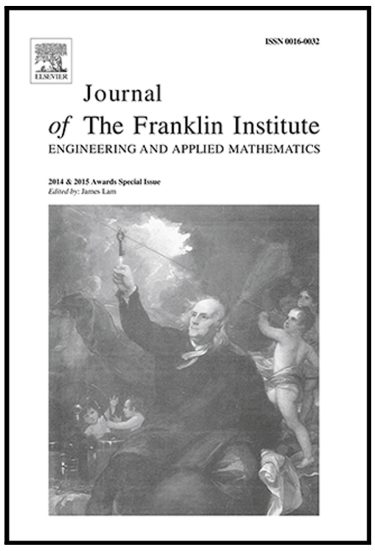

Mixed ℓ_2 and ℓ_1-norm regularization for adaptive detrending with ARMA modeling

L. Giarré, F. Argenti

PII: S0016-0032(18)30003-6
DOI: 10.1016/j.jfranklin.2017.12.009
Reference: FI 3260

To appear in: *Journal of the Franklin Institute*

Received date: 16 February 2017
Revised date: 27 October 2017
Accepted date: 12 December 2017

Please cite this article as: L. Giarré, F. Argenti, Mixed ℓ_2 and ℓ_1-norm regularization for adaptive detrending with ARMA modeling, *Journal of the Franklin Institute* (2018), doi: 10.1016/j.jfranklin.2017.12.009

Mixed ℓ_2 and ℓ_1-norm regularization for adaptive detrending with ARMA modeling

L. Giarré[a,*], F. Argenti[b]

[a] *Dipartimento di Ingegneria "Enzo Ferrari", Università di Modena e Reggio Emilia, Via Vivarelli 10, - 41125, Modena, Italy*
[b] *Dipartimento di Ingegneria dell'Informazione, Università di Firenze, Via di Santa Marta, 3 - 50139, Firenze, Italy*

Abstract

In this paper, the problem of detrending a time series and/or estimating a wandering baseline is addressed. We propose a new methodology that adaptively minimizes different regularized cost functions by introducing an ARMA model of the underlying trend. Mixed ℓ_1/ℓ_2-norm penalty functions are taken into consideration and novel RLS and LMS solutions are derived for the model parameters estimation. The proposed methods are applied to typical trend estimation/removal problems that can be found in the analysis of economic time series or biomedical signal acquisition. Comparisons with standard noncausal filtering techniques are also presented.

Keywords: Trend estimation, regularization, sparsity, ARMA models, RLS, LMS

1. Introduction

The problem of estimating an underlying trend in time series arises in a variety of disciplines, including macroeconomics [1][2], geophysics, financial analysis [3], social sciences, biological and medical sciences [4], environmental measurements modelling [5], or it can be also found in detecting structural changes in time series [6].

*Corresponding author
Email addresses: laura.giarre@unimore.it (L. Giarré), fabrizio.argenti@unifi.it (F. Argenti)

Detrending methods often rely upon optimization of a given functional that may include regularization, or penalty, terms. Such penalty functions are chosen in order to induce a desired behavior or properties into the solution. For example, if the regularization term is based on ℓ_2-norm a minimum energy solution is pursued. Such a solution is effective in many applications and is also convenient from a mathematical standpoint. Regularization based on the ℓ_1-norm, instead, induces a sparse solution and has become popular, with the growing interest in sparse representation modeling and compressed sensing, in several fields including statistics [7], signal processing [8] and machine learning [9].

Examples of regularization-based methods for detrending can be found in the field of macroeconomic time series analysis, where the Hodrick–Prescott algorithm [1] is the classically used. In this approach, a ℓ_2-norm constraint on the second derivative of the unknown trend is imposed. The method has been recently revisited in [10], where detrending is achieved by substituting, in the penalty function, the ℓ_2-norm with the ℓ_1-norm. In this way, the filtering method induces piecewise linear trend estimates and, therefore, is well suited to analyze time series characterized by an underlying trend satisfying such a model. Other examples of ℓ_2-norm regularization application arise in medical sciences. For instance, in electrocardiogram (ECG) acquisitions, one of the major disturbances [11] is the *baseline wandering* (BW), caused by patient movement and respiration, that appears as a random variation of the signal trend. An ℓ_2-norm regularization approach to BW removal, with first order derivative in the penalty term, was proposed in [12]. Algorithms in [1][10][12] present filters that are noncausal and can be used only offline.

In this paper, novel adaptive detrending methods, using a regularization approach, are proposed. The major difference with respect to the algorithms presented in [1][10][12] is that the proposed methods are based on *online* minimization of a regularized cost function, that is adaptation is performed each time a new sample of the signal is acquired instead than using a noncausal approach in which the whole signal is stored and then processed. In order to accomplish this task, the trend signal we would like to estimate is assumed to

follow an auto-regressive moving-average (ARMA) model, whose parameters are adaptively estimated via RLS or LMS techniques.

The problem is solved by using a general formulation of the regularized cost function. In fact, it uses a mixture of both ℓ_1 and ℓ_2-norm penalty functions and is able to cope with any order of the derivative of the trend. This unified framework allows several advantages to be achieved: with a proper selection of the weighting parameters, it includes the single ℓ_1 and ℓ_2-norm penalties as particular cases, whereas the choice of different derivative orders allows various models of the trend to be fit; furthermore, there are more degrees of freedom to induce specific properties - depending on each individual set of data and application - into the solution and the benefits of each single type of penalty can be possibly pursued.

Basic features of the proposed algorithms are also their computational efficiency and, thanks to the online implementation, the use of a limited amount of memory, since only a small number of past and present samples is necessary. Some partial results were preliminary presented in [13] and [14].

As to the use of the ℓ_1-norm, the proposed approach is different from sparse systems estimation methods, where the penalty is directly imposed on the filter coefficients vector, as in the LASSO algorithm, with application to compressive sensing [15], network identification [16][17], sparse channel estimation [18][19], and beamforming design [20].

The paper is organized as follows. In section 2, the cost functions used for the regularization are defined. In section 3, the proposed ARMA modeling of the trend is introduced. In section 4, the novel regularized LMS and RLS solutions are derived. Experimental results, both for synthetic and real data, are presented in Section 5, whereas some conclusions are drawn in Section 6.

2. Regularized least squares methods

Let $y[k]$, $k = 1, 2, \ldots, n$, be the acquired time series affected by a trend $q[k]$, $k = 1, 2, \ldots, n$. It is assumed that q is a lowpass signal that introduces slow variations (or trend) into the time series. The objective of a detrending

algorithm is that of estimating q from y. In some applications, the purpose is just the estimation of the trend, whereas in others it is considered as a disturbance to be removed from the signal, so that $y - q$ has the same shape of y, but a constant baseline. Consider the following penalized mean square error problem to estimate the trend:

$$\hat{q} = \arg\min_{q} J(q) = \arg\min_{q} \{\|y - q\|_2^2 + \mathcal{P}(q)\}, \tag{1}$$

where y and q are n-length column vectors and $\|\cdot\|_p$ is the ℓ_p norm of a vector. The first term is a fidelity term between the acquired time series and the unknown trend. The penalty term $\mathcal{P}(q)$ must be chosen in order to induce smoothness on the signal q. Several choices can be made for $\mathcal{P}(q)$, according also to *a priori* knowledge on the trend model. The choice of $\mathcal{P}(q)$ induces the use of a specific solver. In this paper, we will analyze the following penalty functions.

The ℓ_2-norm penalty $\mathcal{P}(q)$ is defined as

$$\mathcal{P}(q) = \mathcal{P}_{\ell_2}(q) = \lambda_2 \|\Delta^d q\|_2^2 \tag{2}$$

where $\Delta = 1 - z^{-1}$ is the derivative operator (with z^{-1} denoting the unit delay); d is the order of the difference operator, corresponding to a penalty based on the dth order derivative; λ_2 is a given positive constant. For $d = 2$, the solution coincides with Hodrick-Prescott filtering [1]. It is apparent that constraining the minimization with an ℓ_2-norm on the derivative of q induces a degree of smoothness on the estimated trend.

The ℓ_1-norm penalty function is defined as

$$\mathcal{P}(q) = \mathcal{P}_{\ell_1}(q) = \lambda_1 \|\Delta^d q\|_1, \tag{3}$$

It is well-known that a constraint on the ℓ_1-norm induces a *sparse* solution. For example, if we choose $d = 2$ then a trend with a sparse second-derivative, that is an approximately piecewise linear solution, is achieved.

A third choice for the penalty function is based on a mixed ℓ_2/ℓ_1-norm, defined as

$$\mathcal{P}(q) = \mathcal{P}_{\ell_{12}}(q) = \lambda_2 \|\Delta^{d_2} q\|_2^2 + \lambda_1 \|\Delta^{d_1} q\|_1, \tag{4}$$

The aim of this study is finding low-complexity online solutions to the above problems. One example of application is the implementation of baseline removal algorithms on-board of wearable ECG devices, where the computational power and available memory are limited. To achieve this purpose, the trend is supposed to fit an ARMA model and solutions to the problem of estimating its parameters from the observed signal, based on least means squares (LMS) and recursive least squares (RLS) approaches, are proposed.

3. ARMA modeling of the trend

The regularization problems introduced in Section 2 have been solved by assuming that the baseline (or trend) q can be obtained from the observed signal y by means of an ARMA model. Assume that

$$Q(z) = F(z)Y(z) = \frac{B(z)}{A(z)}Y(z), \tag{5}$$

where

$$\begin{aligned} B(z) &= \sum_{k=0}^{M} b_k z^{-k}, \\ A(z) &= 1 + \sum_{k=1}^{N} a_k z^{-k}, \end{aligned} \tag{6}$$

with M and N the orders of the MA and AR components of the model, respectively, and b_k, $k = 0, 1, \ldots, M$, and a_k, $k = 1, \ldots, N$, their parameters. Thus, the trend signal is given by

$$q[n] = \sum_{k=0}^{M} b_k y[n-k] - \sum_{k=1}^{N} a_k q[n-k] = \varphi^T[n]\theta, \tag{7}$$

where

$$\begin{aligned} \varphi[n] &= \left[y[n] \ \ldots \ y[n-M] \, q[n-1] \ \ldots \ q[n-N]\right]^T \\ \theta &= \left[b_0 \ \ldots \ b_M \ -a_1 \ \ldots \ -a_N\right]^T, \end{aligned} \tag{8}$$

In the above expressions, we omitted the presence of additive noise terms.

In order to express the penalty functions, let the causal dth-order difference operator Δ^d be expressed by

$$\Delta^d q = h_d * q, \tag{9}$$

where h_d is the impulse response of a causal filter having $d+1$ coefficients. For example, if $d = 1$ then $h_1 = \{1, -1\}$, whereas for $d = 2$ we have $h_2 = \{1, -2, 1\}$. Thus, the nth sample of the signal $\Delta^d q$, used in the penalty functions defined in Section 2, is given by

$$\Delta^d q[n] = h_d^T \begin{bmatrix} q[n] \\ q[n-1] \\ \cdots \\ q[n-d] \end{bmatrix} = h_d^T \begin{bmatrix} \varphi^T[n] \\ \varphi^T[n-1] \\ \cdots \\ \varphi^T[n-d] \end{bmatrix} \theta = \psi_d^T[n]\theta, \tag{10}$$

where (7) has been used and

$$\psi_d[n] = \big[\varphi[n]\, \varphi[n-1]\, \cdots\, \varphi[n-d]\big] h_d. \tag{11}$$

In order to find an adaptive and computationally low-cost solution to the regularization problems, recursive least square and least mean square algorithms can be devised. The online estimated vector of parameters $\hat{\theta}[n]$ is then used to achieve an estimation of the trend at time n as

$$\hat{q}[n] = \varphi^T[n]\hat{\theta}[n]. \tag{12}$$

4. LMS and RLS solutions

In this section, the adaptive least mean square (LMS) and recursive least squares (RLS) solutions to the trend estimation problem are described. The approach is formulated for the mixed ℓ_1/ℓ_2-norm penalty function, which comprises the solutions to the ℓ_2 and ℓ_1 problems as particular cases.

By using the problem statement in (1) and (4) and the definitions (7) and (10), derived from the ARMA modeling of the trend as described in Section 3,

the function to be minimized can be expressed as

$$\begin{aligned} J(q) &= \|y-q\|_2^2 + \lambda_2\|\Delta^{d_2}q\|_2^2 + \lambda_1\|\Delta^{d_1}q\|_1 \\ &= \sum_{i=0}^{n}(y[i]-\varphi^T[i]\theta)^2 + \lambda_2\sum_{i=0}^{n}(\psi_{d_2}^T[i]\theta)^2 + \lambda_1\sum_{i=0}^{n}|\psi_{d_1}^T[i]\theta| \\ &= J_1(\theta) \end{aligned} \tag{13}$$

For mathematical convenience, the two quadratic terms in (13) can be grouped together. Consider the following vectors

$$z[i] = \begin{bmatrix} y[i] & 0 \end{bmatrix}^T, \qquad z[i] \in \mathbb{R}^2, \tag{14}$$

$$\Upsilon[i] = \begin{bmatrix} \varphi[i] & \sqrt{\lambda_2}\psi_{d_2}[i] \end{bmatrix}, \qquad \Upsilon[i] \in \mathbb{R}^{(M+N+1)\times 2}, \tag{15}$$

$$\varepsilon[i] = z[i] - \Upsilon^T[i]\theta, \qquad \varepsilon[i] \in \mathbb{R}^2. \tag{16}$$

Thus, (13) can be rewritten as

$$J_1(\theta) = \sum_{i=0}^{n}\|\varepsilon[i]\|_2^2 + \lambda_1\sum_{i=0}^{n}|\psi_{d_1}^T[i]\theta| \tag{17}$$

It is apparent that, for $\lambda_1 = 0$, the cost function in (17) coincides with that of the $\mathcal{P}_{\ell_2}(q)$ problem, whereas, for $\lambda_2 = 0$, the second element of the vector $\varepsilon[i]$ is identically null and the cost function collapses to that of the $\mathcal{P}_{\ell_1}(q)$ penalty problem.

4.1. LMS solution

The classical LMS solution [21] coincides with a steepest descent algorithm where the cost function is simply the last observed error (without the presence of any regularization term), that is the first term in (13), given by

$$J_{LMS}(\theta) = (y[i]-\varphi^T[i]\theta)^2 \tag{18}$$

Regularized versions of the LMS algorithm have been introduced in order to promote some properties into the solution. In [22], for example, an ℓ_1-norm penalty function is used to promote sparsity into the adaptive filter coefficients. In this study, in which sparsity of the sequence $\Delta^{d_1}q$ is pursued, the following

regularized LMS (rLMS) cost function, coinciding with the last term in (17), is used:

$$J_{\text{rLMS}}(\theta) = \|\varepsilon[i]\|_2^2 + \lambda_1 |\psi_{d_1}^T[i]\theta|. \tag{19}$$

From the observation of (19), we can see that, unlike [22], the ℓ_1-norm is not applied to the parameter vector, but rather the cost function depends on the absolute value of a linear combination of the parameters to be optimized. The rule that minimizes (19) at each step coincides with the following parameter vector updating:

$$\begin{aligned}\hat{\theta}[i] &= \hat{\theta}[i-1] - \frac{\mu}{2}\nabla\big[\ \|\varepsilon[i]\|_2^2 + \lambda_1 |\psi_{d_1}^T[i]\hat{\theta}[i-1]|\ \big] \\ &= \hat{\theta}[i-1] + \mu\Upsilon[i](z[i] - \Upsilon^T[i]\hat{\theta}[i-1]) - \frac{\mu}{2}\nabla\big[\ \lambda_1 |\psi_{d_1}^T[i]\hat{\theta}[i-1]|\ \big],\end{aligned} \tag{20}$$

where the time index has been added to $\hat{\theta}$ and μ is the updating gain. As can be seen, $\nabla J_{\text{rLMS}}(\theta)$ has been evaluated by using the knowledge of $\hat{\theta}[i-1]$.

Following [23, 24, 25], where adaptive online estimators for sparse parameter vectors are derived, hereafter we use a subgradient analysis, which offers a substitute for the gradient when the minimum of convex, nondifferentiable functions [26] is searched. At any point where the convex function fails to be differentiable, there exist possibly many valid subgradient vectors. According to [25], we approximate it, at each step, with the subdifferential of the quantity $|\psi_{d_1}^T[i]\hat{\theta}[i-1]|$, that is

$$\nabla|\psi_{d_1}^T[i]\hat{\theta}[i-1]| = \text{sign}(\psi_{d_1}^T[i]\hat{\theta}[i-1])\psi_{d_1}[i] \tag{21}$$

Thus, substituting (21) into (20) and using (16) yields

$$\hat{\theta}[i] = \hat{\theta}[i-1] + \mu\Upsilon[i]\varepsilon[i] - \mu_1 \text{sign}(\psi_{d_1}^T[i]\hat{\theta}[i-1])\psi_{d_1}[i], \tag{22}$$

where $\mu_1 = \frac{\mu\lambda_1}{2}$.

4.2. RLS solution

In the classical recursive least squares algorithm, a forgetting factor is introduced to build a cost function that gives more weight to the last errors between the observed samples and the output of an adaptive filter.

Consider now the signal model proposed in this paper and consider the following cost function based on the observed samples of $\varepsilon[i]$:

$$J_{\text{rRLS}}[n] = \frac{1}{2}\sum_{i=1}^{n}\alpha^{n-i}\|\varepsilon[i]\|_2^2, \tag{23}$$

where α is the forgetting factor. The cost function in (23) has been denoted as regularized RLS (rRLS) since, according to the definition of $\varepsilon[i]$ in (16), it embeds the ℓ_2-norm penalty. Note that with $\alpha = 1$ the problem coincides with that in (1) with the penalty $\mathcal{P}_{\ell_2}(q)$ as defined in (2). A solution to the minimization problem of (23) can be found in [13].

In this paper, we generalize that result by adding a penalty derived from the ℓ_1-norm of $\Delta^{d_1}q[n]$ and characterized by a forgetting factor as well. Thus, we redefine $J_{\text{rRLS}}[n]$ as

$$J_{\text{rRLS}}[n] = \frac{1}{2}\sum_{i=1}^{n}\alpha^{n-i}\|\varepsilon[i]\|_2^2 + \lambda_1\sum_{i=1}^{n}\alpha^{n-i}|\psi_{d_1}^T[i]\theta|, \tag{24}$$

where we used (10) to express $\Delta^{d_1}q[n]$.

By using (15) and (16) it can be shown that the gradient of $J_{\text{rRLS}}[n]$ can be written as

$$\begin{aligned}\nabla J_{\text{rRLS}}[n] &= -\sum_{i=1}^{n}\alpha^{n-i}\Upsilon[i]z[i] + \sum_{i=1}^{n}\alpha^{n-i}\Upsilon[i]\Upsilon^T[i]\theta + \lambda_1\sum_{i=1}^{n}\alpha^{n-i}\nabla|\psi_{d_1}^T[i]\theta| \\ &= -r[n] + R[n]\theta + \lambda_1 g[n],\end{aligned} \tag{25}$$

where

$$r[n] = \sum_{i=1}^{n}\alpha^{n-i}\Upsilon[i]z[i] = \alpha r[n-1] + \Upsilon[n]z[n] \tag{26}$$

$$R[n] = \sum_{i=1}^{n}\alpha^{n-i}\Upsilon[i]\Upsilon^T[i] = \alpha R[n-1] + \Upsilon[n]\Upsilon^T[n] \tag{27}$$

$$g[n] = \sum_{i=1}^{n}\alpha^{n-i}\nabla|\psi_{d_1}^T[i]\theta| = \alpha g[n-1] + \nabla|\psi_{d_1}^T[n]\theta|. \tag{28}$$

Consider online updating of the parameters vector by adding a temporal index to θ. The null-gradient condition yields

$$R[n]\theta[n] = f[n], \tag{29}$$

where

$$f[n] = r[n] - \lambda_1 g[n] = \alpha f[n-1] + \Upsilon[n]z[n] - \lambda_1 \nabla|\psi_{d_1}^T[n]\theta[n-1]| \quad (30)$$

Note that, for computational feasibility, we made $f[n]$ dependent on $\theta[n-1]$.

Taking (29) at time $n-1$ and using (27) and (30) yields

$$(R[n] - \Upsilon[n]\Upsilon^T[n])\theta[n-1] = f[n] - \Upsilon[n]z[n] + \lambda_1 \nabla|\psi_{d_1}^T[n]\theta[n-1]| \quad (31)$$

Multiplying both sides of (31) by $R^{-1}[n]$ and using again (29) and (16) yields

$$\theta[n] = \theta[n-1] + R^{-1}[n]\Upsilon[n]\varepsilon[n] - \lambda_1 R^{-1}[n]\nabla|\psi_{d_1}^T[n]\theta[n-1]| \quad (32)$$

Let $P[n] = R^{-1}[n]$. By applying the matrix inversion lemma to $P[n] = (\alpha R[n-1] + \Upsilon[n]\Upsilon^T[n])^{-1}$, the following recursive computation of $P[n]$ from $P[n-1]$ can be achieved [27]:

$$P[n] = \alpha^{-1}P[n-1] - \alpha^{-1}\kappa[n]\Upsilon^T[n]P[n-1], \quad (33)$$

where

$$\kappa[n] = \alpha^{-1}P[n-1]\Upsilon[n](I_2 + \alpha^{-1}\Upsilon^T[n]P[n-1]\Upsilon[n])^{-1}. \quad (34)$$

By using (33) and (34), the parameters update formula (32) becomes

$$\theta[n] = \theta[n-1] + \kappa[n]\varepsilon[n] - \lambda_1 P[n]\nabla|\psi_{d_1}^T[n]\theta[n-1]| \quad (35)$$

and by using (21) we get

$$\theta[n] = \theta[n-1] + \kappa[n]\varepsilon[n] - \lambda_1 P[n]\text{sign}(\psi_{d_1}^T[n]\theta[n-1])\psi_{d_1}^T[n] \quad (36)$$

5. Experimental results

In this section, the results of some experimental tests carried out to evaluate the effectiveness of the proposed methods (implemented in Matlab) are presented. The tests use both synthetic and real data.

In the following, we focus on two applications of our methods, namely, financial time series trend estimation and biomedical signals (electrocardiograms,

ECG) baseline wandering removal. In the former case, the trend is just the information we would like to extract from the observed data for economic analysis purposes, whereas, in the latter problem, the trend is seen as a disturbance we would like to remove in order to facilitate medical diagnosis. Another relevant difference between the two cases is that, in some applications of ECG acquisition (see, e.g., [28]), detrending may need to be performed in real-time so that the algorithms should be fast and causal, while this could not be a strict requirement for financial data. As to the baseline wandering problem several methods and tools have been proposed in the literature, for instance: adaptive filters [29], discrete wavelet transform [30], empirical mode decomposition [31], quadratic variation reduction [32].

In this paper, the results achieved with the proposed methods are compared with those obtained with the Hodrick-Prescott algorithm [1, 33] (referred to as "HP" in the following) and the quadratic variation reduction (QVR) method [12] (here implemented as noncausal LTI filtering [34] and referred to as "QVRLTI" in the following). Both these approaches can be seen as particular cases of the optimization problem in (1) with the choice of the ℓ_2-norm penalty function given in (2) and with derivative order equal to either two or one for the Hodrick-Prescott and QVR algorithms, respectively. The major difference with the proposed methods is that the Hodrick-Prescott and QVR algorithms use both past and future samples to estimate the trend, so that, being noncausal, they are not a feasible solution for real-time applications. On the other hand, since they use a richer information, a better performance is expected and, therefore, they represent a good benchmark to validate our methods.

5.1. Synthetic data tests

In order to assess the performance of the proposed methods, synthetic signals characterized by a synthetic trend are generated. Let q and x be the trend and the no-trend signal, respectively, so that $y = x + q$ is the observed signal. The signals q and x have been generated so that y resembles as much as possible typical realizations of either financial time series or ECG data. In order to

evaluate the tracking capabilities of the proposed methods, some deterministic waveforms (square and triangular periodic waves) have been also used for the trend signal. In all the tests, the spectral occupancy of the superimposed trend q partially overlaps with that of the no-trend signal x.

Since the trend is known, the implemented methods can be compared in terms of the mean square error (MSE), defined as

$$MSE = \frac{1}{N_q} \sum_n (q[n] - \hat{q}[n])^2, \tag{37}$$

where q is the synthetically generated trend, $\hat{q}$ is its estimation obtained from a trend estimation algorithm and N_q is their length. Adaptive algorithms are evaluated at their convergence, i.e., the head portion of the trend was discarded.

The LMS algorithms proposed in the previous sections for the ℓ_2, ℓ_1 and mixed ℓ_1/ℓ_2 norm penalties are denoted as LMS-L2, LMS-L1 and LMS-12, in that order, whereas the RLS counterparts are denoted as RLS-L2, RLS-L1 and RLS-12.

As to the selection of the parameters necessary to run the algorithms, we made the following choices. The derivative order of the ℓ_1/ℓ_2-norm terms was selected according to the assumed model of the trend: for instance, for a piecewise linear trend, a sparse second derivative (and therefore $d_1 = 2$) is expected. In some cases, the choice followed that of already proposed algorithms: for example, in "HP" $d_2 = 2$ and in "QVRLTI" $d_2 = 1$. The selection of the weighting parameters was instead based on experimental trial-and-test: for each method, after choosing an interval of variation for the parameters λ_1 and λ_2, the mean square error (MSE) between the true and estimated trend was computed on a grid of values on those intervals; the values of the parameters giving rise to the minimum MSE were selected.

5.1.1. Deterministic waveforms for the trend

Aim of these set of experiments is evaluating the tracking capabilities of the proposed methods in response to piecewise constant and linear trends. In these tests, the trend q is a periodic, either square or triangular, wave.

Consider a periodic square wave with a changing amplitude and frequency. The trend signal is composed of $N_q = 32768$ samples: in the first half, the amplitude varies in (-2,2) and the period is 1280 samples; in the second half, the amplitude varies in (-1,1) and the period is 640 samples. The no-trend signal x is a stochastic process obtained by bandpass filtering a white process distributed as $\mathcal{N}(0,1)$. A portion of a realization of the observed signal y is shown in Fig. 1.

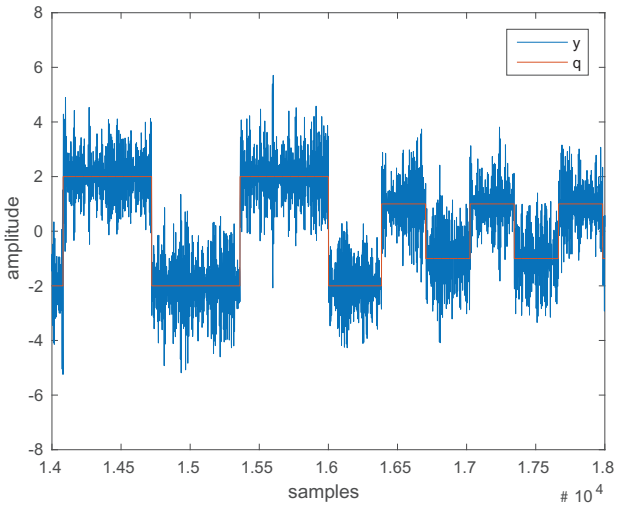

Figure 1: Synthetic pseudo-random signal with superimposed square wave trend.

In Fig. 2, the true trend q and $\hat{q}$, estimated by the proposed algorithms for a particular realization, are presented. The results shown in Fig. 2 were obtained with: order of the derivatives (see (4)) $d_1 = d_2 = 1$; orders of the ARMA model (see (6)) $M = 8$ and $N = 1$; value of the updating gain in the LMS algorithms $\mu = 5 \cdot 10^{-5}$; value of the forgetting factor in the RLS algorithms $\alpha = 0.999$. The values of the regularizing constants for the RLS algorithm were set to $\lambda_2 = 25$ for penalty $\mathcal{P}_{\ell_2}$, $\lambda_1 = 3$ for penalty $\mathcal{P}_{\ell_1}$ and $\lambda_2 = 15$, $\lambda_1 = 3$ for penalty $\mathcal{P}_{\ell_{12}}$, whereas for the LMS algorithms they were set to $\lambda_2 = 25$ for penalty $\mathcal{P}_{\ell_2}$, $\lambda_1 = 8$ for penalty $\mathcal{P}_{\ell_1}$ and $\lambda_2 = 15$, $\lambda_1 = 3$ for penalty $\mathcal{P}_{\ell_{12}}$. The corresponding MSE errors, evaluated by using (37) and averaged over fifty realizations, are reported

Table 1: MSE values for square and triangular wave trends.

Trend	RLS-L2	RLS-L1	RLS-12	LMS-L2	LMS-L1	LMS-12
square wave	0.0492	0.0524	0.0501	0.0314	0.0329	0.0312
triangular wave	0.0050	0.0042	0.0052	0.0047	0.0036	0.0038

in Table 1.

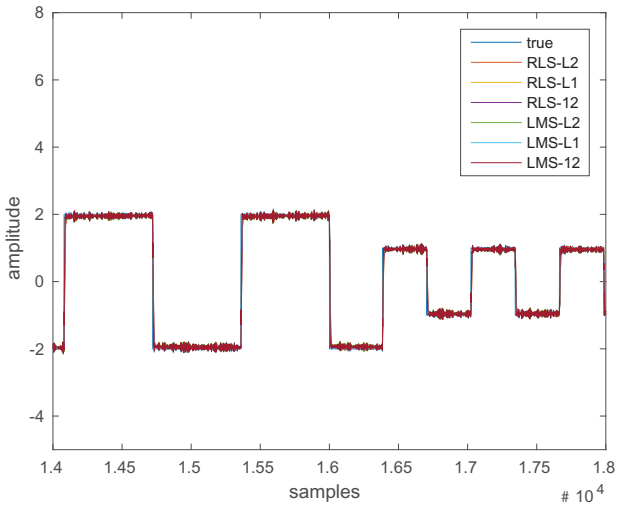

Figure 2: Trends $\hat{q}$ estimated with the proposed algorithms.

A periodic triangular wave was also used as synthetic trend. The no-trend signal was generated in the same way as in the square wave case previously described. A portion of a realization of the signal is depicted in Fig. 3. In the algorithms, the derivative parameters were set to $d_2 = 1$ and $d_1 = 2$; the latter was selected since the triangular wave possesses a sparse second derivative. The values of the regularizing constants for the RLS algorithm were set to $\lambda_2 = 25$ for penalty $\mathcal{P}_{\ell_2}$, $\lambda_1 = 3$ for penalty $\mathcal{P}_{\ell_1}$ and $\lambda_2 = 15$, $\lambda_1 = 1$ for penalty $\mathcal{P}_{\ell_{12}}$, whereas for the LMS algorithms they were set to $\lambda_2 = 30$ for penalty $\mathcal{P}_{\ell_2}$, $\lambda_1 = 5$ for penalty $\mathcal{P}_{\ell_1}$ and $\lambda_2 = 15$, $\lambda_1 = 3$ for penalty $\mathcal{P}_{\ell_{12}}$. All the other parameters were the same as in the square wave case. The trends estimated by the various algorithms are shown, for one realization, in Fig. 4, while the corresponding

MSE errors are reported in Table 1.

Figures 2 and 4 show that the proposed algorithms are able to correctly estimate the underlying trends. From Table 1, we can observe that best performance depends on the type of signals, i.e., the use of $\mathcal{P}_{\ell_2}$ and $\mathcal{P}_{\ell_1}$ are beneficial in the square and traingular wave cases, respectively, whereas the advantages of using a mixed norm are not always manifest.

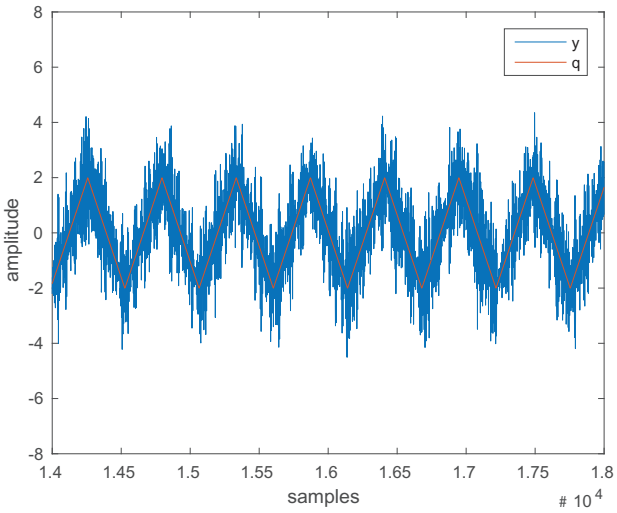

Figure 3: Synthetic pseudo-random signal with superimposed triangular wave trend.

5.1.2. Stochastic waveform for the trend

We consider two different ways to generate a stochastic trend. The first method creates a random piecewise linear trend, as suggested in [10], where time series resembling economic data were aimed at. The second assumes the trend to be a random lowpass signal as proposed in several papers on ECG baseline wandering removal algorithms, e.g., [12][31].

The first method to create a synthetic stochastic trend aims at generating a piecewise constant-slope signal according to the following expression:

$$q[n+1] = q[n] + v[n], \tag{38}$$

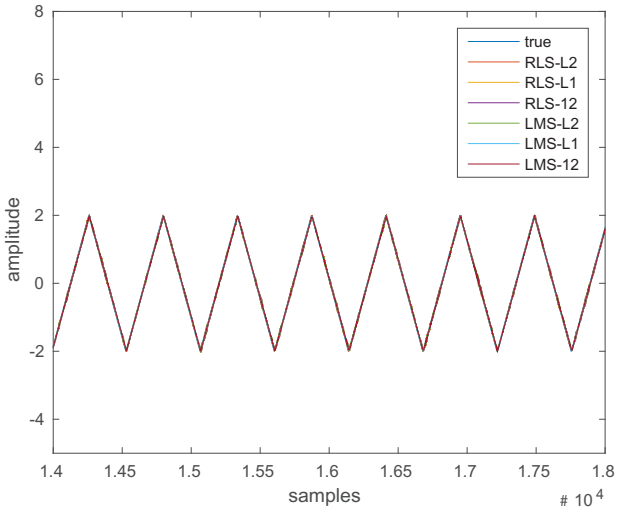

Figure 4: Trends $\hat{q}$ estimated with the proposed algorithms.

where $v[n]$ is the trend slope generated by a Markov chain, i.e., the slope remains unchanged ($v[n+1] = v[n]$) with probability p, whereas, with probability $1-p$, a new slope value is generated from a uniform distribution on $(-b, b)$. The length of q was $N_q = 20000$ samples. The observed time series was achieved by adding to the trend q a no-trend stochastic process obtained by bandpass filtering a white process distributed as $\mathcal{N}(0, \sigma_x)$. In the simulations we used $b = 2$, $p = 0.01$, and $\sigma_x = 2$.

The estimated trends are depicted in Fig. 5. These estimates were obtained by setting the order of the ARMA model to $M = 3$ and $N = 1$ and derivative orders to $d_2 = 1$ and $d_1 = 2$. The values of the regularizing constants for the RLS algorithm were set to $\lambda_2 = 4$ for penalty $\mathcal{P}_{\ell_2}$, $\lambda_1 = 3$ for penalty $\mathcal{P}_{\ell_1}$ and $\lambda_2 = 1$, $\lambda_1 = 2.5$ for penalty $\mathcal{P}_{\ell_{12}}$, whereas for the LMS algorithms they were set to $\lambda_2 = 70$ for penalty $\mathcal{P}_{\ell_2}$, $\lambda_1 = 140$ for penalty $\mathcal{P}_{\ell_1}$ and $\lambda_2 = 70$, $\lambda_1 = 10$ for penalty $\mathcal{P}_{\ell_{12}}$. The MSE values obtained averaging fifty realizations are reported in Table 2. To cope with the variability in the signal, the value of μ was set to 10^{-7} in the LMS algorithms. As to the parameter λ used for the HP and QVRLTI algorithm, the results presented here were obtained by choosing

Table 2: MSE values for stochastic piecewise constant-slope trend.

RLS-L2	RLS-L1	RLS-12	LMS-L2	LMS-L1	LMS-12	HP	QVRLTI
0.5283	0.4871	0.5006	0.5275	0.5659	0.5404	0.0374	0.3405

$\lambda = 1600$ and $\lambda = 100$ for the HP and the QVRLTI methods, respectively.

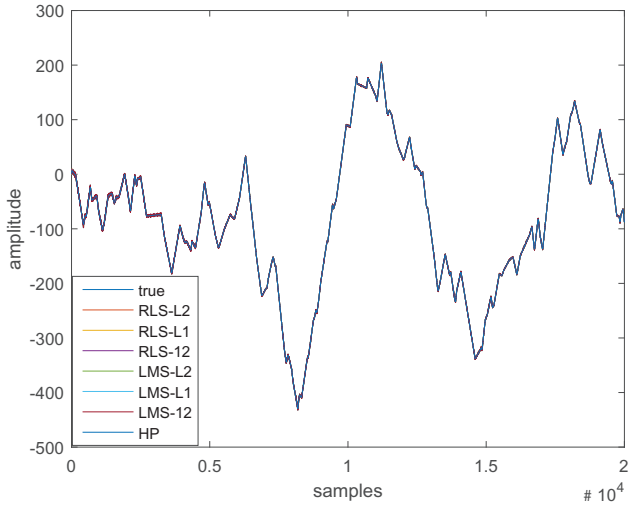

Figure 5: Trends $\hat{q}$ estimated with the proposed and HP algorithms.

A second dataset of synthetic signals was created to resemble the acquisitions from an electrocardiograph. The synthetic baseline-free ECG signal x was obtained by using the algorithm in [35] (a Matlab implementation of which is available in PhysioNet [36]). The data were generated by setting the heart rate to 60 bpm and the sampling frequency to $f_s = 256$ Hz. White Gaussian noise with standard deviation $\sigma_n = 0.01$ was also added. The output is an ECG-like signal normalized between -0.4 and 1.2 mV.

A pseudo-random synthetic baseline (the trend signal q) was then added. It was generated by lowpass filtering a white Gaussian process with a fourth-order Butterworth filter having a 3-dB cutoff frequency set to a given value f_t. The amplitude of the baseline was adjusted so that its standard deviation σ_b was equal to 0.5 mV.

Table 3: MSE values for stochastic low-pass trend.

RLS-L2	RLS-L1	RLS-12	LMS-L2	LMS-L1	LMS-12	HP	QVRLTI
0.0180	0.0271	0.0181	0.0290	0.0295	0.0291	0.0309	0.0204

The BW removal algorithms proposed in this paper were run by setting the order of the ARMA model to $M = 1$ and $N = 3$ and the derivative orders to $d_2 = 1$ and $d_1 = 1$. The values of the regularizing constants for the RLS algorithm were set to $\lambda_2 = 90$ for penalty $\mathcal{P}_{\ell_2}$, $\lambda_1 = 2$ for penalty $\mathcal{P}_{\ell_1}$ and $\lambda_2 = 90$, $\lambda_1 = 2$ for penalty $\mathcal{P}_{\ell_{12}}$, whereas for the LMS algorithm were set to $\lambda_2 = 80$ for penalty $\mathcal{P}_{\ell_2}$, $\lambda_1 = 11$ for penalty $\mathcal{P}_{\ell_1}$ and $\lambda_2 = 70$, $\lambda_1 = 1$ for penalty $\mathcal{P}_{\ell_{12}}$. The parameter λ from the HP and QVRLTI algorithms was set to $\lambda = 1600$ and $\lambda = 10^4$, respectively.

In Table 3, the MSE values obtained in the case of $f_t = 0.4$ Hz averaged over fifty realizations of the pseudo-random baseline are reported. A portion of a realization of the synthetic ECG signal is given in Fig. 6. In Fig. 7, examples of BW removal, obtained with the RLS-12 and the QVRLTI algorithms, are presented.

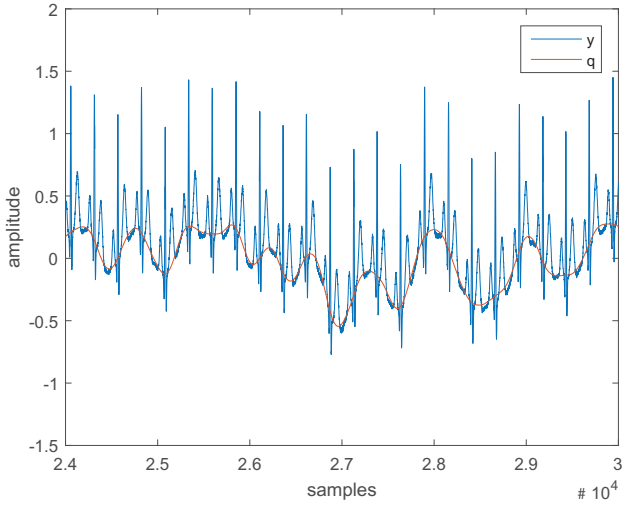

Figure 6: Sinthetic ECG with synthetic baseline.

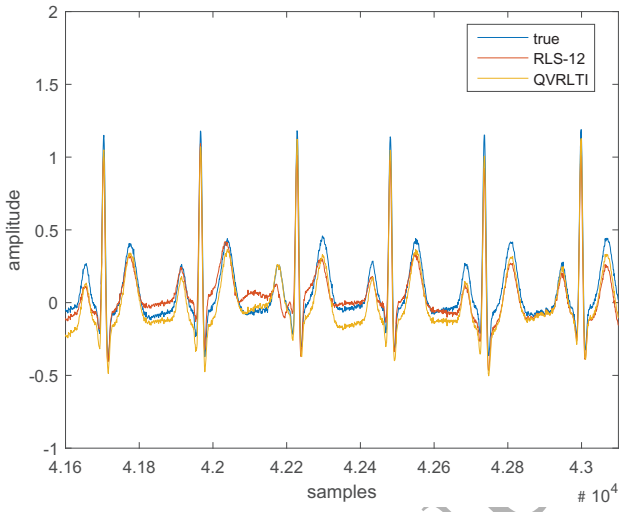

Figure 7: BW removal results.

Table 4: MSE values for LMS algorithms varying the μ parameter.

μ	LMS-L2	LMS-L1	LMS-12
5.e-5	0.0311	0.0333	0.0315
1.e-4	0.0290	0.0295	0.0291
5.e-4	0.0285	0.0366	0.0284
1.e-3	0.0282	0.0451	0.0282

From Tables 2 and 3, we observe only a slight (in most of the cases) degradation of the proposed methods with respect to the HP and QVR methods. Nonetheless, the underlying trend is well extracted by all the methods, as shown in Fig. 5, or the BW removal is effective, as shown in Fig. 7.

In Table 4, the MSE values obtained for the LMS algorithms varying the updating gains are shown. As can be seen, the average MSE is only slightly dependent on the selected gains.

5.1.3. Convergence and complexity

Many tests have been carried out to ascertain the convergence of the proposed methods, especially for the LMS algorithms, which are, in general, more critical at this regard.

In order to evaluate the rate of convergence, we computed the learning curves by averaging the cost defined in equation (19) over one thousand realizations for the LMS-12 method and various choices of the parameter μ. Such a method was selected because it comprises the LMS-L2 and LMS-L1 as special cases. In Fig. 8, focusing on the case of a piecewise linear stochastic trend, the curves have been plotted to demonstrate the influence of the parameter μ onto the convergence rate of the methods (all the other parameters were set as described in Section (5.1.2)). It is apparent that all the curves converge to about the same value with faster convergence relative to greater values, as expected.

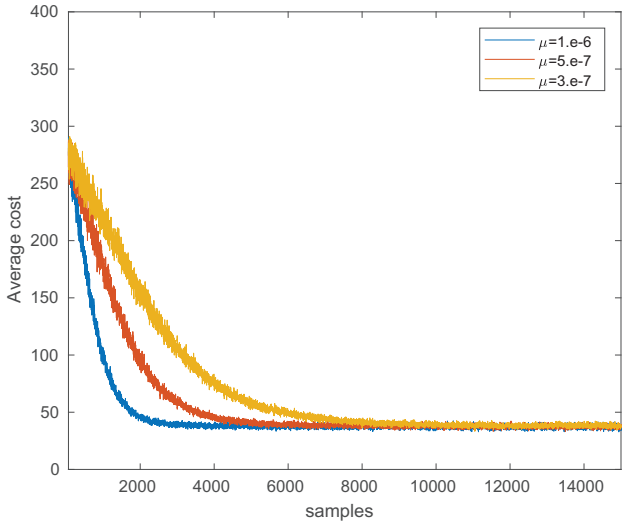

Figure 8: Learning curves relative to LMS algorithms (case of piecewise linear stochastic trend).

As to the complexity of the proposed methods, the cost of the LMS method is due to the computation of (22), which in turn depends on (11), whose order is $O(N_\theta \times (d+1))$, where $d = \max(d_1, d_2)$. The cost of the RLS method is due to the computation of (33) and (34) as well as of (36), whose order is $O(N_\theta^2)$, where $N_\theta = M + N + 1$ (assuming $d < N_\theta$).

5.2. Real data tests

In order to test the presented methods on real time series, we used two different types of dataset. The first type was the economic time series representing the Standard & Poor Dow-Jones index SP500, taken daily on a 10 years interval (from September 29th, 2006 to October 3rd, 2016) [37]. The second type of dataset were real ECG signals acquired by using a prototype of wearable ECG device developed in our laboratory. In both cases, since the real trend is unknown, the results were evaluated only by means of visual inspection.

5.2.1. Results obtained from the SP500 dataset

Fig. 9 plots the estimated trends obtained by using the RLS-12, LMS-12 and HP algorithms as well as the real data. These results were obtained by setting the order of the derivatives $d_1 = 1$ and $d_2 = 2$, $\lambda_1 = 50$, $\lambda_2 = 100$, $\mu = 10^{-8}$; the ARMA model was identified with $M = 3$ and $N = 1$. As can be seen, the trend is well estimated by the presented method, even though the noncausal HP method presents a smoother behavior.

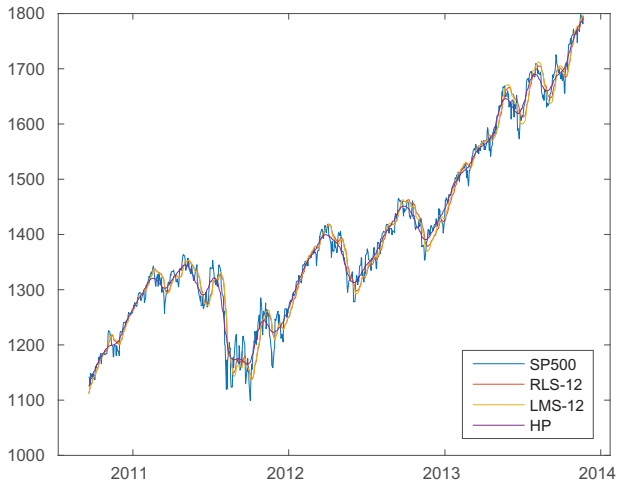

Figure 9: Real SP500 data and estimated trends with different methods.

5.2.2. Results obtained from real ECG dataset

A prototype of ECG acquisition device was developed in our laboratory [14, 13]. Its features are the following: acquisition of 3 ECG bipolar derivations (DI, DII, DIII) and 1 pre-cordial derivation (V1), by using 5 standard electrodes; analog front-end and ADC at 24 bit (Texas Instruments ADS1293), sampling frequency up to 25.6 ksps; micro-controller ARM STM32F411; storage onto microSD; transmission of ECG signals in real time by means of wireless Bluetooth 4.0 Low Energy (Nordic Semiconductor nRF8001) or by means of USB connection; PCB dimension of 44x60 mm; long duration battery with capacity of 1300 mAh; standard ECG connectors DIN, diameter 1.5mm.

In Fig. 10, an example of a real ECG signal acquired with the prototype device is shown as well as the results obtained after baseline removal by means of RLS-12, LMS-12, and QVRLTI algorithms. As to the proposed methods, we used ARMA orders $M = 1$ and $N = 3$, order of the derivatives $d_1 = d_2 = 1$, $\lambda_1 = 20$, $\lambda_2 = 200$, $\mu = 10^{-4}$. The methods were run in Matlab after importing the data. As can be seen, all the methods allow the baseline wandering to be cancelled, with the main difference that the proposed algorithms work online, whereas the QVRLTI method is noncausal.

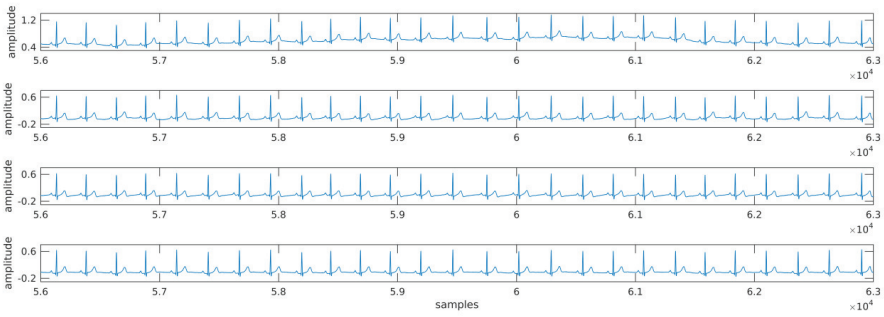

Figure 10: ECG acquired with the laboratory prototype (top plot) and BW removal results obtained with the RLS-12, LMS-12 and QVRLTI algorithms (from up to bottom, in that order).

5.3. Discussion

As already previously mentioned, the key features of the proposed methods are the online implementation and a limited usage of resources (memory and computation). This must be taken into consideration when our algorithms are compared to classical ones, performing offline and using the dataset in a non-causal fashion, that may possibly lead to better performance. Another issue in comparing the various algorithms stands in the fact that the cost to be minimized is different for each method, depending on the choice of several parameters (order of derivatives, weighting factors, etc.). For these reasons, we considered as metric of performance the MSE between true and estimated trend, which is computable only for synthetic datasets, whose order of magnitude depends also on the range of the signal amplitude.

As to the case of a deterministic trend, addressed to in Section 5.1.1, we notice that, despite the square wave signal presents abrupt changes and high-frequency components, all the proposed methods estimated quite well the trend, even when the model changes (see Fig. 2). The same holds for the case of the triangular wave trend, with lower MSE due to the higher regularity and narrower frequency content.

As to the case of a stochastic trend, addressed to in Section 5.1.2, we notice for ECG-like signals that the ℓ_2-norm yields a better MSE coherently with a nonsparse trend model, while for the piecewise-linear stochastic trend the ℓ_1-norm is better in the RLS case, coherently with a sparse trend model.

6. Conclusion

In this paper, we have presented some methods for trend estimation based on regularized cost functions. The basic features of the proposed algorithms are the ARMA modeling of the underlying trend, the use of ℓ_1 and ℓ_2-norm penalty functions (separately or jointly) as well as online estimation of model parameters. Novel RLS and LMS solutions to cope with the mixed penalty function have been derived. The methods are applied to typical trend estimation/removal

problems that can be found in the analysis of economic time series or biomedical signal acquisition. The proposed algorithms are also characterized by limited computational burden and memory requirements so that its implementation on low-cost devices is feasible. The presented experimental results, performed on synthetic and real data, validate the effectiveness of the proposed methods and compare them with classical regularized smoothing approaches that, however, use the whole set of data in a noncausal fashion.

Acknowledgments

This paper has been supported by the European Fund for Regional Development for the 20072013 programming period (POR FESR 2007-2013 CReO, ASSO Project). The authors would like to thank Dr Alessio Carpini for developing part of the firmware used in the aforementioned ECG prototype. The authors wish also to sincerely thank the anonymous reviewers for their constructive comments that helped them to improve the quality of the original submission.

References

[1] R. Hodrick, E. C. Prescott, Postwar U.S. business cycles: An empirical investigation, Journal of Money, Credit, and Banking 29 (1) (1997) 1–16.

[2] K. J. Singleton, Econometric issues in the analysis of equilibrium business cycle models, Journal of Monetary Economics 21 (23) (1988) 361–386.

[3] R. Tsay, Analysis of Financial Time Series, Wiley-Interscience, Hoboken, NJ, 2005.

[4] S. Greenland, M. Longnecker, Methods for trend estimation from summarized dose-response data, with applications to meta-analysis, Amer. J. Epidemiology 135 (11) (1992) 1301–1309.

[5] Y.-N. Jeng, T.-M. Yang, Y.-C. Cheng, A class of fast and accurate deterministic trend decomposition in the spectral domain using simple and sharp diffusive filters, Journal of the Franklin Institute 349 (6) (2012) 2065–2092.

[6] B. Wang, J. Sun, A. E. Motter, Detecting structural breaks in seasonal time series by regularized optimization, Vol. Safety, Reliability, Risk and Life-Cycle Performance of Structures & Infrastructures, Taylor & Francis Group, London, 2013, pp. 3621–3628.

[7] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society 58 (1) (1996) 267–288.

[8] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, SIAM Journal on Scientific Computing 20 (1) (1998) 33–61.

[9] S. Boyd, L. Vandenberghe, Convex optimization, Cambridge Univ. Press, 2004.

[10] S.-J. Kim, K. Koh, S. Boyd, D. Gorinevsky, ℓ_1 trend filtering, SIAM Review 51 (2009) 339–360.

[11] G. Clifford, ECG statistics, noise, artifacts, and missing data, in: G. Clifford, F. Azuaje, P. McSharry (Eds.), Advanced Methods for ECG Analysis, Artech House, London, 2006, Ch. 3, pp. 55–99.

[12] A. Fasano, V. Villani, L. Vollero, Baseline wander estimation and removal by quadratic variation reduction, in: International Conference on Engineering in Medicine and Biology Society, EMBC,, 2011, pp. 977–980.

[13] F. Argenti, L. Facheris, L. Giarré, Adaptive quadratic regularization for baseline wandering removal in wearable ecg devices, in: 2016 24th European Signal Processing Conference (EUSIPCO), 2016, pp. 1718–1722.

[14] F. Argenti, B. Bamieh, L. Giarré, Regularized lms methods for baseline wandering removal in wearable ecg devices, in: 2016 IEEE 55th Conference on Decision and Control (CDC), 2016, pp. 5029–5034.

[15] E. J. Candes, M. B. Wakin, An introduction to compressive sampling, IEEE Signal Processing Magazine 25 (2) (2008) 21–30.

[16] A. Chiuso, G. Pillonetto, A bayesian approach to sparse dynamic network identification, Automatica 48 (8) (2012) 1553 – 1565.

[17] D. Materassi, G. Innocenti, L. Giarré, M. Salapaka, Model identification of a network as compressing sensing, Systems & Control Letters 62 (8) (2013) 664 – 672.

[18] Y. Liu, C. Li, Z. Zhang, Diffusion sparse least-mean squares over networks, IEEE Transactions on Signal Processing 60 (8) (2012) 4480–4485.

[19] L. Liu, Y. Zhang, D. Sun, VFF l_1-norm penalised WL-RLS algorithm using DCD iterations for underwater acoustic communication, IET Communications 11 (5) (2017) 615–621.

[20] J. F. de Andrade, M. L. R. de Campos, J. A. Apolinrio, L_1-constrained normalized LMS algorithms for adaptive beamforming, IEEE Transactions on Signal Processing 63 (24) (2015) 6524–6539.

[21] V. Solo, X. Kong, Adaptive Signal Processing Algorithms: Stability and Performance, Prentice Hall, 1994.

[22] Y. Chen, Y. Gu, A. O. Hero, Sparse LMS for system identification, in: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, pp. 3125–3128.

[23] E. Eksioglu, A. Tanc, RLS algorithm with convex regularization, IEEE Signal Processing Letters 18 (8) (2011) 470–473.

[24] D. Angelosante, G. Giannakis, RLS-weighted Lasso for adaptive estimation of sparse signals, in: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2009, pp. 3245–3248.

[25] P. Garrigues, L. E. Ghaoui, An homotopy algorithm for the lasso with online observations, in: D. Koller, D. Schuurmans, Y. Bengio, L. Bottou (Eds.), Advances in Neural Information Processing Systems 21, Curran Associates, Inc., 2009, pp. 489–496.

[26] D. Bertsekas, A. Nedic, A. Ozdaglar, Convex Analysis and Optimization, Cambridge, MA: Athena Scientific, 2003.

[27] D. Manolakis, V. Ingle, S. Kogon, Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing, Artech House, 2005.

[28] F. Tay, D. Guo, L. Xu, M. Nyan, K. Yap, Memswear-biomonitoring system for remote vital signs monitoring, Journal of the Franklin Institute 346 (6) (2009) 531–542.

[29] P. Laguna, R. Jané, P. Caminal, Adaptive filtering of ECG baseline wander, in: 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 2, 1992, pp. 508–509.

[30] K. L. Park, K. J. Lee, H. R. Yoon, Application of a wavelet adaptive filter to minimise distortion of the ST-segment, Medical and Biological Engineering and Computing 36 (5) (1998) 581–586.

[31] M. Blanco-Velasco, B. Weng, K. E. Barner, ECG signal denoising and baseline wander correction based on the empirical mode decomposition, Computers in Biology and Medicine 38 (1) (2008) 1–13.

[32] A. Fasano, V. Villani, L. Vollero, Fast ECG baseline wander removal preserving the ST segment, in: Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies, ISABEL '11, ACM, New York, NY, USA, 2011, pp. 56:1–56:5.

[33] T. McElroy, Exact formulas for the Hodrick-Prescott filter, Econometrics Journal 11 (2008) 209–217.

[34] A. Fasano, V. Villani, Baseline wander removal in ECG and AHA recommendations, in: Computing in Cardiology Conference (CinC), 2013, 2013, pp. 1171–1174.

[35] P. McSharry, G. Clifford, L. Tarassenko, L. Smith, A dynamical model for generating synthetic electrocardiogram signals, IEEE Transactions on Biomedical Engineering 50 (3) (2003) 289–294.

[36] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, H. E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) e215–e220.

[37] http://us.spindices.com/indices/equity/sp-500, availabe on line.