lambda_0) \quad (8)$$

where  $\text{tvd}(\cdot, \lambda)$  is the solution to the total variation (TV) denoising problem ( $\lambda_0 = 0$  in (5)). TV denoising can be solved exactly in finite-time by fast solvers, e.g. [5]. In the context of convex analysis, TV denoising constitutes the proximal operator of  $\|D \cdot\|_1$ . Likewise, the proximal operator of the fused lasso penalty is given by (8).

### B. The Generalized Moreau Envelope

In this section, we define the generalized Moreau envelope of a convex function. This will be used to define the proposed non-convex penalty. For a description of the generalized Moreau envelope and its properties, see Refs. [13], [21].

We will use results from convex analysis [1]. We denote by  $\Gamma_0(\mathbb{R}^N)$  the set of proper lower semicontinuous convex function from  $\mathbb{R}^N$  to  $\mathbb{R} \cup \{+\infty\}$ .

The Moreau envelope of a convex function  $f \in \Gamma_0(\mathbb{R}^N)$ , denoted  $f^M: \mathbb{R}^N \rightarrow \mathbb{R}$ , is defined as

$$f^M(x) = \inf_{v \in \mathbb{R}^N} \left\{ \frac{1}{2} \|x - v\|_2^2 + f(v) \right\}. \quad (9)$$

Similarly, we define the generalized Moreau envelope (GME) of  $f$  as follows.

**Definition 1.** Let  $f \in \Gamma_0(\mathbb{R}^N)$  and  $B \in \mathbb{R}^{M \times N}$ . We define the generalized Moreau envelope  $f_B^M: \mathbb{R}^N \rightarrow \mathbb{R}$  as

$$f_B^M(x) = \inf_{v \in \mathbb{R}^N} \left\{ \frac{1}{2} \|B(x - v)\|_2^2 + f(v) \right\}. \quad (10)$$

The function is parameterized by matrix  $B$ .

The generalized Moreau envelope of a convex function  $f$  is differentiable, even if  $f$  itself is not.

**Lemma 1.** *The generalized Moreau envelope of  $f \in \Gamma_0(\mathbb{R}^N)$  is differentiable and its gradient is given by*

$$\nabla f_B^M(x) = B^T B \left( x - \arg \min_{v \in \mathbb{R}^N} \left\{ \frac{1}{2} \|B(x-v)\|_2^2 + f(v) \right\} \right). \quad (11)$$

*Proof.* Inasmuch as  $f$  is a convex function with unique critical point,  $f$  is coercive (Corollary 8.7.1 in [19]); and  $\|B \cdot\|_2^2$  is bounded below, therefore  $f_B^M$  is exact in  $\Gamma_0(\mathbb{R}^N)$  (Proposition 12.14 (ii) in [1]), that is

$$f_B^M(x) = \min_{v \in \mathbb{R}^N} \left\{ \frac{1}{2} \|B(x-v)\|_2^2 + f(v) \right\}.$$

Since  $\|B \cdot\|_2^2$  is Fréchet differentiable everywhere, by Proposition 18.7 in [1],  $f_B^M$  is Fréchet differentiable. The gradient is given by (Theorem 3.8 (e) in [25])

$$\begin{aligned} \nabla f_B^M(x) &= \nabla \left( \frac{1}{2} \|B(x-v)\|_2^2 \right) \\ &= B^T B(x-v). \end{aligned}$$

Because  $f_B^M$  is exact,  $v$  is achieved by

$$v = \arg \min_{v \in \mathbb{R}^N} \left\{ \frac{1}{2} \|B(x-v)\|_2^2 + f(v) \right\}.$$

This completes the proof.  $\square$

### III. TRANSIENT ARTIFACT SUPPRESSION

We formulate the problem of transient artifacts suppression (TAS) for signal model (1) as

$$\arg \min_{x, f \in \mathbb{R}^N} \left\{ \frac{1}{2} \|y - x - f\|_2^2 + \frac{\alpha}{2} \|D_k f\|_2^2 + \|Ax\|_1 \right\} \quad (12)$$

where  $D_k f$  is the order- $k$  difference operator and  $A$  is given by (7). The low-pass signal  $f$  is regularized using standard (quadratic) Tikhonov regularization. The transient artifact signal  $x$  is penalized using the  $\ell_1$  norm fused lasso penalty.

Since the objective function in (12) is quadratic with respect to  $f$ , the solution for  $f$  can be written in closed-form,

$$f^* = (I + \alpha D_k^T D_k)^{-1} (y - x) \quad (13)$$

where  $(I + \alpha D_k^T D_k)^{-1}$  represents a low-pass filter. The parameter  $\alpha > 0$  is related to the cut-off frequency  $f_c$  of the low-pass filter by

$$\alpha = \frac{1}{4^d \sin^{2d}(\pi f_c)}. \quad (14)$$

This relation is obtained by setting the frequency response equal to one half at the frequency  $\omega_c = 2\pi f_c$ .

We rewrite the objective function in (12) in terms of  $x$  alone by substituting  $f$  in (13). Then the objective function  $F: \mathbb{R}^N \rightarrow \mathbb{R}$  is given by

$$F(x) := \frac{1}{2} \|y - x - f\|_2^2 + \frac{\alpha}{2} \|D_k f\|_2^2 + \|Ax\|_1 \quad (15)$$

$$\begin{aligned} &= \frac{1}{2} \|(I - (I + \alpha D_k^T D_k)^{-1})(y - x)\|_2^2 \\ &\quad + \frac{\alpha}{2} \|D_k (I + \alpha D_k^T D_k)^{-1}(y - x)\|_2^2 + \|Ax\|_1. \end{aligned} \quad (16)$$

The objective function  $F$  can be written more simply as

$$F(x) = \frac{1}{2} \|H(y - x)\|_2^2 + \|Ax\|_1 \quad (17)$$

where

$$H^T H := \alpha (I + \alpha D_k^T D_k)^{-1} D_k^T D_k. \quad (18)$$

It turns out we will not need  $H$  itself. It will be sufficient to work with  $H^T H$ .

The implementation of the forward backward splitting algorithm for minimizing the objective function  $F$  requires the Lipschitz constant of the quadratic term in (17), given by the maximum eigenvalue of  $H^T H$ .

**Proposition 1.** *The maximum eigenvalue of the linear operator  $H^T H$  in (18) is upper bounded by*

$$\rho = (1 + (\alpha 4^d)^{-1})^{-1}. \quad (19)$$

*Proof.* The linear operator  $H^T H$  is the matrix form of an LTI filter, thus its maximum eigenvalue is upper bounded by the maximum value of its frequency response,

$$\begin{aligned} (H^T H)^f(\omega) &= \frac{(D_k^T D_k)^f(\omega)}{1/\alpha + (D_k^T D_k)^f(\omega)} \\ &= \frac{[4 \sin^2(\omega/2)]^d}{1/\alpha + [4 \sin^2(\omega/2)]^d} \end{aligned}$$

Setting the derivative to zero, we obtain the value in (19).  $\square$

### IV. THE GENERALIZED FUSED LASSO PENALTY

In this section, we propose a non-convex generalization of the fused lasso penalty. We first define a special case of the generalized Moreau envelope, with  $f(v) = \|Av\|_1$  in (10).

**Definition 2.** *We define  $S_B: \mathbb{R}^N \rightarrow \mathbb{R}$  to be the generalized Moreau envelope of the fused lasso penalty,*

$$S_B(x) := \min_{v \in \mathbb{R}^N} \left\{ \frac{1}{2} \|B(x-v)\|_2^2 + \|Av\|_1 \right\} \quad (20)$$

where  $A$  is given by (7). The function is parameterized by matrix  $B \in \mathbb{R}^{M \times N}$ .

As it is a generalized Moreau envelope, the function  $S_B$  has several properties such as convexity, exactness, and differentiability. We now define the non-convex generalization of the fused lasso penalty.

**Definition 3.** *We define the generalized fused lasso penalty  $\psi_B: \mathbb{R}^N \rightarrow \mathbb{R}$  as*

$$\psi_B(x) := \|Ax\|_1 - S_B(x) \quad (21)$$

where  $S_B$  is given by (20). The function is parameterized by matrix  $B \in \mathbb{R}^{M \times N}$ .

The generalized fused lasso penalty generalizes the generalized minimax concave penalty [21], and enjoys similar properties.

## V. TRANSIENT ARTIFACT SUPPRESSION USING THE GENERALIZED FUSED LASSO PENALTY

In this section, we formulate the problem of suppressing transient artifacts, using the non-convex generalized fused lasso penalty. We provide a condition to ensure the convexity of the objective function to be minimized. We also derive an algorithm for its minimization, using forward backward splitting.

We formulate the transient artifact suppression (TAS) problem as

$$x^* = \arg \min_{x \in \mathbb{R}^N} \left\{ J(x) = \frac{1}{2} \|H(y-x)\|_2^2 + \psi_B(x) \right\} \quad (22)$$

where  $H$  is given by (18) and  $\psi_B$  is the generalized fused lasso penalty defined in (21). In this formulation, we replace the  $\ell_1$  norm fused lasso penalty in (17) with the generalized fused lasso penalty defined in (21).

**Theorem 4.** Let  $y \in \mathbb{R}^N$  and  $H \in \mathbb{R}^{(N-1) \times N}$ . Define  $J: \mathbb{R}^N \rightarrow \mathbb{R}$  as

$$J(x) = \frac{1}{2} \|H(y-x)\|_2^2 + \psi_B(x) \quad (23)$$

where  $\psi_B$  is the generalized fused lasso penalty (21). If

$$B^T B \preceq H^T H, \quad (24)$$

then  $J$  is a convex function.

*Proof.* We write  $J(x)$  as

$$\begin{aligned} J(x) &= \frac{1}{2} \|H(y-x)\|_2^2 + \|Ax\|_1 \\ &\quad - \min_{v \in \mathbb{R}^N} \left\{ \|Av\|_1 + \frac{1}{2} \|B(x-v)\|_2^2 \right\} \\ &= \frac{1}{2} \|H(y-x)\|_2^2 + \|Ax\|_1 - \frac{1}{2} \|Bx\|_2^2 \\ &\quad - \min_{v \in \mathbb{R}^N} \left\{ \|Av\|_1 + \frac{1}{2} \|Bv\|_2^2 - v^T B^T B x \right\} \\ &= \frac{1}{2} x^T (H^T H - B^T B) x + \frac{1}{2} \|Hy\|_2^2 - y^T H^T H x + \|Ax\|_1 \\ &\quad + \max_{v \in \mathbb{R}^N} \left\{ -\|Av\|_1 - \frac{1}{2} \|Bv\|_2^2 + v^T B^T B x \right\} \end{aligned}$$

Consider the final expression for  $J$ . The first term is convex if  $H^T H - B^T B$  is positive semidefinite. The expression inside the curly braces is affine in  $x$ , hence convex in  $x$ . Therefore, the entire last term is convex in  $x$ , because the maximum of a set of convex functions (here indexed by  $v$ ) is convex (Proposition 8.14 in [1]). The remaining terms are convex in  $x$ .  $\square$

The convexity condition can be easily satisfied. A simple choice for  $B^T B$  is

$$B^T B = \gamma H^T H, \quad 0 \leq \gamma \leq 1. \quad (25)$$

Then (24) is satisfied. With  $B^T B$  chosen in this way, the parameter  $\gamma$  controls the non-convexity of  $\psi_B$ . As in the case of  $H$ , we will not need  $B$  itself. The objective function and its minimization depends only on  $B^T B$  and not  $B$  itself.

The following iterative algorithm minimizes the convex function  $J$ . It is derived as an application of the forward-backward splitting (FBS) algorithm.

**Proposition 2.** Let  $y \in \mathbb{R}^N$ ,  $\alpha > 0$ , and  $H^T H$  be given by (18). Set  $0 \leq \gamma \leq 1$  and  $B^T B = \gamma H^T H$ . Then the iteration

$$v^{(i)} = \arg \min_{v \in \mathbb{R}^N} \left\{ \frac{\gamma}{2} \|H(x^{(i)} - v)\|_2^2 + \|Av\|_1 \right\} \quad (26a)$$

$$z^{(i)} = H^T H (x^{(i)} - y - \gamma(x^{(i)} - v^{(i)})) \quad (26b)$$

$$x^{(i+1)} = \text{soft}(\text{tvd}(x^{(i)} - z^{(i)}, \lambda_1), \lambda_0) \quad (26c)$$

converges to the minimizer of  $J$  in (23).

*Proof.* We write  $J$  as the sum of two convex functions,

$$J(x) = f_1(x) + f_2(x)$$

where

$$f_1(x) = \frac{1}{2} \|H(y-x)\|_2^2 - S_B(x),$$

$$f_2(x) = \|Ax\|_1.$$

Since both  $f_1$  and  $f_2$  are convex and additionally  $\nabla f_1$  is  $\rho$ -Lipschitz continuous, the FBS algorithm, i.e.,

$$x^{(i+1)} = \text{prox}_{\mu f_2}(x^{(i)} - \mu \nabla f_1(x^{(i)})),$$

converges to the minimizer of  $J$  [1]. The parameter  $\mu$  should be set such that  $0 \leq \mu \leq 2/\rho$  where  $\rho$  is given by (19). In particular, we can set  $\mu = 1$  because  $\rho > 1$  in (19). It remains to determine the gradient of  $f_1$  and the proximal operator of  $f_2$ . The gradient of  $f_1$  is given by

$$\nabla f_1(x) = H^T H (x - y) - \nabla S_B(x). \quad (27)$$

Using (11), we have

$$\nabla S_B(x) = B^T B (x - \arg \min_{v \in \mathbb{R}^N} \left\{ \frac{1}{2} \|B(x-v)\|_2^2 + \|Av\|_1 \right\}).$$

Using (25), we have

$$\nabla S_B(x) = \gamma H^T H (x - \arg \min_{v \in \mathbb{R}^N} \left\{ \frac{\gamma}{2} \|H(x-v)\|_2^2 + \|Av\|_1 \right\}).$$

Hence,  $\nabla f_1(x^{(i)})$  is computed in (26b). From (8), the proximal operator of  $f_2$  is computed in (26c).  $\square$

The minimization in (26a) can be solved by FBS (i.e., ISTA) or other algorithms (e.g., FISTA, FASTA).

## VI. NUMERICAL EXAMPLES

We show results on two numerical examples to demonstrate the performance of the proposed method: one on synthetic data, and one on near infrared spectroscopic (NIRS) data.

**Example 1.** We apply the proposed method to the synthetic data illustrated in Fig. 2(a). The noisy signal consists of two low-frequency sinusoids, several piecewise constant transients, and additive white Gaussian noise (standard deviation  $\sigma = 0.5$ ). We show the solutions using both the  $\ell_1$  norm fused lasso penalty and the proposed non-convex generalized fused lasso penalty (both of which are formulated as convex optimization problems).

We set  $k = 2$  in (18) and  $\alpha$  using (14) with  $f_c = 0.033$ . We set  $\lambda_0 = \beta_0 \|H^T H\|_2 \sigma$  and  $\lambda_1 = \beta_1 \|H^T H\|_2 \sqrt{N} \sigma$  with  $\beta_0 = 0.1$  and  $\beta_1 = 0.25$ . Such a parameter-setting strategy is a variation of 3-sigma rule [18]. Additionally, for the generalized

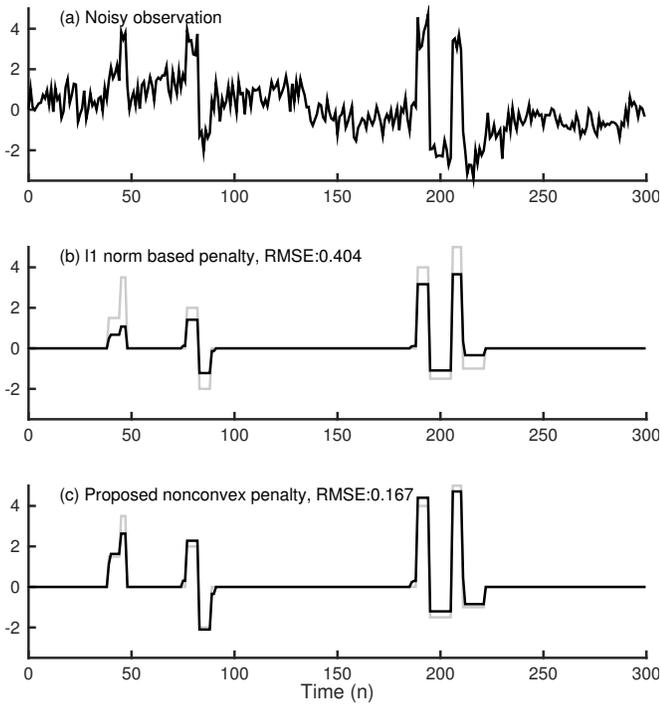


Fig. 2. Example 1. (a) Noisy observation. Transient artifacts as estimated using (b) the  $\ell_1$ -norm and (c) the proposed non-convex generalized penalty.

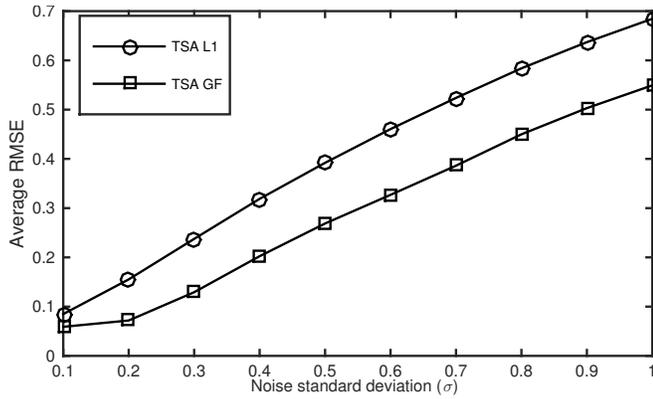


Fig. 3. Example 1. RMSE as a function of noise level.

fused lasso penalty, we set  $\gamma$ , the index of non-convexity, to  $\gamma = 0.6$ . The proposed method is implemented using algorithm (26).

Figure 2 shows the estimated transient artifacts of the two aforementioned methods, compared with the ground truth. As shown in Fig. 2(b), the  $\ell_1$  norm regularizer systematically underestimates the true signal values. The non-convex generalized fused lasso penalty estimates the signal values more accurately (Fig. 2(c)). The improvement attained by the generalized penalty is also reflected in the lower RMSE value.

To further compare the estimation accuracy of the two methods, we calculate the average RMSE as a function of the noise level. We let the noise standard deviation span the interval  $0.1 \leq \sigma \leq 1.0$ . For each value, we calculate the average RMSE of 20 noise realizations with the best parameter

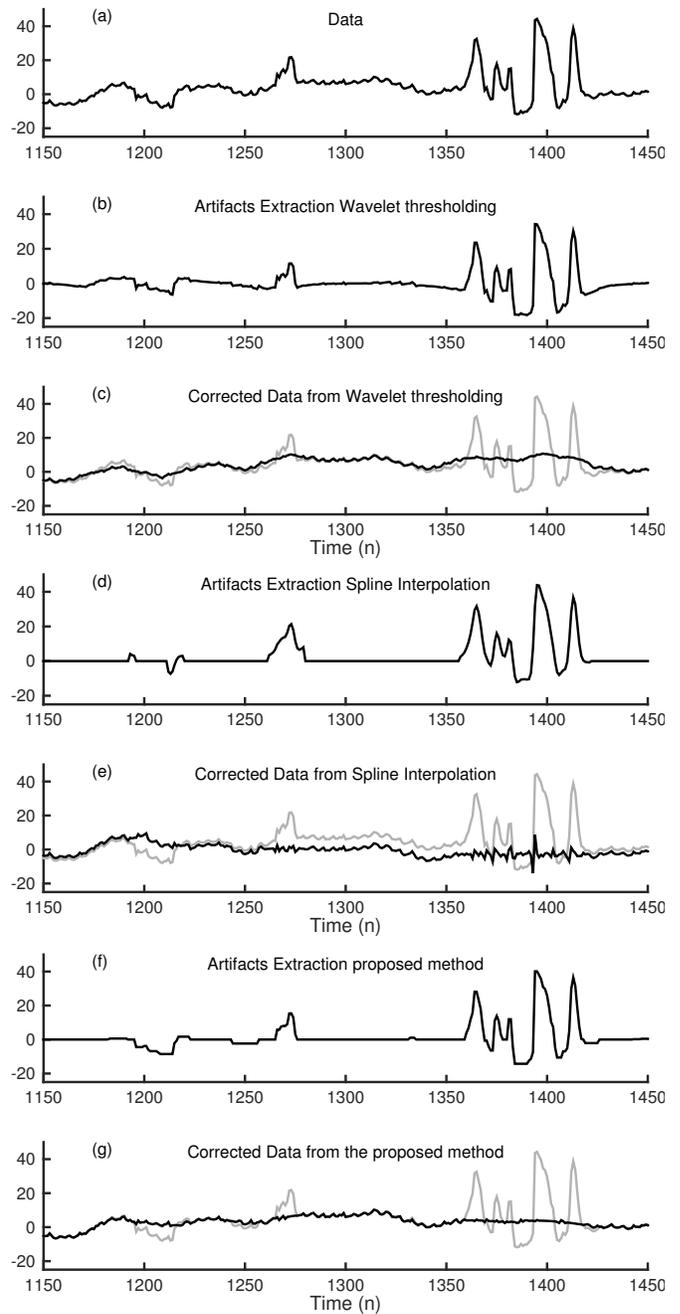


Fig. 4. Example 2. NIRS data artifact suppression. (a) Original data. (b) Transient artifacts estimated using wavelet thresholding. (c) Corrected data using (b). (d) Transient artifacts estimated using spline interpolation. (e) Corrected data using (d). (f) Transient artifacts estimated using proposed method. (g) Corrected data using (f).

settings. Fig. 3 shows that the proposed method yields the lower average RMSE of the two methods.

**Example 2.** In this example, we apply the method to real near infrared spectroscopic (NIRS) time series data, and compare it to wavelet thresholding and spline interpolation artifact-suppression methods. The data shown in Fig. 4(a) were acquired using a pair of optodes (one source and one detector) on the subject's forehead near the left eye. In this

case, it is susceptible to artifacts due to eye blinks (in addition to other artifacts ordinarily present) which are of variable amplitude, width, and shape. The time series has a length of 1900 samples. We show a segment of 300 samples to better present its detail.

For the wavelet thresholding artifact-suppression method [15], we use the undecimated Haar wavelet transform the non-negative garrote threshold function. The artifacts are estimated by the reconstruction of the all thresholded subbands (the low-pass subband is set to zero). For the spline interpolation artifact suppression method [20], we first identify the segments containing the transient artifacts using the moving standard deviation: the transient artifacts surge in short time periods, segments containing artifacts have large standard deviation. Then, each identified segment is modeled using cubic spline interpolation, and use the strategy presented in Ref. [20] is used to reconstruct the complete signal. For the proposed method with the non-convex generalized fused lasso penalty, we set the  $\lambda$  parameters as in Example 1 and we set  $\gamma = 0.8$ . The corrected data is obtained by subtracting the estimated artifact signal from the noisy data.

The corrected data is illustrated in Figs. 4(c), 4(e), and 4(g). The result using the proposed method preserves the baseline of the original data. In contrast, the wavelet thresholding method changes the baseline in the neighborhood of artifacts. The spline interpolation method also substantially alters the baseline. The result using the proposed method, shown in Fig. 4(g), exhibits less distortion of the baseline than the other two methods, since it estimates the amplitude of the transient artifacts more accurately than the other two methods shown in Fig. 4(b) and 4(d).

## VII. CONCLUSION

For the purpose of suppressing transient artifacts in biomedical time series data, we propose a non-convex generalized fused lasso penalty for the estimation of signals comprising a low-pass signal, a sparse piecewise constant signal, and additive white Gaussian noise. The proposed non-convex penalty is designed so as to preserve the convexity of the total cost function to be minimized, thereby realizing the benefits of a convex optimization framework (reliable, robust algorithms, etc.). The benefit of the proposed non-convex penalty, relative to the classical  $\ell_1$  norm penalty, is that it overcomes the tendency of the  $\ell_1$  norm to underestimate the true amplitude of signal values. We apply the proposed method to the suppression of artifacts in near infrared spectroscopic (NIRS).

## VIII. ACKNOWLEDGMENT

The authors thank Randall Barbour for important discussions. This work was supported by NSF (grant CCF-1525398).

## REFERENCES

- [1] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2011, vol. 408.
- [2] I. Bayram, P.-Y. Chen, and I. Selesnick, "Fused lasso with a non-convex sparsity inducing penalty," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, May 2014.

- [3] A. Blake and A. Zisserman, *Visual Reconstruction*. MIT Press, 1987.
- [4] M. E. Calkins, J. Katsanis, M. A. Hammer, and W. G. Iacono, "The misclassification of blinks as saccades: Implications for investigations of eye movement dysfunction in schizophrenia," *Psychophysiology*, vol. 38, no. 5, pp. 761–767, 2001.
- [5] L. Condat, "A direct algorithm for 1-d total variation denoising," *IEEE Signal Processing Letters*, vol. 20, no. 11, pp. 1054–1057, 2013.
- [6] R. Cooper, J. Selb, L. Gagnon, D. Phillip, H. W. Schytz, H. K. Iversen, M. Ashina, and D. A. Boas, "A systematic comparison of motion artifact correction techniques for functional near-infrared spectroscopy," *Frontiers in neuroscience*, vol. 6, no. 147, 2012.
- [7] C. Damon, A. Liutkus, A. Gramfort, and S. Essid, "Non-negative matrix factorization for single-channel EEG artifact rejection," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1177–1181, 2013.
- [8] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, "Pathwise coordinate optimization," *The Annals of Applied Statistics*, vol. 1, no. 2, pp. 302–332, 2007.
- [9] H. L. Graber, Y. Xu, and R. L. Barbour, "Optomechanical imaging system for breast cancer detection," *Journal of the Optical Society of America A*, vol. 28, no. 12, pp. 2473–2493, 2011.
- [10] M. K. Islam, A. Rastegarnia, A. T. Nguyen, and Z. Yang, "Artifact characterization and removal for in vivo neural recording," *Journal of Neuroscience Methods*, vol. 226, pp. 110–123, 2014.
- [11] M. Izzetoglu, A. Devaraj, S. Bunce, and B. Onaral, "Motion artifact cancellation in NIR spectroscopy using Wiener filtering," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 5, pp. 934–938, 2005.
- [12] S. Jahani, S. K. Setarehdan, D. A. Boas, and M. A. Yücel, "Motion artifact detection and correction in functional near-infrared spectroscopy: a new hybrid method based on spline interpolation method and Savitzky–Golay filtering," *Neurophotonics*, vol. 5, no. 1, p. 015003, 2018.
- [13] A. Lanza, S. Morigi, I. Selesnick, and F. Sgallari, "Sparsity-inducing non-convex non-separable regularization for convex image processing," 2017, preprint.
- [14] A. J. Metz, M. Wolf, P. Achermann, and F. X. Scholkmann, "A new approach for automatic removal of movement artifacts in near-infrared spectroscopy time series by means of acceleration data," *Algorithms*, vol. 8, no. 4, pp. 1052–1075, 2015.
- [15] B. Molavi and G. A. Dumont, "Wavelet-based motion artifact removal for functional near-infrared spectroscopy," *Physiological measurement*, vol. 33, no. 2, pp. 259–270, 2012.
- [16] M. K. I. Molla, M. R. Islam, T. Tanaka, and T. M. Rutkowski, "Artifact suppression from EEG signals using data adaptive time domain filtering," *Neurocomputing*, vol. 97, pp. 297–308, 2012.
- [17] M. Nikolova, "Energy minimization methods," in *Handbook of mathematical methods in imaging*. Springer, 2011, pp. 139–185.
- [18] A. Parekh and I. W. Selesnick, "Convex fused lasso denoising with non-convex regularization and its use for pulse detection," *IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pp. 1–6, 2015.
- [19] R. T. Rockafellar, *Convex analysis*. Princeton university press, 1972.
- [20] F. Scholkmann, S. Spichtig, T. Muehlemann, and M. Wolf, "How to detect and reduce movement artifacts in near-infrared imaging using moving standard deviation and spline interpolation," *Physiological measurement*, vol. 31, no. 5, pp. 649–662, 2010.
- [21] I. Selesnick, "Sparse regularization via convex analysis," *IEEE Transactions on Signal Processing*, vol. 65, no. 17, pp. 4481–4494, 2017.
- [22] I. W. Selesnick, H. L. Graber, Y. Ding, T. Zhang, and R. L. Barbour, "Transient artifact reduction algorithm (TARA) based on sparse optimization," *IEEE Transactions on Signal Processing*, vol. 62, no. 24, pp. 6596–6611, 2014.
- [23] I. W. Selesnick, H. L. Graber, D. S. Pfeil, and R. L. Barbour, "Simultaneous low-pass filtering and total variation denoising," *IEEE Transactions on Signal Processing*, vol. 62, no. 5, pp. 1109–1124, 2014.
- [24] J.-L. Starck, D. Donoho, and M. Elad, "Redundant multiscale transforms and their application for morphological component separation," *Advances in Imaging and Electron Physics*, vol. 132, pp. 287–348, 2004.
- [25] S. Thomas, "The operation of infimal convolution," Ph.D. dissertation, University of Lund, 1996.
- [26] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 1, pp. 91–108, 2005.