



# Sparsity-based correction of exponential artifacts



Yin Ding\*, Ivan W. Selesnick

School of Engineering, New York University, 6 Metrotech Center, Brooklyn, NY 11201, USA

## ARTICLE INFO

### Article history:

Received 26 June 2015

Received in revised form

17 September 2015

Accepted 18 September 2015

Available online 30 September 2015

### Keywords:

Sparsity

Signal decomposition

Artifact removal

## ABSTRACT

This paper describes an exponential transient excision algorithm (ETEA). In biomedical time series analysis, e.g., *in vivo* neural recording and electrocorticography (ECoG), some measurement artifacts take the form of piecewise exponential transients. The proposed method is formulated as an unconstrained convex optimization problem, regularized by smoothed  $\ell_1$ -norm penalty function, which can be solved by majorization–minimization (MM) method. With a slight modification of the regularizer, ETEA can also suppress more irregular piecewise smooth artifacts, especially, ocular artifacts (OA) in electroencephalography (EEG) data. Examples of synthetic signal, EEG data, and ECoG data are presented to illustrate the proposed algorithms.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

This work is motivated by the problem of suppressing various types of artifacts in recordings of neural activity. In a recent study [25], typical artifacts in *in vivo* neural recordings are classified into four types (Type 0–3, see Section 2.2 and Figure 3 in [25]). This classification covers many artifacts in the scope of human brain activity recordings, e.g., electroencephalography (EEG) and electrocorticography (ECoG). In this paper, we consider the suppression of Type 0 and Type 1 artifacts. For the purpose of flexibility and generality, we redefine them in terms of morphological characteristics:

- Type 0: a smooth protuberance that can be modeled as  $\hat{x}(t) = te^{-\alpha t}$ , when  $t \geq t_0$ .
- Type 1: an abrupt jump followed by an exponential decay that can be modeled as  $\hat{x}(t) = e^{-\alpha t}$ , when  $t \geq t_0$ .

Fig. 1 shows examples of the two types of artifacts. We do not consider the other two types in this work because our previous works have addressed efficient algorithms to

remove such artifacts. For instance, low-pass filtering/total variation denoising (LPF/TVD) [61] suppresses Type 2 artifacts (Fig. 1(c)), and lowpass filtering/compound sparse denoising (LPF/CSD) [60,61] can remove sparse and blocky spikes (Type 3 shown in Fig. 1(d)).

The approach proposed in this paper is based on an optimization problem intended to capture the primary morphological characteristics of the artifacts using sparsity-inducing regularization. To formulate the problem, we model the observed time series as

$$y(t) = f(t) + x(t) + w(t), \quad (1)$$

where  $f$  is a lowpass signal,  $x$  is a piecewise smooth transient signal (i.e., Type 0 or Type 1 artifacts), and  $w$  is stationary white Gaussian noise. More specifically,  $f$  is assumed to be restricted to a certain range of low frequencies. In other words,  $\mathbf{H}(f) \approx \mathbf{0}$ , where  $\mathbf{H}$  is a high-pass filter. Note that in the signal model (1), conventional LTI filtering is not suitable to estimate either  $f$  or  $x$  from  $y$ , because component  $x$ , as a piecewise smooth signal comprised of transients, is not band limited.

In order to estimate the components, we combine LTI filtering with sparsity-based techniques. We formulate an optimization problem for both decomposition and denoising. A computationally efficient algorithm is derived to

\* Corresponding author.

E-mail addresses: [yd372@nyu.edu](mailto:yd372@nyu.edu) (Y. Ding), [selesi@nyu.edu](mailto:selesi@nyu.edu) (I.W. Selesnick).

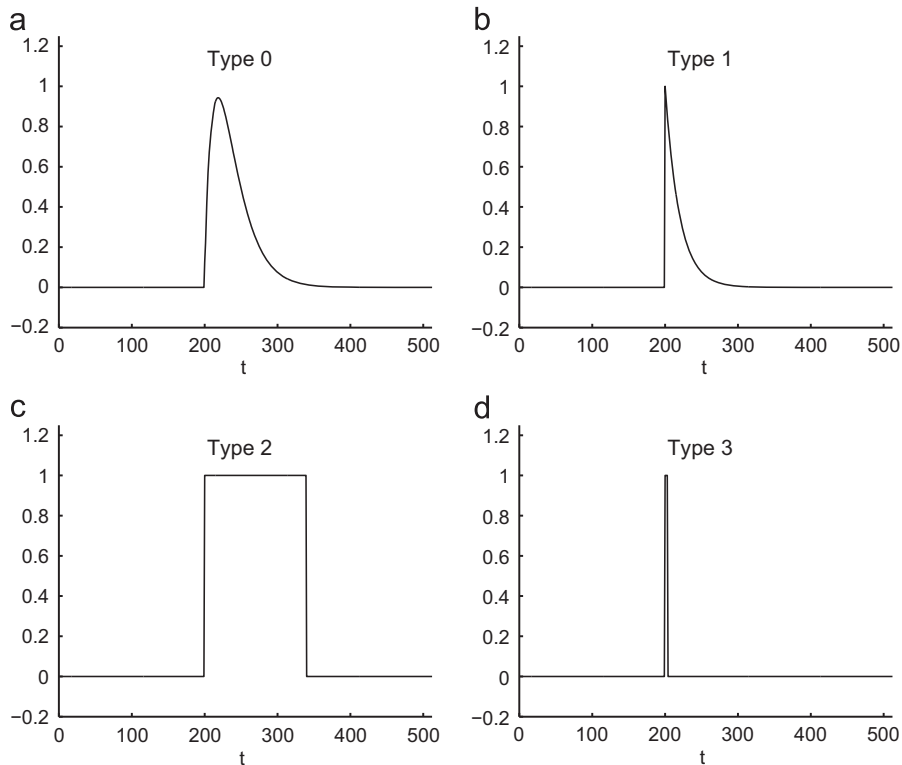


Fig. 1. Examples of (a) Type 0 artifact, and (b) Type 1 artifact, and (c) Type 2 artifact, and (d) Type 3 artifact.

solve the optimization problem, based on the theory of majorization–minimization (MM) [15,24,36].

In addition, this paper specifies how to generate a smoothed penalty function and its majorizer from a non-smooth one, in order to overcome a numerical issue that arises when the penalty function is not differentiable.

### 1.1. Related works

Some recent works recover signals with transients by various algorithms. In [19], a slowly varying signal is modeled as a local polynomial and an optimization problem using Tikhonov regularization is formulated to capture it. In [47], the slowly varying trend is modeled as a higher-order sparse-derivative signal (e.g., the third-order derivative is sparse).

Instead of estimating the slowly varying component via regularization, the LPF/TVD method [61] estimates a lowpass component by LTI filtering and a piecewise constant component by optimization. In this case, an optimization problem is formulated to estimate the piecewise constant component. The approach proposed here uses a similar technique to recover the lowpass component, but in contrast to LPF/TVD, it is more general—the regularization is more flexible with a tunable parameter, so that LPF/TVD can be considered as a special case.

Another algorithm related to the approach taken in this paper is the transient artifact reduction algorithm (TARA) [60] which is utilized to suppress additive piecewise constant artifacts and spikes (similar to a hybrid of Type 2 and Type 3 artifacts). The approaches proposed in this work

target different types of artifacts (Type 0 and Type 1) and applied in different applications.

The irregularity of Type 0 transients leads to a more complicated artifact removal problem, where the artifact are irregular fluctuations. A typical example in EEG are ocular artifacts (OA) caused by the blink and/or movement of eyes. To suppress OA, there are approaches based on empirical mode decomposition (EMD) [41–43,69], and on independent component analysis (ICA) methods [1,12,20,26,37,48]. The concept of spatial-frequency in acoustic analysis is also used to remove OA from multichannel signals [44,64]. In this work, we present a new method to suppress ocular artifacts by proposing a specific model and using sparse optimization.

This paper adopts a regularizer inspired by the generalized 1-D total variation [27], wherein the derivative operator in conventional total variation regularizer is generalized to a recursive filter. The regularizers adopted in ETEA and second-order ETEA coincide with first-order and second-order cases of generalized 1-D total variation, respectively. Some differences to the problem discussed in [27] are as follows. Firstly, the signal model (1) allows a lowpass baseline as a component, hence, ETEA can be seen as a combination of conventional LTI filtering and generalized 1-D total variation. Secondly, we consider a formulation in terms of banded matrices, for computational efficiency. Thirdly, we give optimality conditions of the proposed problems, and use these conditions as a guide to set the regularization parameters.

## 2. Preliminaries

### 2.1. Notation

We use bold uppercase letters for matrices, e.g.,  $\mathbf{A}$  and  $\mathbf{B}$ , and bold lowercase letters for vectors, e.g.,  $\mathbf{x}$  and  $\mathbf{y}$ . We use column vectors for one-dimensional series. For example, a vector  $\mathbf{x} \in \mathbb{R}^N$  is written as

$$\mathbf{x} = [x(0), x(1), \dots, x(N-1)]^T \quad (2)$$

where  $[\cdot]^T$  denotes matrix transpose. The  $\ell_1$ -norm and squared  $\ell_2$ -norm of  $\mathbf{x}$  are defined as

$$\|\mathbf{x}\|_1 := \sum_n |x(n)|, \quad \|\mathbf{x}\|_2^2 := \sum_n x^2(n). \quad (3)$$

The inverse transpose of a matrix is denoted as  $\mathbf{A}^{-T}$ . The first-order difference operator  $\mathbf{D}$  of size  $(N-1) \times N$  is

$$\mathbf{D} := \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}. \quad (4)$$

We use  $f(x; a)$  to denote a function of  $x$  determined by parameter  $a$ , to distinguish it from a function with two ordered arguments, e.g.,  $f(x, y)$ .

### 2.2. LTI filter with matrix formulation

A discrete-time filter can be described as

$$\sum_k a_k y(n-k) = \sum_k b_k x(n-k), \quad (5)$$

with transfer function  $H(z) = B(z)/A(z)$ . For a finite length signal, (5) can be rewritten as

$$\mathbf{A}\mathbf{y} = \mathbf{B}\mathbf{x}, \quad (6)$$

where  $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{N \times N}$  are both banded matrices. When  $\mathbf{H}$  is a zero-phase filter with  $a_{-k} = a_k$  and  $b_{-k} = b_k$ , then  $\mathbf{A}$  and  $\mathbf{B}$  are both symmetric Toeplitz matrices

$$\mathbf{A} = \begin{bmatrix} a_1 & a_0 & & & \\ a_0 & a_1 & a_0 & & \\ & \ddots & \ddots & \ddots & \\ & & a_0 & a_1 & a_0 \\ & & & a_0 & a_1 \end{bmatrix}, \quad (7a)$$

**Table 1**  
Sparsity-promoting penalty functions ( $a > 0$ ).

Penalty	$\phi(u)$	$\phi_\epsilon(u)$	$\psi(u) = u/\phi'_\epsilon(u)$
abs	$ u $	$\sqrt{u^2 + \epsilon}$	$\sqrt{u^2 + \epsilon}$
log	$\frac{1}{a} \log(1 + a u )$	$\frac{1}{a} \log(1 + a\sqrt{u^2 + \epsilon})$	$\sqrt{u^2 + \epsilon}(1 + a\sqrt{u^2 + \epsilon})$
atan	$\frac{2}{a\sqrt{3}} \left( \tan^{-1} \left( \frac{1+2a u }{\sqrt{3}} \right) - \frac{\pi}{6} \right)$	$\frac{2}{a\sqrt{3}} \left( \tan^{-1} \left( \frac{1+2a\sqrt{u^2 + \epsilon}}{\sqrt{3}} \right) - \frac{\pi}{6} \right)$	$\sqrt{u^2 + \epsilon} (1 + a\sqrt{u^2 + \epsilon} + a^2(u^2 + \epsilon))$

$$\mathbf{B} = \begin{bmatrix} b_1 & b_0 & & & \\ b_0 & b_1 & b_0 & & \\ & \ddots & \ddots & \ddots & \\ & & b_0 & b_1 & b_0 \\ & & & b_0 & b_1 \end{bmatrix}. \quad (7b)$$

In this case, disregarding a few samples on both edges,  $\mathbf{y}$  can be re-written as

$$\mathbf{y} = \mathbf{B}\mathbf{A}^{-1}\mathbf{x} = \mathbf{H}\mathbf{x}. \quad (8)$$

A detailed description of such zero-phase filters is given in Section V of Ref. [61], where two parameters,  $d$  and  $f_c$ , determine the order of the filter and the cut-off frequency, respectively. In this paper,  $\mathbf{B}$  is a square matrix, in contrast to [61]. The simplest case ( $d=1$ ) is given by (7).

### 2.3. Majorization–minimization

Majorization–minimization (MM) [15,35,57] is a procedure to replace a difficult optimization problem by a sequence of simpler ones [15,24,36,57]. Suppose a minimization problem has an objective function  $J: \mathbb{R}^N \rightarrow \mathbb{R}$ , then the majorizer  $G(\mathbf{u}, \mathbf{v}): \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$  satisfies

$$G(\mathbf{u}, \mathbf{v}) = J(\mathbf{u}), \quad \text{when } \mathbf{u} = \mathbf{v}, \quad (9a)$$

$$G(\mathbf{u}, \mathbf{v}) \geq J(\mathbf{u}), \quad \text{when } \mathbf{u} \neq \mathbf{v}. \quad (9b)$$

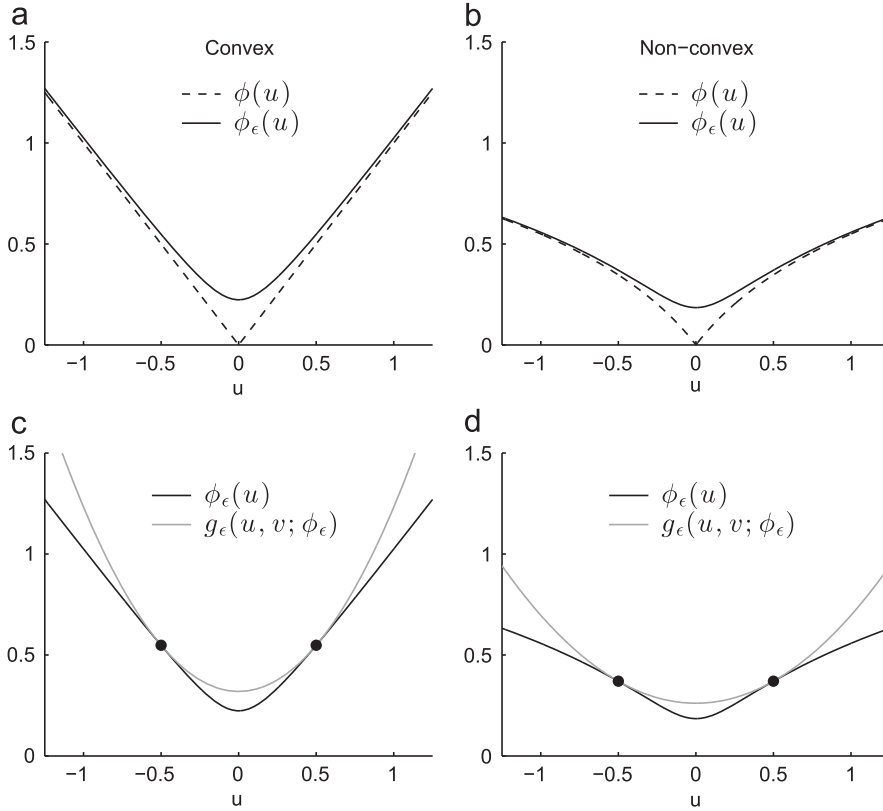
Then MM iteratively solves

$$\mathbf{u}^{(k+1)} = \arg \min_{\mathbf{u}} G(\mathbf{u}, \mathbf{u}^{(k)}) \quad (10)$$

until convergence, where  $k$  is iteration index. The detailed derivation and proof of convergence of MM are given in [35].

## 3. Majorization of smoothed penalty function

When using sparse-regularized optimization to recover a signal, the  $\ell_1$ -norm is widely utilized. To further enhance sparsity, some methods use iteratively re-weighting procedures [4,34,45,63], or a non-convex pseudo  $\ell_p$ -norm ( $0 < p < 1$ ), or a mixed-norm in the regularizer [5,68,66,58,33]. Other non-convex penalty functions can also be used. In [46] a logarithm penalty function is discussed ('log' function in Table 1), and in [59], an arctangent function is used ('atan' function Table 1). However, each of these functions is non-differentiable at zero, where  $\phi'(0^-) \neq \phi'(0^+)$ . A method, to avoid the discontinuity in the derivative, is to smooth the function.



**Fig. 2.** (a) Convex non-smooth penalty function and corresponding smoothed penalty function. (b) Non-convex and smoothed penalty functions. (c) Majorizer (gray) of the smoothed penalty function in (a). (d) Majorizer (gray) of the smoothed non-convex penalty function in (b). For all the figures we set  $\epsilon = 0.05, v = 0.5, a = 2$ .

Consider the smoothed function  $\phi_\epsilon: \mathbb{R} \rightarrow \mathbb{R}_+$

$$\phi_\epsilon(u) := \phi(\rho(u; \epsilon)), \tag{11}$$

where function  $\rho: \mathbb{R} \rightarrow \mathbb{R}_+$  is defined as

$$\rho(u; \epsilon) := \sqrt{u^2 + \epsilon}, \quad \epsilon > 0. \tag{12}$$

Because  $\sqrt{u^2 + \epsilon}$  is always greater than zero,  $\phi_\epsilon$  has a continuous first-order derivative

$$\phi'_\epsilon(u) = \frac{u}{\sqrt{u^2 + \epsilon}} \phi'(\sqrt{u^2 + \epsilon}). \tag{13}$$

Table 1 gives the smoothed penalty functions  $\phi_\epsilon$  corresponding to  $\phi$  in the first column, and Fig. 2(a) and (b) illustrates the penalty function and its smoothed version. The parameter  $\epsilon$  controls the similarity of  $\phi_\epsilon$  to non-smooth function  $\phi$ . When  $\epsilon \rightarrow 0$ ,  $\phi_\epsilon(u) \approx \phi(u)$ . In practice, we set  $\epsilon$  to a very small number (e.g.,  $10^{-10}$ ).

### 3.1. Majorization of smoothed penalty function

We assume the non-smooth function  $\phi$  satisfies the following conditions:

1.  $\phi(u)$  is continuous on  $\mathbb{R}$ .
2.  $\phi(u)$  is twice continuously differentiable on  $\mathbb{R} \setminus \{0\}$ .
3.  $\phi(u)$  is even symmetric:  $\phi(u) = \phi(-u)$ .
4.  $\phi(u)$  is increasing and concave on  $\mathbb{R}_+$ .

Under such assumptions, we find a quadratic function  $g: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ ,

$$g(u, v; \phi) = \frac{\phi'(v)}{2v} u^2 + \phi(v) - \frac{v}{2} \phi'(v), \tag{14}$$

majorizing  $\phi$  when  $v \neq 0$  [10, Lemma 1]. However,  $\phi$  might not be differentiable at zero, e.g., all functions  $\phi$  in Table 1. To avoid this issue, we use (11) to ensure that the objective function is continuous and differentiable. Then the MM method can be applied without the occurrence of numerical issues (e.g., divide by zero). The following proposition indicates a suitable majorizer of the smoothed penalty functions.

**Proposition 1.** *If function  $\phi$  satisfies the listed four conditions in the previous paragraph and  $g$  in (14) is a majorizer of  $\phi$ , then*

$$g_\epsilon(u, v; \phi_\epsilon) = g(\rho(u; \epsilon), \rho(v; \epsilon); \phi), \tag{15}$$

*is majorizer of  $\phi_\epsilon(u) = \phi(\rho(u; \epsilon))$  for  $u, v \in \mathbb{R}$ , and if writing explicitly, then  $g_\epsilon$  in (15) is*

$$g_\epsilon(u, v; \phi_\epsilon) = \frac{u^2}{2\psi(v)} + \phi_\epsilon(v) - \frac{v}{2} \phi'_\epsilon(v), \tag{16}$$

where  $\psi(v) = v/\phi'_\epsilon(v) \neq 0$ .

The proof of Proposition 1 is given in Appendix A. Fig. 2(c) and (d) gives examples of using function (16) to majorize the smoothed penalty functions in Fig. 2(a) and (b), respectively.

#### 4. Exponential transient excision algorithm

To formulate the problem of exponential transient excision, we define  $\mathbf{R}$  as

$$\mathbf{R} := \begin{bmatrix} -r & 1 & & & \\ & -r & 1 & & \\ & & \ddots & \ddots & \\ & & & -r & 1 \end{bmatrix}. \quad (17)$$

The first-order derivative operator  $\mathbf{D}$  is a special case of  $\mathbf{R}$  with  $r=1$ . In this paper we restrict  $0 < r < 1$  to distinguish them. As a filter,  $\mathbf{R}$  can be seen to be a first-order filter attenuating low frequencies, with a transfer function

$$R(z) := 1 - rz^{-1}. \quad (18)$$

Driving the system  $R(z)$  with the step exponential

$$x(n) = \begin{cases} 0, & n < n_0, \\ r^{n-n_0}, & n \geq n_0, \end{cases} \quad (19)$$

produces an impulse  $\delta(n-n_0)$  as an output. If input signal  $\mathbf{x}$  is a step exponential with rate  $r$ , i.e., a Type 1 artifact, then

$$\mathbf{v} = \mathbf{R}\mathbf{x} \quad (20)$$

will be sparse.

Note that  $r$  should be known beforehand. In practice, we can estimate it from the time-constant of a Type 1 artifact. Using observation data  $\mathbf{y}$ , if we can measure the time  $N_0$  over which a Type 1 transient decays to half its initial height, then  $r$  can be found by solving  $r^{N_0} = 0.5$ . We ignore the influence of the slowly varying baseline, as we assume the exponential decays much faster. If the transients have an approximately equal decay rate, then we can use an average measured form multiple transients.

##### 4.1. Problem definition

In signal model (1), if an estimate  $\hat{\mathbf{x}}$  is known, then we can estimate  $f$  by

$$\hat{\mathbf{f}} = (\mathbf{y} - \hat{\mathbf{x}}) - \mathbf{H}(\mathbf{y} - \hat{\mathbf{x}}), \quad (21)$$

where  $\mathbf{H} = \mathbf{B}\mathbf{A}^{-1}$  is a highpass filter, where the  $\mathbf{A}$  and  $\mathbf{B}$  matrices are both banded with a structure as (7). Correspondingly, noise  $w$  is  $\mathbf{y} - (\hat{\mathbf{x}} + \hat{\mathbf{f}}) = \mathbf{H}(\mathbf{y} - \hat{\mathbf{x}})$ , which indicates a suitable data fidelity term  $\|\mathbf{H}(\mathbf{y} - \hat{\mathbf{x}})\|_2^2$ . Moreover, when the component  $x$  is modeled as (19), we can use  $\mathbf{v} = \mathbf{R}\mathbf{x}$  as its sparse representation. Hence, we propose to formulate the estimation of  $x$  from noisy data  $\mathbf{y}$  as the unconstrained optimization problem:

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \left\{ P_1(\mathbf{x}) = \|\mathbf{H}(\mathbf{y} - \mathbf{x})\|_2^2 + \lambda \sum_n \phi_\epsilon([\mathbf{R}\mathbf{x}]_n) \right\}, \quad (22)$$

where  $\lambda > 0$  is the regularization parameter related to noise level.

##### 4.2. Algorithm derivation

Using the majorizer of smoothed penalty function (16), and recalling that  $\mathbf{H} = \mathbf{B}\mathbf{A}^{-1}$  illustrated in Section 2.2, we

majorize the objective function (22),

$$\begin{aligned} G(\mathbf{x}, \mathbf{z}; P_1) &= \|\mathbf{H}(\mathbf{y} - \mathbf{x})\|_2^2 + \lambda \sum_n g_\epsilon([\mathbf{R}\mathbf{x}]_n, [\mathbf{R}\mathbf{z}]_n; \phi_\epsilon) \\ &= \|\mathbf{B}\mathbf{A}^{-1}(\mathbf{y} - \mathbf{x})\|_2^2 + \mathbf{x}^T \mathbf{R}^T [\Lambda(\mathbf{R}\mathbf{z})] \mathbf{R}\mathbf{x} + C \end{aligned} \quad (23)$$

where  $C$  does not depend on  $\mathbf{x}$ , and  $[\Lambda(\mathbf{R}\mathbf{z})] \in \mathbb{R}^{(N-1) \times (N-1)}$  is a diagonal matrix

$$[\Lambda(\mathbf{R}\mathbf{z})]_{n,n} = \frac{\lambda}{2\psi([\mathbf{R}\mathbf{z}]_n)} \quad (24)$$

where  $\psi(u) := u/\phi'_\epsilon(u)$ , which is listed in Table 1. Then the minimizer of (23) is given by

$$\mathbf{x} = \left[ \mathbf{A}^{-T} \mathbf{B}^T \mathbf{B} \mathbf{A}^{-1} + \mathbf{R}^T [\Lambda(\mathbf{R}\mathbf{z})] \mathbf{R} \right]^{-1} \mathbf{A}^{-T} \mathbf{B}^T \mathbf{B} \mathbf{A}^{-1} \mathbf{y}. \quad (25)$$

Note that calculating (25) directly requires the solution to a large dense system of equations. However, since  $\mathbf{A}$  is invertible, we can factor  $\mathbf{A}$  out and write (25) as

$$\mathbf{x}^{(k+1)} = \mathbf{A} \left[ \mathbf{B}^T \mathbf{B} + \mathbf{A}^T \mathbf{R}^T [\Lambda(\mathbf{R}\mathbf{x}^{(k)})] \mathbf{R} \mathbf{A} \right]^{-1} \mathbf{B}^T \mathbf{B} \mathbf{A}^{-1} \mathbf{y}, \quad (26)$$

where the MM procedure (10) is adopted with  $\mathbf{z} = \mathbf{x}^{(k)}$ . Note that the matrix to be inverted in (26) is banded, so that the solution can be computed efficiently by fast banded system solvers (e.g., [28,29]). Moreover, in (26), all matrix multiplications are between banded ones. Table 2 summarizes the algorithm to solve (22).

##### 4.3. Optimality condition and parameter selection

For problem (22), it is difficult to derive optimality conditions directly. Here, we consider the optimality condition indirectly via an equivalent problem, similar to the discussion of optimality condition for total variation problems in [2,17] and for LPF/TVD in [61].

To facilitate our derivation, we firstly denote the optimal solution of (22) by  $\mathbf{x}^*$ . We then define  $\mathbf{v}^* = \mathbf{R}\mathbf{x}^*$  where  $\mathbf{R}$  is given by (17). We define  $\mathbf{G} \in \mathbb{R}^{N \times (N-1)}$ ,

$$\mathbf{G} := \begin{bmatrix} 0 & & & & & \\ 1 & 0 & & & & \\ r & 1 & 0 & & & \\ r^2 & r & 1 & 0 & & \\ \vdots & & & \ddots & \ddots & \\ r^{N-3} & \dots & \dots & r & 1 & 0 \\ r^{N-2} & \dots & \dots & r^2 & r & 1 \end{bmatrix} \quad (27)$$

which satisfies

$$\mathbf{R}\mathbf{G} = \mathbf{I}. \quad (28)$$

The operator  $\mathbf{G}$  acts as an inverse filter of  $\mathbf{R}$ . Using the above assumptions, we derive the following proposition regarding optimality conditions for (22).

**Proposition 2.** If  $\mathbf{x}^*$  is an optimal solution to (22), then  $\mathbf{v}^* = \mathbf{R}\mathbf{x}^*$  is an optimal solution to problem

$$\mathbf{v}^* = \arg \min_{\mathbf{v}} Q(\mathbf{v}), \quad (29)$$

where

$$Q(\mathbf{v}) = \|\mathbf{H}(\mathbf{y}_0 - \mathbf{G}\mathbf{v})\|_2^2 + \lambda \sum_n \phi_\epsilon([\mathbf{v}]_n). \quad (30)$$

and  $\mathbf{y}_0 = \mathbf{y} - (\mathbf{x}^* - \mathbf{G}\mathbf{R}\mathbf{x}^*)$ .

**Proof.** We define  $\mathbf{x}_0 \in \mathbb{R}^N$  as

$$\mathbf{x}_0 := \mathbf{G}\mathbf{v}^* = \mathbf{G}\mathbf{R}\mathbf{x}^* \quad (31)$$

where  $\mathbf{x}^*$  is the optimal solution to (22). Note that  $\mathbf{x}_0$  and  $\mathbf{x}^*$  share an identical  $\mathbf{v}^* = \mathbf{R}\mathbf{x}_0 = \mathbf{R}\mathbf{x}^*$ . Moreover, we define the difference between  $\mathbf{x}^*$  and  $\mathbf{x}_0$  as

$$\mathbf{d} := \mathbf{x}^* - \mathbf{x}_0. \quad (32)$$

It follows that  $\mathbf{R}\mathbf{d} = \mathbf{R}\mathbf{x}^* - \mathbf{R}\mathbf{x}_0 = \mathbf{0}$ . As a consequence,  $\mathbf{x}_0$  is the optimal solution to problem

$$\mathbf{x}_0 = \arg \min_{\mathbf{x}} P_1(\mathbf{u}) \quad (33a)$$

$$\text{s.t. } \mathbf{x} = \mathbf{u} - \mathbf{d}. \quad (33b)$$

Note  $x_0(0) = 0$  must be satisfied in (33) because of the definition of  $\mathbf{G}$ . Thus, we can write an equivalent problem to (15) using (31),

$$\{\mathbf{x}_0, \mathbf{v}^*\} = \arg \min_{\mathbf{x}, \mathbf{v}} P_1(\mathbf{u}) \quad (34a)$$

$$\text{s.t. } \mathbf{x} = \mathbf{u} - \mathbf{d} \quad (34b)$$

$$\mathbf{x} = \mathbf{G}\mathbf{v} \quad (34c)$$

where  $\{\mathbf{x}_0, \mathbf{v}^*\}$  must be the optimal solution. In problem (34),  $\mathbf{u}$  is uniquely determined by linear functions of  $\mathbf{v}$  and  $\mathbf{d}$ , so problem (34) can be simplified by substituting the variables,

$$\{\mathbf{x}_0, \mathbf{v}^*\} = \arg \min_{\mathbf{x}, \mathbf{v}} Q(\mathbf{v}) \quad (35a)$$

$$\text{s.t. } \mathbf{x} = \mathbf{G}\mathbf{v} \quad (35b)$$

where  $Q(\mathbf{v}) = P_1(\mathbf{u})$ , with  $\mathbf{u} = \mathbf{G}\mathbf{v} + \mathbf{d}$ , and can be written explicitly,

$$Q(\mathbf{v}) = \|\mathbf{H}(\mathbf{y} - \mathbf{G}\mathbf{v} - \mathbf{d})\|_2^2 + \lambda \sum_n \phi_\epsilon([\mathbf{R}\mathbf{G}\mathbf{v} + \mathbf{R}\mathbf{d}]_n). \quad (36)$$

Because  $\mathbf{R}\mathbf{G} = \mathbf{I}$  and  $\mathbf{R}\mathbf{d} = \mathbf{0}$ , we simplify (36) as

$$Q(\mathbf{v}) = \|\mathbf{H}(\mathbf{y}_0 - \mathbf{G}\mathbf{v})\|_2^2 + \lambda \sum_n \phi_\epsilon([\mathbf{v}]_n) \quad (37)$$

where  $\mathbf{y}_0 = \mathbf{y} - \mathbf{d}$ .

In this case, the equality constraint in (35) is redundant. We can solve the unconstrained problem (29) first, and then compute  $\mathbf{x}_0$  by (31). This implies that  $\mathbf{v}^*$  is the optimal solution to (29) and satisfies  $\nabla Q(\mathbf{v}^*) = \mathbf{0}$ .

Using Proposition 2, instead of problem (22), we alternatively consider the optimality condition of problem (29) where as long as  $\mathbf{x}^*$  is an optimal solution of (22),  $\mathbf{v}^*$  is an optimal solution of (29). Moreover, we can rewrite the optimality condition of (29) as

$$p(n) = \lambda \phi'_\epsilon([\mathbf{v}^*]_n) = \lambda \phi'_\epsilon([\mathbf{R}\mathbf{x}^*]_n), \quad (38)$$

where  $\mathbf{p} = 2\mathbf{G}^T \mathbf{H}^T \mathbf{H}(\mathbf{y}_0 - \mathbf{x}_0)$ , which can be rewritten as

$$\mathbf{p} = 2\mathbf{G}^T \mathbf{H}^T \mathbf{H}(\mathbf{y} - \mathbf{x}^*). \quad (39)$$

Note that, writing  $\mathbf{p}$  as (39) does not require the solution to problem (29) or to compute  $\mathbf{d}$  and  $\mathbf{y}_0$ . Therefore, although we derived an indirect way to verify the optimality condition for (22), the final procedure is direct.

*Setting parameter  $\lambda$ :* Equation (38) can be used as a guide to set the regularization parameter  $\lambda$ . Suppose the observation data is composed of Gaussian noise only, then

a proper value of  $\lambda$  should make the solution of (22) almost identically zero. Thus, its sparse representation  $\mathbf{v}^*$ , given by (20), should be all zero as well. In this case, we can calculate a vector  $\mathbf{q}$ , which depends only on noise,

$$\mathbf{q} = 2\mathbf{G}^T \mathbf{H}^T \mathbf{H}\mathbf{w}. \quad (40)$$

When  $\mathbf{p}$  is calculated by (39), we find a constraint that  $-\lambda < p(n) < \lambda$ , by the property of  $\phi_\epsilon$ . Furthermore, since  $\mathbf{x}^* \approx \mathbf{0}$  is expected when the observation is pure noise (i.e.,  $\mathbf{y} = \mathbf{w}$ ), the values of  $\mathbf{q} \approx \mathbf{p}$  can be considered bounded:

$$q(n) \in [-\lambda, +\lambda] \quad \text{when } |v(n)| \leq \beta_\epsilon \quad \text{for } n \in \mathbb{Z}_N \quad (41)$$

with high probability. The value  $\beta_\epsilon$  here is related to  $\epsilon$ , and since  $\epsilon$  is extremely small, it can be simply assumed to be 0. As a consequence, the value of  $\lambda$  needs to satisfy

$$\lambda \geq \max\{|2[\mathbf{G}^T \mathbf{H}^T \mathbf{H}\mathbf{w}]_n|, \quad n \in \mathbb{Z}_N\}. \quad (42)$$

Further, if the statistical property of the noise can be exploited, e.g.,  $w(n) \sim \mathcal{N}(0, \sigma_w^2)$ ,  $\lambda$  can be set statistically, such that

$$\lambda = 2.5\sigma_w \|\mathbf{2}\mathbf{h}_1\|_2 \quad (43)$$

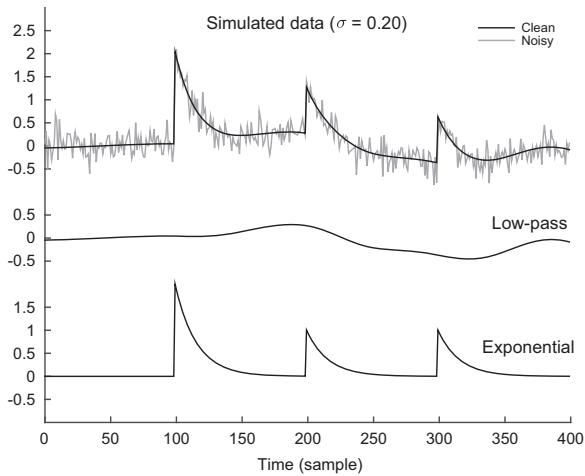
where  $\sigma_w$  is the standard deviation of the noise and  $\mathbf{h}_1$  is the impulse response corresponding to system  $\mathbf{G}^T \mathbf{H}^T \mathbf{H}$ . Note that,  $\mathbf{G}^T \mathbf{H}^T \mathbf{H}$  is a filter with a transfer function  $P_1(z) = H^2(z)/R(z)$ . Hence, although we give the matrix  $\mathbf{G}$  to derive the approach to set  $\lambda$ , it is not required to compute  $\mathbf{h}_1$  because the filter  $P_1(z)$  can be implemented directly using  $H(z)$  in (8), and  $R(z)$  in (18).

#### 4.4. Example: synthetic signal

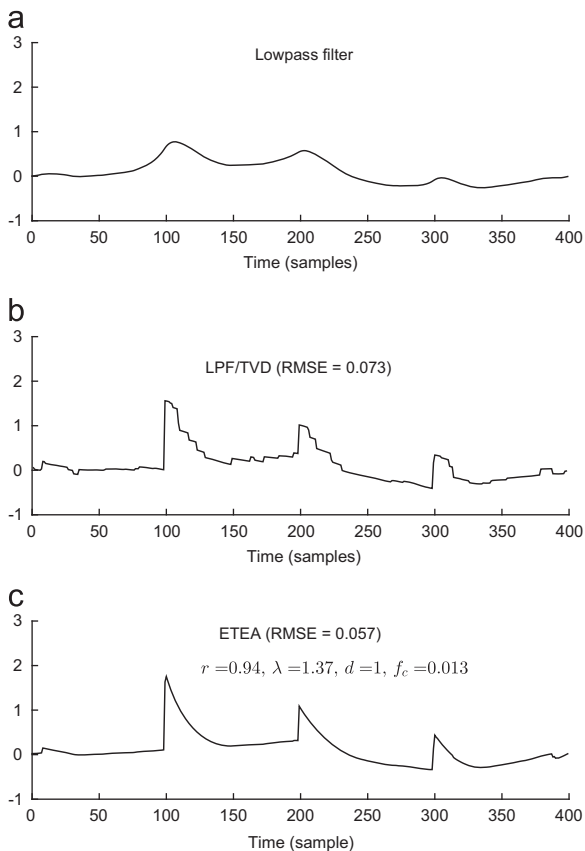
To illustrate ETEA, the algorithm is tested on a simulated signal and compared to other methods. Fig. 3 shows the test signal and its components: a lowpass signal  $f$ , a transient component  $x$ , composed of three exponential decays, and white Gaussian noise with  $\sigma_w = 0.20$ . The noisy observation is shown in Fig. 3 in gray.

*Denoising:* The result of conventional lowpass filtering is shown in Fig. 4(a). The result is smooth, but the step edges of each pulse are diminished. Moreover, if compared to the true low-pass baseline, the result is also distorted. The result of using LPF/TVD [61] is illustrated in Fig. 4(b). Although the abrupt jumps have been preserved, serious staircase effects appear. LPF/TVD uses the first-order derivative operator  $\mathbf{D}$  in regularization. The derivative of an exponential is another exponential, therefore the derivative is unlikely to have a sufficiently sparse representation. A compensative method is to widen the bandwidth of the lowpass filter, however, this leads to a noisy result. The result presented in Fig. 4(b) is selected by tuning all necessary parameters of LPF/TVD to optimize RMSE (root-mean-square error) value.

The result of the proposed ETEA method is shown in Fig. 4(c). The step edges and decay behavior are well preserved (RMSE=0.057). We use a smoothed  $\ell_1$ -norm penalty function. In this example, we set  $\epsilon = 10^{-10}$ ,  $r = 0.94$ , filter order parameter  $d = 1$ , cut-off frequency  $f_c = 0.013$  cycle/sample, and regularization parameter is calculated by (43) based on the noise principle.

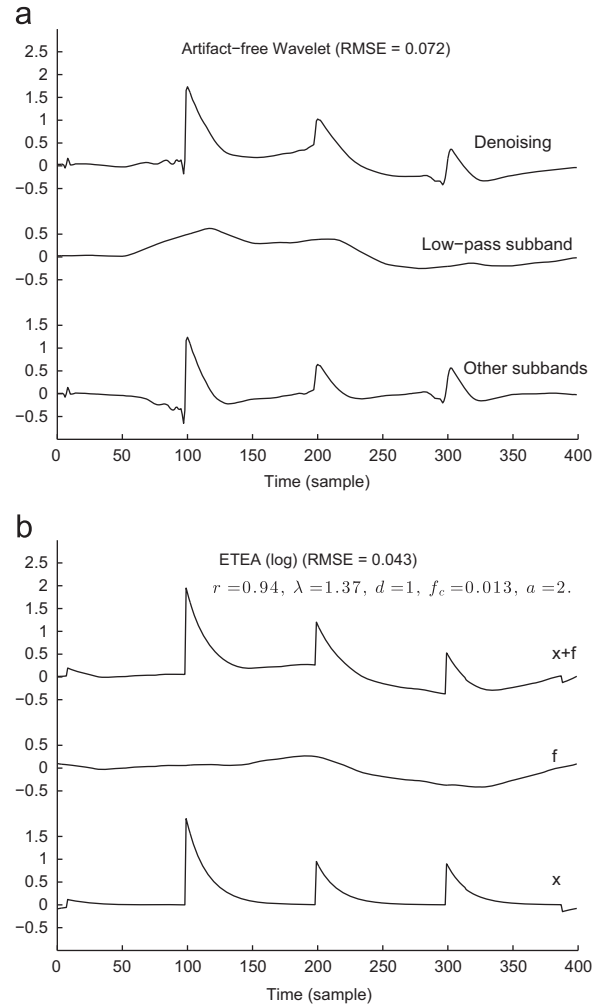


**Fig. 3.** Example 1: Simulated data comprising a lowpass component and exponential transients.



**Fig. 4.** Example 1: The denoising results of (a) lowpass filtering, (b) LPF/TVD, (c) ETEA (proposed method).

**Decomposition:** Wavelet-based methods for artifact correction have been described in [1,25,40,56]. In Fig. 5(a), we illustrate denoising and decomposition results obtained using the stationary (un-decimated) wavelet transform [11] with Haar wavelet using hard-threshold determined by the  $\sqrt{2} \log N \sigma_w$  thresholding scheme. The denoised result is



**Fig. 5.** Example 1: Comparison of denoising and decomposition results using (a) Wavelet-based method and, (b) ETEA with non-convex regularization.

further enhanced by the artifact-free denoising method [13,14] which uses total variation minimization to overcome pseudo-Gibbs oscillations. Although the denoising output is relatively smooth and captures the discontinuities, the decomposed components are both distorted. Compared to the true components in Fig. 3, the lowpass subband deviates from true lowpass component (especially at about  $n=100$ ). The transient component, reconstructed from other subbands after denoising, has to compensate for the error. Hence, the estimated transient component does not have a zero baseline (see bottom of Fig. 5(a)).

Non-convex penalty functions can be used to improve the result of ETEA. Here we use the smoothed non-convex logarithm penalty function in Table 1 ( $a=2$ ). The filtering result is shown in Fig. 5(b), where discontinuities are more accurately preserved, and the RMSE is reduced to 0.043 compared with the result in Fig. 4(c). We also illustrate the decomposed  $f$  and  $x$  components in Fig. 5(b). ETEA recovers both of the components accurately. The algorithm run-time is not affected by the choice of penalty function. In this example, 50 iterations of the algorithm takes about

25 ms on a MacBook Pro 2012 with a 2.7 GHz CPU, implemented in Matlab 8.4.

4.5. Example: artifact removal of ECoG data

Conventional EEG studies and most clinical EEG applications are restricted below 75 Hz [51]. Advanced measuring methods such as ECoG records multichannel cortical potentials from the micro-electrodes located inside human skull with a higher sampling rate. In this example, we use a 5-second ECoG signal epoch, with a sampling rate of 2713 Hz, recorded from a mesial temporal lobe epilepsy (MTLE) patient.

Fig. 6(a) shows the raw data in gray and the corrected data by ETEA in black. There are two Type 1 artifacts identified in this 5-s epoch. One is at about  $t=1.20$  s, and the other is at about  $t=3.60$  s. In signal model (1), suppressing component  $x$  which represents artifacts, the corrected data ( $f+w$ ) should preserve the components. We show the corrected data in Fig. 6(a), where the two Type 1 spikes are correctly removed. The decomposed artifact is illustrated in Fig. 6(a) as well, which adheres a zero-baseline.

Since wavelet-based methods have been successfully applied to suppress artifacts [1,25,40]. A comparison with a wavelet-based method is shown in Fig. 6(b). We use the stationary (un-decimated) wavelet transform [11] with Haar wavelet filter and the non-negative garrote threshold function [18], as recommended in [25]. The thresholding has been applied to all the subbands except the lowpass band, and the artifact component is obtained by subtracting the corrected data from the raw data. As shown, this approach estimates transient pulses but the estimated pulse adheres to the shape of the Haar wavelet filter: a positive-negative pulse (see the bottom square box in Fig. 6(b)), but not the true Type 1 artifact in Fig. 1(b).

To more clearly illustrate the estimated results by the two methods, we show the details of the corrected data and the artifact from  $t=3.50$  to  $3.65$  s in dashed line boxes in Fig. 1(a) and (b), respectively. Since the wavelet-based method cannot correctly estimate Type 1 artifact in this case, a ‘bump’ can be observed in Fig. 6(b) to compensate the error. Moreover, this inappropriate estimation will cause the corrected data to change before Type 1 artifact actually occurs (see Fig. 1(b)). In contrast, ETEA estimates the artifact as the abrupt drift with a decay, then it excises the artifact without influencing the data before the transient occurs (see Fig. 1(a)).

In this example, for the signal with a length of 13,565 samples, the proposed algorithm converges within 40 iterations, and takes about 330 ms on a MacBook Pro 2012 with 2.7 GHz CPU, implemented in Matlab 8.4. The cost function history is shown in Fig. 7.

5. Higher-order ETEA

In the previous section, we have presented ETEA based on signal model (1), suitable for Type 1 artifacts, where the component  $x$  has discontinuous step exponential transients. Here we illustrate another version of ETEA based on

signal model (1), where the component  $x$  models Type 0 artifacts.

First, we consider the operator

$$\mathbf{R}_2 := \begin{bmatrix} r^2 & -2r & 1 & & & & \\ & r^2 & -2r & 1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & r^2 & -2r & 1 & \end{bmatrix}, \tag{44}$$

where  $\mathbf{R}_2 \in \mathbb{R}^{(N-2) \times N}$  has three non-zero coefficients in each row. As a special case, when  $r=1$ ,  $\mathbf{R}_2$  is the second-order difference operator, and  $\|\mathbf{R}_2 \mathbf{x}\|_1$  is the same as the regularizer used in [30] for  $\ell_1$  detrending.

Taking  $\mathbf{R}_2$  as a filter, the transfer function is

$$R_2(z) = (1 - rz^{-1})^2, \tag{45}$$

which has a double zero at  $z=r$ . The impulse response of system  $R_2^{-1}(z)$  is  $(n+1)r^n$ , when  $n \geq 0$ . It has the same shape as Fig. 1(a), which is a suitable model for the Type 0 artifacts in [25]. Therefore,  $\mathbf{R}_2 \mathbf{x}$  is sparse when  $x$  is composed of such piecewise smooth transients. Similar to (22), we formulate the problem

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \left\{ P_2(\mathbf{x}) = \|\mathbf{H}(\mathbf{y} - \mathbf{x})\|_2^2 + \lambda \sum_n \phi_\epsilon([\mathbf{R}_2 \mathbf{x}]_n) \right\} \tag{46}$$

to estimate the transient component  $x$ , so that the corrected data is  $y - \hat{x}$ , and the low-pass trend  $f$  can be estimated by (21). Additionally, (46) can be easily solved by ETEA (in Table 2) substituting  $\mathbf{R}$  by  $\mathbf{R}_2$ .

An example of complicated and irregular artifacts is the eye blink/movement artifacts in EEG data, which may have various morphologies and durations. In this case, we assume that the artifacts can be estimated by a continuous piecewise smooth waveform, generated by applying a certain sparse impulse sequence to  $R_2^{-1}(z)$  (45). In other words, we broaden our signal model so that the artifact is not only equivalent to an isolated Type 0 transient as in [25], but also a superposition of multiple such transients, with some freedom of scaling and shifting. The problem of estimating  $x$  can be solved by (46) in this case as well.

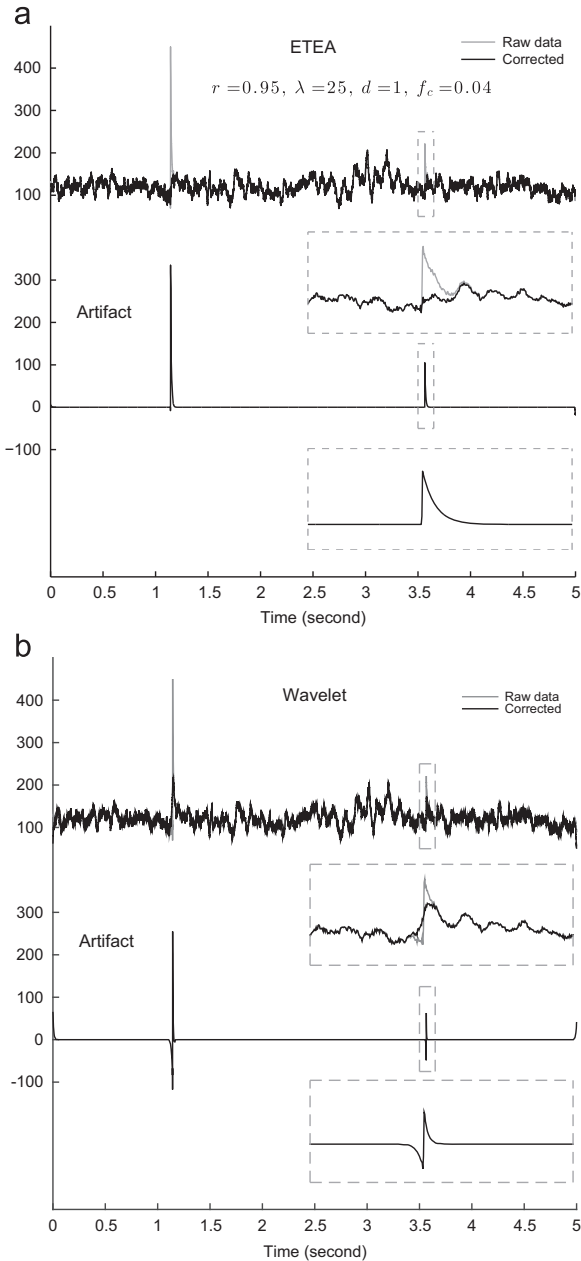
5.1. Example: simulated data

Fig. 8 shows the simulated data and its lowpass and transient components. The transient component has several pulses with different heights and widths. They are obtained by feeding a sequence of impulses into system  $R_2^{-1}(z)$ . The filter is given in (5) with  $a_k = [1, -2r, r^2]$ ,  $b_0 = 1$ , and  $r=0.950$ .

Empirical mode decomposition (EMD) [22,23], a powerful method for analyzing signals, has been successfully utilized in different fields, including neuroscience, biometrics, speech recognition, electrocardiogram (ECG) analysis, and fault detection [8,16,21,39,54,62,65,67].

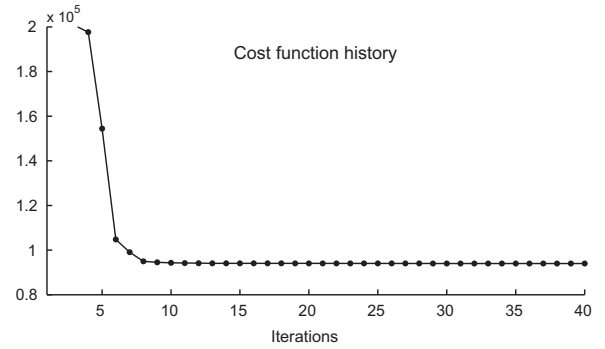
As a comparison to the proposed approach, we use the EMD based denoising algorithm in [32], which uses wavelet coefficient thresholding techniques on decomposed IMFs. More specifically, we use clear first iterative interval thresholding (CIIT) with smoothly clipped absolute deviation (SCAD) penalty thresholding [31,32], with 20 iterations, and the result is shown in Fig. 9(b). In this





**Fig. 6.** Example 2: Comparison of artifacts removal for ECoG data. (a) ETEA (proposed method), (b) wavelet-based method.

example, in order to perform decomposition, among the entire eight IMFs, the thresholding is performed on IMFs 1–5, their summation is considered as the transients illustrated in Fig. 9(a), and IMFs 6–8 are considered as the estimation of the lowpass component. The EMD based method achieves a decent denoising performance, but does not accurately estimate the components. For instance, the simulated data has a smooth dip at about  $n=700$  (circled in Fig. 9(a)). EMD decomposes it into higher IMFs since they are varying slowly, which degrades the estimation. We may group the lowpass and transient



**Fig. 7.** Example 2: Cost function history.

components differently to avoid this problem, for instance, grouping IMF 1–6 together in order to include more oscillations into transient component, but this causes other distortion, where the decomposed transient component contains a lowpass signal and does not adhere to a baseline of zero.

The result obtained using second-order ETEA is illustrated in Fig. 9(b). ETEA estimates both the low-pass and transient components well, and recovers the signal by  $x+f$  precisely with  $\text{RMSE}=0.87$  (with the smoothed  $\ell_1$ -penalty function). The regularization parameter  $\lambda$  for problem (46) was similar to (43),

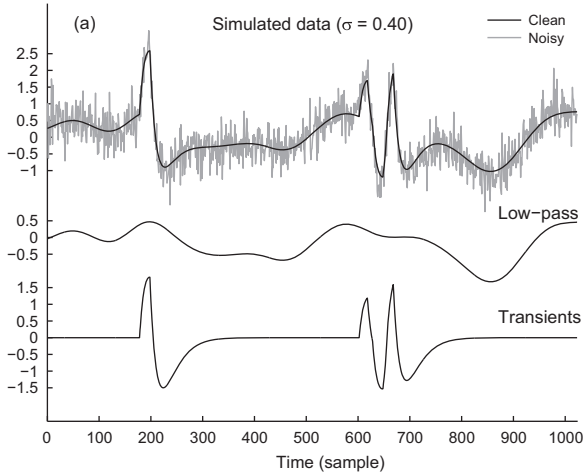
$$\lambda \approx 2.5\sigma_w \|\mathbf{2h}_2\|_2, \quad (47)$$

where  $\mathbf{h}_2$  is the impulse response of system  $H^2(z)/R_2(z)$ , and  $R_2(z)$  is defined in (45). The decomposition is accurate. There are no compensating waveforms between the estimated  $x$  and  $f$  at  $n=700$ , comparing to the estimation in Fig. 9(a).

Through numerical experiments, we found that second-order ETEA is not very sensitive to parameter  $r$ . Fig. 9(c) shows the RMSE of denoising the data in Fig. 8(a), using  $r \in [0.85, 0.99]$ . In this test, all filter parameters ( $f_c$  and  $d$ ) are the same and  $\lambda$  is set by (47). In most cases, the results are better than EMD-CIIT. In addition,  $r$  should not be too small or very close to 1. If  $r$  has to be very small to yield a good estimation of component  $x$ , it must fluctuate extremely rapidly, then it must be closer to a sequence of sparse spikes (i.e., Type 3 artifacts in [25]), which differs from the signal model. For such a signal, other algorithms are more suitable, e.g., LPF/CSD [60,61]. Similarly, if  $r$  has to be very close to 1 to fit the transients, then the transients must be very close to a piecewise linear signal, which is also not how we model the signal initially. As a consequence, we suggest to set  $r$  in the range  $0.90 < r < 0.98$ .

## 5.2. Example: ocular artifacts suppression

In this example, we use ETEA with  $\mathbf{R}_2$  to correct EEG with eye blink/movement artifacts. Fig. 10a shows a 10-s signal from channel Fp1, with sampling rate  $f_s = 256$  Hz, downloaded from [38]. As a channel located on forehead, Fp1 is very sensitive to the motion of eyes and eyebrows, and in this example, eye movement artifacts with large amplitudes are present through the entire signal. Applying



**Fig. 8.** Example 3: The simulated test data, generated by driving an AR filter with a sparse sequence and adding noise.

second-order ETEA with  $r=0.95$ , the results for corrected data and extracted OA are illustrated in Fig. 10(b).

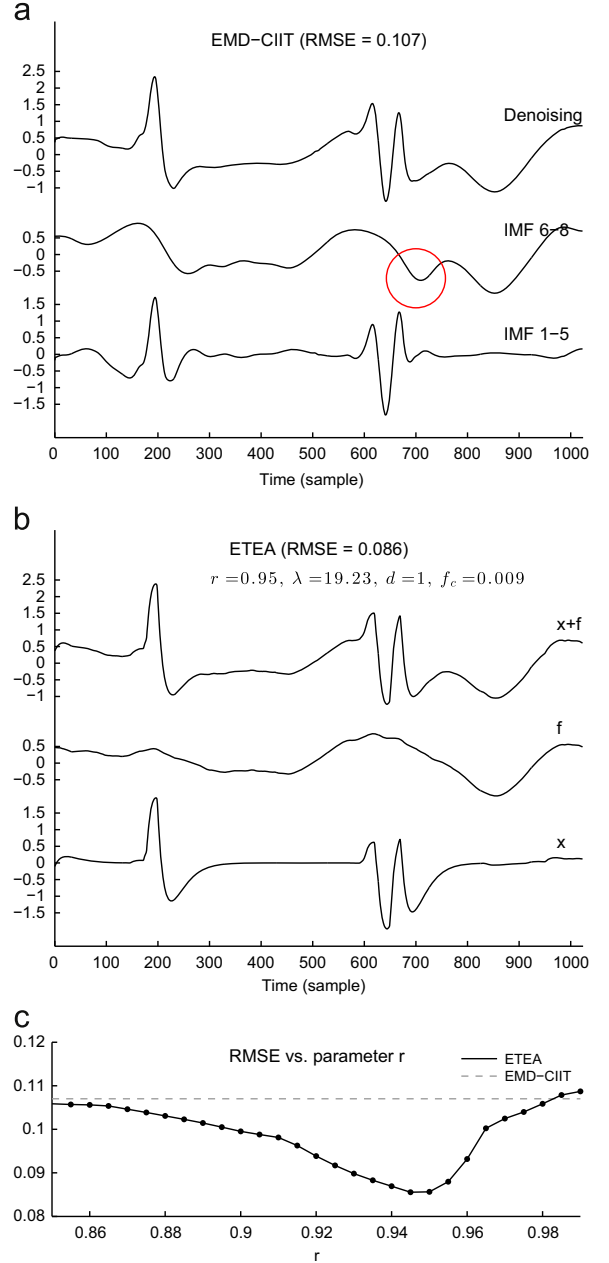
As a comparison, we use multivariate empirical mode decomposition (MEMD) [52] to correct the data. MEMD is a recently developed algorithm extending conventional EMD. It has been used in different aspects of EEG signal analysis and applications [9,50,49], including removing the ocular artifacts (OA) from multichannel EEG data [53,42]. In this example, we use 4 EEG channels (Fp1, Fp2, C3, C4) measured simultaneously as the input, and decompose the higher-index IMFs (low-frequency subbands) considered to be artifacts [53, Section V]. More specifically, among all 14 IMFs decomposed in this example, we use IMF 1–4 as the corrected data, and the rest as artifacts. The corrected data and estimated OA are shown in Fig. 10(c).

From the results in Fig. 10(b) and (c), ETEA estimates artifacts more clearly than MEMD. In the MEMD result, some small-amplitude higher frequency oscillations leak into the artifact (e.g., about  $t=5.5$  and  $9.0$  s). Moreover, some artifacts are introduced after applying MEMD method to correct the data, (e.g., about  $t=3.0$  and  $9.5$  s in Fig. 10(c)). Some oscillations are generated as transients where the abrupt artifacts occur. In contrast, in Fig. 10(b), there are no oscillations introduced either in the estimated artifacts or the corrected data.

**Computational efficiency:** Fig. 11 shows the average computation time as a function of the signal length. In this experiment, we calculate the time of computation of ETEA (22) and second-order ETEA (46) with different filter settings (controlled by parameter  $d$ ), using input signal lengths from 5000 to  $10^5$ . For each length, we average the computation time over 10 trials. For each trial, run 40 iterations of each algorithm. The experiment is implemented in Matlab 8.4 on a MacBook Pro 2012 with 2.7 GHz CPU.

As shown, the proposed algorithms have a run time of order  $O(N)$ . Most of the computation time is consumed by the step of solving the linear system  $\mathbf{Q}^{-1}\mathbf{b}$  in Table 2, where  $\mathbf{Q}$  is a banded matrix and we use a fast banded solver.

Additionally, fast iterative shrinkage-thresholding algorithm (FISTA) [3,7], which is an acceleration scheme for

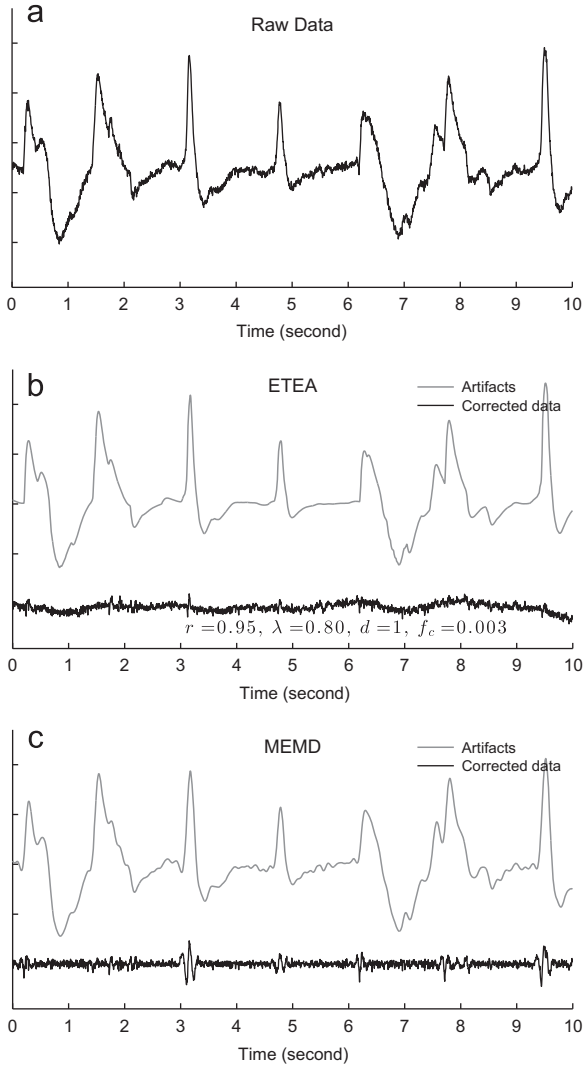


**Fig. 9.** Example 3: Denoising and decomposition results by (a) EMD based denoising (EMD-CIIT [32]) and (b) ETEA.

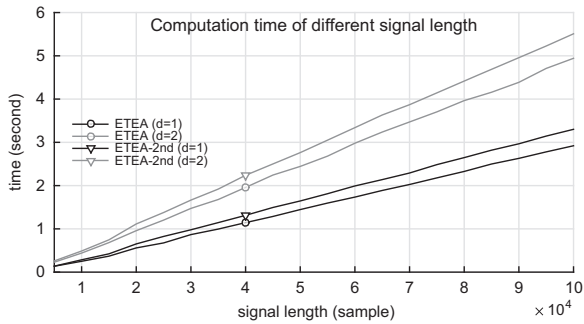
iterative shrinkage-thresholding algorithm (ISTA) [6], (a special formulation of MM) may be used to further accelerate the algorithm.

## 6. Conclusion

This paper proposes a new algorithm for denoising and artifact removal for signals comprising artifacts arising in measured data, e.g., neural time-series recordings, using sparse optimization. The first algorithm, ETEA, assumes



**Fig. 10.** Example 4: (a) EEG data with ocular artifacts, and corrected data using (b) ETEA and (c) MEMD.



**Fig. 11.** Comparison of computation time with different signal lengths.

that the signal is composed of a lowpass signal and an exponential transients (Type 1). It is formulated as an optimization problem regularized by differentiable and smooth penalty function. The second algorithm is an

extension of ETEA, using a higher-order recursive filter, which is applicable for correction of continuous protuberance transients (Type 0), and more irregular artifacts. As applications, we have shown that ETEA with different regularizers ( $\mathbf{R}$  and  $\mathbf{R}_2$ ) are suitable for the suppression of Type 1 artifact in ECoG data and ocular artifacts (OA) (as sequential Type 0 artifacts) in conventional EEG data, with detailed comparisons to some state-of-the-art methods. Both of the above algorithms are computationally efficient because they are formulated in terms of banded matrices. A promising future work is to extend the above data correcting methods to multichannel data.

## Acknowledgments

The authors would like to thank Jonathan Viventi of the Department of Biomedical Engineering of Duke University, and Justin Blanco of the Electrical and Computer Engineering Department of United States Naval Academy, for providing the data and giving useful comments.

## Appendix A. Proof of Proposition 1

**Proof.** Substitute the variables  $u$  and  $v$  in (14) by

$$u = \sqrt{x^2 + \epsilon}, \quad \text{and} \quad v = \sqrt{z^2 + \epsilon}. \quad (\text{A.1})$$

Therefore, for  $x, z \in \mathbb{R}$ , the inequality holds

$$\begin{aligned} \frac{\phi'(\sqrt{z^2 + \epsilon})}{2\sqrt{z^2 + \epsilon}}(x^2 + \epsilon) + \phi(\sqrt{z^2 + \epsilon}) - \frac{\sqrt{z^2 + \epsilon}}{2} \phi'(\sqrt{z^2 + \epsilon}) \\ \geq \phi(\sqrt{x^2 + \epsilon}). \end{aligned} \quad (\text{A.2})$$

The right of the inequality is the majorizer of the smoothed penalty function  $\phi(\sqrt{x^2 + \epsilon})$ .

(1) If  $z \neq 0$ , we can multiply  $z$  to the numerator and denominator of the first term on the left side of (A.2),

$$\begin{aligned} \frac{\phi'(\sqrt{z^2 + \epsilon})}{2\sqrt{z^2 + \epsilon}}(x^2 + \epsilon) &= \frac{x^2}{2z} \left( \phi'(\sqrt{z^2 + \epsilon}) \frac{z}{\sqrt{z^2 + \epsilon}} \right) \\ &+ \frac{\epsilon}{2z} \left( \phi'(\sqrt{z^2 + \epsilon}) \frac{z}{\sqrt{z^2 + \epsilon}} \right) = \frac{x^2}{2z} \phi'_\epsilon(z) + \frac{\epsilon}{2z} \phi'_\epsilon(z). \end{aligned} \quad (\text{A.3})$$

Multiplying the numerator and denominator of the third term on the left side of (A.2) by  $\sqrt{z^2 + \epsilon}$ ,

$$\begin{aligned} \frac{\sqrt{z^2 + \epsilon}}{2} \phi'(\sqrt{z^2 + \epsilon}) &= \frac{z^2}{2\sqrt{z^2 + \epsilon}} \phi'(\sqrt{z^2 + \epsilon}) \\ &+ \frac{\epsilon}{2\sqrt{z^2 + \epsilon}} \phi'(\sqrt{z^2 + \epsilon}) = \frac{z}{2} \left( \phi'(\sqrt{z^2 + \epsilon}) \frac{z}{\sqrt{z^2 + \epsilon}} \right) \\ &+ \frac{\epsilon}{2z} \left( \phi'(\sqrt{z^2 + \epsilon}) \frac{z}{\sqrt{z^2 + \epsilon}} \right) = \frac{z}{2} \phi'_\epsilon(z) + \frac{\epsilon}{2z} \phi'_\epsilon(z). \end{aligned} \quad (\text{A.4})$$

Using the results in (A.3) and (A.4), the inequality (A.2) can be rewritten as

$$\frac{x^2}{2z} \phi'_\epsilon(z) + \frac{\epsilon}{2z} \phi'_\epsilon(z) + \phi_\epsilon(z) - \frac{z}{2} \phi'_\epsilon(z) - \frac{\epsilon}{2z} \phi'_\epsilon(z) \geq \phi_\epsilon(x), \quad (\text{A.5})$$

which can be reorganized into

$$g_\epsilon(x, z; \phi_\epsilon) = \frac{\phi'_\epsilon(z)}{2z}x^2 + \phi_\epsilon(z) - \frac{z}{2}\phi'_\epsilon(z) \geq \phi_\epsilon(x). \quad (\text{A.6})$$

(2) If  $z=0$  and  $x \neq 0$ , by *Lagrange's Mean Value Theorem* [55, Theorem 5.10], since function  $\phi(x)$  is continuous and  $\phi'(x) \leq 0$  on  $x > 0$ , there exists a value  $\xi$  in the range  $\sqrt{\epsilon} < \xi < \sqrt{x^2 + \epsilon}$  that  $\phi'(\sqrt{\epsilon}) \geq \phi'(\xi) \geq \phi'(\sqrt{x^2 + \epsilon})$  satisfying

$$\phi'(\xi)(\sqrt{x^2 + \epsilon} - \sqrt{\epsilon}) = \phi(\sqrt{x^2 + \epsilon}) - \phi(\sqrt{\epsilon}). \quad (\text{A.7})$$

Moreover, consider the square that is always positive

$$(\sqrt{x^2 + \epsilon} - \sqrt{\epsilon})^2 = x^2 + \epsilon - 2\sqrt{\epsilon(x^2 + \epsilon)} + \epsilon > 0, \quad (\text{A.8})$$

which implies the inequality

$$x^2 > 2\sqrt{\epsilon(x^2 + \epsilon)} - 2\epsilon. \quad (\text{A.9})$$

Furthermore, we can multiply both sides of (A.9) by a positive term  $\phi'(\sqrt{\epsilon})/2\sqrt{\epsilon}$ , and then adopt the result from (A.7), so that

$$\frac{\phi'(\sqrt{\epsilon})}{2\sqrt{\epsilon}}x^2 > \frac{\phi'(\sqrt{\epsilon})}{2\sqrt{\epsilon}}\left(2\sqrt{\epsilon(x^2 + \epsilon)} - 2\epsilon\right) \quad (\text{A.10a})$$

$$= \phi'(\sqrt{\epsilon})(\sqrt{x^2 + \epsilon} - \sqrt{\epsilon}) \quad (\text{A.10b})$$

$$> \phi'(\xi)(\sqrt{x^2 + \epsilon} - \sqrt{\epsilon}) \quad (\text{A.10c})$$

$$= \phi(\sqrt{x^2 + \epsilon}) - \phi(\sqrt{\epsilon}), \quad (\text{A.10d})$$

which leads to

$$\frac{\phi'(\sqrt{\epsilon})}{2\sqrt{\epsilon}}x^2 + \phi(\sqrt{\epsilon}) > \phi(\sqrt{x^2 + \epsilon}). \quad (\text{A.11})$$

Because  $\phi_\epsilon$  in (13) is differentiable on  $\mathbb{R}$ , we can find its second-order derivative at zero,

$$\lim_{z \rightarrow 0} \phi''_\epsilon(z) = \frac{\phi'(\sqrt{\epsilon})}{\sqrt{\epsilon}}, \quad (\text{A.12})$$

and by L'Hôpital's rule [55, Theorem 5.13], we have

$$\lim_{z \rightarrow 0} \frac{\phi'_\epsilon(z)}{2z} = \lim_{z \rightarrow 0} \frac{\phi''_\epsilon(z)}{2} = \frac{\phi'(\sqrt{\epsilon})}{2\sqrt{\epsilon}}, \quad (\text{A.13})$$

which implies that (A.11) is in the same form of the majorizer (A.6) at  $z=0$ .

(3) If  $x = z = 0$ , then the condition (9) follows immediately. □

**References**

[1] M.T. Akhtar, W. Mitsuhashi, C.J. James, Employing spatially constrained ICA and wavelet denoising, for automatic removal of artifacts from multichannel EEG data, *Signal Process.* 92 (2) (2012) 401–416.  
 [2] F. Bach, R. Jenatton, J. Mairal, G. Obozinski, Optimization with sparsity-inducing penalties, *Found. Trends Mach. Learn.* 4 (1) (2012) 1–106.  
 [3] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imag. Sci.* 2 (1) (2009) 183–202.  
 [4] E.J. Candès, M.B. Wakin, S. Boyd, Enhancing sparsity by reweighted  $l_1$  minimization, *J. Fourier Anal. Appl.* 14 (December (5)) (2008) 877–905.

[5] M. Cetin, W.C. Karl, Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization, *IEEE Trans. Image Process.* 10 (April (4)) (2001) 623–631.  
 [6] A. Chambolle, R.A. de Vore, N.-Y. Lee, B.J. Lucier, Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage, *IEEE Trans. Image Process.* 7 (March (3)) (1998) 319–335.  
 [7] A. Chambolle, V.R. Dossal, On the convergence of the iterates of FISTA, hal-01060130, September 2014, preprint.  
 [8] C.-P. Chang, J.-C. Lee, Y. Su, P.S. Huang, T.-M. Tu, Using empirical mode decomposition for iris recognition, *Comput. Stand. Interfaces* 31 (4) (2009) 729–739.  
 [9] H.-C. Chang, P.-L. Lee, M.-T. Lo, Y.-T. Wu, K.-W. Wang, G.-Y. Lan, Inter-trial analysis of post-movement beta activities in EEG signals using multivariate empirical mode decomposition, *IEEE Trans. Neural Syst. Rehabil. Eng.* 21 (July (4)) (2013) 607–615.  
 [10] P.-Y. Chen, I.W. Selesnick, Group-sparse signal denoising: non-convex regularization, convex optimization, *IEEE Trans. Signal Process.* 62 (July (13)) (2014) 3464–3478.  
 [11] R.R. Coifman, D.L. Donoho, Translation-invariant denoising, in: A. Antoniadis (Ed.), *Lecture Notes in Wavelets and Statistics*, Springer-Verlag, 1995.  
 [12] J. Dammers, M. Schiek, F. Boers, C. Silex, M. Zvyagintsev, U. Pietrzyk, K. Mathiak, Integration of amplitude and phase statistics for complete artifact removal in independent components of neuromagnetic recordings, *IEEE Trans. Biomed. Eng.* 55 (October (10)) (2008) 2353–2362.  
 [13] S. Durand, J. Froment, Artifact free signal denoising with wavelets, in: *Proceedings of ICASSP*, 2001.  
 [14] S. Durand, J. Froment, Reconstruction of wavelet coefficients using total variation minimization, *SIAM J. Sci. Comput.* 24 (5) (2003) 1754–1767.  
 [15] M. Figueiredo, J. Bioucas-Dias, R. Nowak, Majorization–minimization algorithms for wavelet-based image restoration, *IEEE Trans. Image Process.* 16 (December (12)) (2007) 2980–2991.  
 [16] J. Fleureau, A. Kachenoura, L. Albera, J.-C. Nunes, L. Senhadji, Multivariate empirical mode decomposition and application to multi-channel filtering, *Signal Process.* 91 (12) (2011) 2783–2792.  
 [17] J.-J. Fuchs, On sparse representations in arbitrary redundant bases, *IEEE Trans. Inf. Theory* 50 (6) (2004) 1341–1344.  
 [18] H. Gao, Wavelet shrinkage denoising using the non-negative garrote, *J. Comput. Graph. Stat.* 7 (4) (1998) 469–488.  
 [19] A. Gholami, S.M. Hosseini, A balanced combination of Tikhonov and total variation regularizations for reconstruction of piecewise-smooth signals, *Signal Process.* 93 (7) (2013) 1945–1960.  
 [20] C. Guerrero-Mosquera, A. Navia-Vázquez, Automatic removal of ocular artifacts using adaptive filtering and independent component analysis for electroencephalogram data, *IET Signal Process.* 6 (2) (2012) 99–106.  
 [21] H. Huang, J. Pan, Speech pitch determination based on Hilbert–Huang transform, *Signal Process.* 86 (April (4)) (2006) 792–803.  
 [22] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. Lond. A*, 454 (March) (1998) 903–995.  
 [23] N.E. Huang, M.-L.C. Wu, S.R. Long, S.S.P. Shen, W. Qu, P. Gloersen, K.L. Fan, A confidence limit for the empirical mode decomposition and Hilbert spectral analysis, *Proc. R. Soc. Lond. Ser. A: Math. Phys. Eng. Sci.* 459 (September (2037)) (2003) 2317–2345.  
 [24] D.R. Hunter, K. Lange, A tutorial on MM algorithms, *Am. Stat.* 58 (2004) 30–37.  
 [25] M.K. Islam, A. Rastegarnia, A.T. Nguyen, Z. Yang, Artifact characterization and removal for in vivo neural recording, *J. Neurosci. Methods* 226 (0) (2014) 110–123.  
 [26] C.A. Joyce, I.F. Gorodnitsky, M. Kutas, Automatic removal of eye movement and blink artifacts from EEG data using blind component separation, *Psychophysiology* 41 (2) (2004) 313–325.  
 [27] F.I. Karahanoglu, I. Bayram, D. Van De Ville, A signal processing approach to generalized 1-d total variation, *IEEE Trans. Signal Process.* 59 (November (11)) (2011) 5265–5274.  
 [28] E. Kilic, Explicit formula for the inverse of a tridiagonal matrix by backward continued fractions, *Appl. Math. Comput.* 197 (1) (2008) 345–357.  
 [29] E. Kilic, P. Stanica, The inverse of banded matrices, *J. Comput. Appl. Math.* 237 (1) (2013) 126–135.  
 [30] S. Kim, K. Koh, S. Boyd, D. Gorinevsky,  $\ell_1$  trend filtering, *SIAM Rev.* 51 (2) (2009) 339–360.

- [31] Y. Kopsinis, S. McLaughlin, Empirical mode decomposition based soft-thresholding, in: 16th European Signal Processing Conference, 2008.
- [32] Y. Kopsinis, S. McLaughlin, Development of EMD-based denoising methods inspired by wavelet thresholding, *IEEE Trans. Signal Process.* 57 (April (4)) (2009) 1351–1362.
- [33] M. Kowalski, Sparse regression using mixed norms, *Appl. Comput. Harmonic Anal.* 27 (3) (2009) 303–324.
- [34] I. Kozlov, A. Petukhov, Sparse solutions of underdetermined linear systems, in: W. Freeden, et al., (Eds.), *Handbook of Geomathematics*, Springer, 2010.
- [35] K. Lange, *Optimization*, Springer, New York, 2004.
- [36] K. Lange, D. Hunter, I. Yang, Optimization transfer using surrogate objective functions, *J. Comput. Graph. Stat.* 9 (2000) 1–20.
- [37] N. Mammone, F. L. a Foresta, F.C. Morabito, Automatic artifact rejection from multichannel scalp EEG by wavelet ICA, *IEEE J. Sens.* 12 (March (3)) (2012) 533–542.
- [38] D. Mandic, Empirical mode decomposition, multivariate EMD, matlab and data sources.
- [39] D.P. Mandic, N.U. Rehman, Z. Wu, N.E. Huang, Empirical mode decomposition-based time-frequency analysis of multivariate signals: the power of adaptive data analysis, *IEEE Signal Process. Mag.* 30 (November (6)) (2013) 74–86.
- [40] B. Molavi, G.A. Dumont, Wavelet-based motion artifact removal for functional near-infrared spectroscopy, *Physiol. Meas.* 33 (2) (2012) 259.
- [41] M.K.I. Molla, M.R. Islam, T. Tanaka, T.M. Rutkowski, Artifact suppression from EEG signals using data adaptive time domain filtering, *Neurocomputing* 97 (2012) 297–308.
- [42] M.K.I. Molla, T. Tanaka, T.M. Rutkowski, Multivariate EMD based approach to EOG artifacts separation from EEG, in: 2012 Proceedings of ICASSP, 2012, pp. 653–656.
- [43] M.K.I. Molla, T. Tanaka, T.M. Rutkowski, A. Cichocki, Separation of EOG artifacts from EEG signals using bivariate EMD, in: 2010 Proceedings of ICASSP, 2010, pp. 562–565.
- [44] K. Nazarpour, Y. Wongsawat, S. Sanei, J.A. Chambers, S. Orintara, Removal of the eye-blink artifacts from EEGs via STF-TS modeling and robust minimum variance beamforming, *IEEE Trans. Biomed. Eng.* 55 (9) (2008) 2221–2231.
- [45] D. Needell, Noisy signal recovery via iterative reweighted  $\ell_1$ -minimization, in: Proceedings of Forty-Third Asilomar Conference on Signals, Systems and Computers, 2009, pp. 113–117.
- [46] M. Nikolova, Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares, *Multiscale Model. Simul.* 4 (3) (2005) 960–991.
- [47] X. Ning, I.W. Selesnick, ECG enhancement and QRS detection based on sparse derivatives, *Biomed. Signal Process. Control* 8 (6) (2013) 713–723.
- [48] B. Nouredin, P.D. Lawrence, G.E. Birch, Online removal of eye movement and blink EEG artifacts using a high-speed eye tracker, *IEEE Trans. Biomed. Eng.* 59 (8) (2012) 2103–2110.
- [49] A. Omidvarnia, G. Azemi, P.B. Colditz, B. Boashash, A time-frequency based approach for generalized phase synchrony assessment in nonstationary multivariate signals, *Digit. Signal Process.* 23 (3) (2013) 780–790.
- [50] C. Park, M. Plank, J. Snider, S. Kim, H.C. Huang, S. Gepshtein, T. P. Coleman, H. Poizner, EEG gamma band oscillations differentiate the planning of spatially directed movements of the arm versus eye: multivariate empirical mode decomposition analysis, *IEEE Trans. Neural Syst. Rehabil. Eng.* 22 (September (5)) (2014) 1083–1096.
- [51] R.M. Rangayyan, *Biomedical Signal Analysis—A Case-study Approach*, IEEE and Wiley, New York, NY, 2002.
- [52] N.U. Rehman, D.P. Mandic, Multivariate empirical mode decomposition, *Proc. R. Soc. A*, 466 (2117) (2010) 1291–1302.
- [53] N.U. Rehman, D.P. Mandic, Filter bank property of multivariate empirical mode decomposition, *IEEE Trans. Signal Process.* 59 (May (5)) (2011) 2421–2426.
- [54] G. Rilling, P. Flandrin, One or two frequencies? The empirical mode decomposition answers, *IEEE Trans. Signal Process.* 56 (1) (2008) 85–95.
- [55] W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, 1976.
- [56] H. Sato, N. Tanaka, M. Uchida, Y. Hirabayashi, M. Kanai, T. Ashida, I. Konishi, A. Maki, Wavelet analysis for detecting body-movement artifacts in optical topography signals, *NeuroImage* 33 (2) (2006) 580–587.
- [57] E.D. Schifano, R.L. Strawderman, M.T. Wells, Majorization-minimization algorithms for nonsmoothly penalized objective functions, *Electron. J. Stat.* 4 (2010) 1258–1299.
- [58] I.W. Selesnick, S. Arnold, V. Dantham, Polynomial smoothing of time series with additive step discontinuities, *IEEE Trans. Signal Process.* 60 (December (12)) (2012) 6305–6318.
- [59] I.W. Selesnick, I. Bayram, Sparse signal estimation by maximally sparse convex optimization, *IEEE Trans. Signal Process.* 62 (March (5)) (2014) 1078–1092.
- [60] I.W. Selesnick, H.L. Graber, Y. Ding, T. Zhang, R.L. Barbour, Transient artifact reduction algorithm (TARA) based on sparse optimization, *IEEE Trans. Signal Process.* 62 (December (24)) (2014) 6596–6611.
- [61] I.W. Selesnick, H.L. Graber, S. Douglas, S. Pfeil, R.L. Barbour, Simultaneous low-pass filtering and total variation denoising, *IEEE Trans. Signal Process.* 62 (March (5)) (2014) 1109–1124.
- [62] B. Tang, S. Dong, T. Song, Method for eliminating mode mixing of empirical mode decomposition based on the revised blind source separation, *Signal Process.* 92 (1) (2012) 248–258.
- [63] D. Wipf, S. Nagarajan, Iterative reweighted  $\ell_1$  and  $\ell_2$  methods for finding sparse solutions, *IEEE J. Sel. Top. Signal Process.* 4 (April (2)) (2010) 317–329.
- [64] Y. Wongsawat, Efficient implementation of RMVB for eyeblink artifacts removal of EEG via STF-TS modeling, in: 2008 Proceedings of ROBIO, 2008, pp. 1567–1572.
- [65] X. Xie, Illumination preprocessing for face images based on empirical mode decomposition, *Signal Process.* 103 (0) (2014) 250–257.
- [66] M. Yaghoobi, T. Blumensath, M.E. Davies, Dictionary learning for sparse approximations with the majorization method, *IEEE Trans. Signal Process.* 57 (June (6)) (2009) 2178–2191.
- [67] J. Yan, L. Lu, Improved Hilbert–Huang transform based weak signal detection methodology and its application on incipient fault diagnosis and ECG signal analysis, *Signal Process.* 98 (2014) 74–87.
- [68] S. Yu, A.S. Khwaja, J. Ma, Compressed sensing of complex-valued data, *Signal Process.* 92 (2) (2012) 357–362.
- [69] H. Zeng, A. Song, R. Yan, H. Qin, EOG artifact correction from EEG recording using stationary subspace analysis and empirical mode decomposition, *Sensors* 13 (11) (2013) 14839–14859.