



Online Filtering Using Piecewise Smoothness Priors: Application to Normal and Abnormal Electrocardiogram Denoising

Reza Sameni

► To cite this version:

Reza Sameni. Online Filtering Using Piecewise Smoothness Priors: Application to Normal and Abnormal Electrocardiogram Denoising. Signal Processing, 2016, 10.1016/j.sigpro.2016.10.019 . hal-01352281v3

HAL Id: hal-01352281

<https://hal.science/hal-01352281v3>

Submitted on 30 Oct 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Online Filtering Using Piecewise Smoothness Priors: Application to Normal and Abnormal Electrocardiogram Denoising

Reza Sameni

School of Electrical & Computer Engineering, Shiraz University, Shiraz, Iran
E-mail: rsameni@shirazu.ac.ir, Web: www.sameni.info, Phone: +98 713 6133169, Fax: +98 713 6473367

Abstract

In this work, a block-wise extension of Tikhonov regularization is proposed for denoising smooth signals contaminated by wide-band noise. The proposed method is derived from a constrained least squares problem in two forms: 1) a block-wise fixed-lag smoother with smooth inter-block transitions applied in matrix form, and 2) a fixed-interval smoother applied as a forward-backward zero-phase filter. The filter response is maximally flat and monotonically decreasing, without any ripples in its pass-band. The method is also extended to smoothness of multiple smoothness orders, and its relationship with Lipschitz regularity and block-wise Wiener smoothing is also studied.

The denoising of normal and abnormal electrocardiogram (ECG) signals in different stationary and non-stationary noise levels is studied as case study. While most ECG denoising techniques benefit from the pseudo-periodicity of the ECG, the developed technique is merely based on the smoothness assumption, which makes it a powerful method for both normal and abnormal ECG. The performance of the method is assessed by Monte-Carlo simulations over three standard normal and abnormal ECG databases of different sampling rates, in comparison with bandpass filtering, wavelet denoising with various parameters, and Savitzky-Golay filters using *Stein's unbiased risk estimate* shrinkage scheme.

Keywords: Tikhonov regularization, forward-backward filtering, electrocardiogram filtering, wavelet denoising, Lipschitz regularity, Wiener smoothing.

1. Introduction

Signal denoising is a prominent branch of signal processing, which is based on a *mixing model* and *prior assumptions* on the signal and noise structure. While the most widely studied mixing models are additive mixtures of signal and noise, the signal/noise priors are totally subjective. Deterministic vs. stochastic, temporal vs. frequency domain or spatial, static vs. dynamical, linear vs. periodic, are among the different priors used for designing denoising schemes [1].

In absence of reliable priors regarding the signal and noise structure, generic filtering techniques such as bandpass filtering, wavelet denoising [2], Tikhonov

regularization and smoothing splines [3, 4], Savitzky-Golay filters [5], *Stein's unbiased risk estimate* (SURE) regularization [6], blind or semi-blind source separation (for multi-channel signals) [7], or even empirical techniques [8] have been used, which are merely based on prior assumptions such as signal smoothness, time or transform-domain sparsity, or stochastic independence.

Specifically, Tikhonov regularization has found great interest in various fields of science and engineering (including biomedical engineering) for smoothing noisy data and signal detrending [9, 4, 10]. Despite its simplicity and good performance over short stationary data blocks, Tikhonov regularization becomes computationally demanding for long data. Moreover, it is an offline procedure and does not adapt to signal and noise non-stationarities.

Herein, a generic filtering scheme is proposed based on smoothness priors and a constrained least squares problem, which extends Tikhonov regularization to a block-wise formulation that is computationally efficient

This is the author's version of the manuscript accepted for publication in Elsevier Signal Processing, October 2016. The paper can be found in its final format at:
<http://dx.doi.org/10.1016/j.sigpro.2016.10.019>

and applicable to nonstationary scenarios. Since the method is applied over rather short segments of data, it is shown that it is computationally more efficient than the classical Tikhonov regularization and it can be implemented as an online fixed-lag smoother with smooth inter-segment transitions, for both stationary and non-stationary mixtures of signal and noise. The smoother is also studied in the frequency domain and is shown to be equivalent to an adaptive filter, which tunes its bandwidth according to the signal/noise contents of each data segment. It is shown that the proposed method bridges between various concepts and filtering schemes, including smoothing splines, Lipschitz regularization [11], Wiener filtering, and closely related SURE regularization [6], and Savitzky-Golay filters [5].

The denoising of Electrocardiogram (ECG) signals is used as a case study for the proposed method. The problem of ECG denoising is of major importance in biomedical signal processing. Due to the large diversity of ECG waveforms, the current gold standard for ECG acquisition system front-ends are based on rather simple bandpass filters [12]. More effective techniques for ECG denoising, such as wavelet denoising, are based on the concept of wavelet shrinkage using signal smoothness priors [13]. Considering the broad range of parameters, which can be used for wavelet denoising, various studies have been carried-out to study the best set of wavelet denoising parameters for ECG signals [14]. More recently, Savitzky-Golay filters using the SURE for optimal order selection have also been used for ECG denoising and compared with various wavelet-based schemes [15, 16].

More advanced (yet sophisticated) methods for ECG denoising have used the pseudo-periodicity of the ECG for constructing model-based filters using the extended Kalman filter and its extensions [17, 18, 19]. The idea has been further utilized for ECG segmentation [20], decomposition [21], anomaly detection [22], fiducial point extraction [23], and fetal ECG extraction [24, 25]. Despite the effectiveness of these highly model-based filtering schemes (and their outperformance versus wavelet denoising [18]) they are mainly applicable to normal ECG with pseudo-periodic behavior and their performance degrades in cardiac anomalies with highly abnormal beat morphologies. However, in the hereby studied case studies, it is shown that the proposed method, which is merely based on smoothness priors (similar to wavelet denoising), is equally applicable to normal and abnormal ECG contaminated by broad-band noise, without requiring the ECG R-peaks. Due to its minimal assumptions on the ECG and noise structure, it can be potentially used in a broad range

of ECG-based applications, even in very low signal-to-noise ratios (SNR).

The organization of the paper is as follows. In Section 2, a piecewise smooth matrix form of the proposed algorithm together with the parameter selection details are presented. The proposed scheme is related to the concept of Lipschitz regularity in Section 3. In Section 4 using a sample-wise forward-backward smoothing formulation, a frequency domain interpretation is presented, which gives a Wiener smoothing perspective to Tikhonov regularization and the proposed method. A generalization of the developed framework to multiple-order smoothness is presented in Section 5; followed by a detailed case study for ECG denoising, in Section 6. The paper ends with a summary and promising future directions, in Section 7.

2. Piecewise smooth Tikhonov regularization

2.1. Discrete smoothing operator

Consider \mathbf{d}_d , a length M finite impulse response approximation for the d th order derivative operator. As a simple realization of \mathbf{d}_d , one can use the following recursion, which is based on the first order difference approximation of the derivative operator:

$$\mathbf{d}_d = \mathbf{d}_{d-1} * \mathbf{d}_1, \quad d \geq 2 \quad (1)$$

where $\mathbf{d}_1 \triangleq (1, -1)$ is the first-order difference impulse response and $*$ is the convolution operator. Throughout the paper, the second-order difference impulse response $\mathbf{d}_2 = (1, -2, 1)$ is of special interest.

For integers $T > M$, $\tilde{\mathbf{D}}_d \in \mathbb{R}^{(T-M+1) \times T}$ is defined as the Toeplitz matrix form of \mathbf{d}_d . As a special case we have

$$\tilde{\mathbf{D}}_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix} \quad (2)$$

The left multiplication of $\tilde{\mathbf{D}}_d$ in any column-wise signal of length T is equivalent to the convolution of the signal with \mathbf{d}_d , which is its d th order approximate derivative in our case. The effectiveness of finite difference approximations of the derivative depends on the sampling frequency of the data and the desired accuracy. In practice, (2) can be replaced by discrete-time differentiators with a desired impulse response or even a weighted difference (cf. [5] for a Savitzky-Golay perspective).

2.2. Data model

Let us consider a signal $\theta(t)$ with a decaying Fourier transform $\Theta(\omega)$, i.e.,

$$\int_{-\infty}^{+\infty} |\Theta(\omega)|(1+|\omega|^\nu)d\omega < +\infty \quad (3)$$

where $\nu > 0$ determines the rate of decay. Equation (3) implies that $\theta(t)$ is globally $d = \lfloor \nu \rfloor$ times *continuously differentiable* (cf. [11, Theorem 2.5]).

Further consider $\theta[m] = \theta(mT_s)$ as discrete-time samples of $\theta(t)$, $x[m]$ ($0 \leq m < L$) as noisy samples of $\theta[m]$ contaminated by uncorrelated additive noise, and a strictly monotone sequence of user-defined (or automatically selected) knot points $\mathcal{K} = \{p_0, p_1, \dots, p_N\}$, $p_k \in [0, L]$. The knots are used to segmentize the noisy signal into N non-overlapping segments. Accordingly, the segment \mathcal{S}_k ($0 \leq k < N$) is defined in vector form as

$$\mathbf{x}_k \triangleq (x[p_k], x[p_k + 1], \dots, x[p_{k+1} - 1])^T \quad (4)$$

with length $n_k \triangleq p_{k+1} - p_k$. We fix $p_0 = 0$ and $p_N = L$ to assure that each sample of $x[m]$ falls in exactly one segment.

In each segment \mathcal{S}_k , the following signal model is considered

$$\mathbf{x}_k = \boldsymbol{\theta}_k + \boldsymbol{\eta}_k \quad (5)$$

where $\boldsymbol{\theta}_k \in \mathbb{R}^{n_k}$ is the noiseless (desired) signal, and the noise vector $\boldsymbol{\eta}_k \in \mathbb{R}^{n_k}$ is assumed to be uncorrelated with $\boldsymbol{\theta}_k$ and zero-mean with variance $\boldsymbol{\eta}_k^T \boldsymbol{\eta}_k / n_k = \delta_k^2$ (a segment dependent variance).

In order to consider the inter-segment boundary behavior of the proposed method, two vectors containing q neighboring samples of \mathcal{S}_k from segments \mathcal{S}_{k-1} and \mathcal{S}_{k+1} are defined:

$$\begin{aligned} \boldsymbol{\alpha}_k &\triangleq (\theta[p_k - q], \theta[p_k - q + 1], \dots, \theta[p_k - 1])^T \in \mathbb{R}^q \\ \boldsymbol{\beta}_k &\triangleq (\theta[p_{k+1}], \theta[p_{k+1} + 1], \dots, \theta[p_{k+1} + q - 1])^T \in \mathbb{R}^q \end{aligned} \quad (6)$$

Exceptionally, for \mathcal{S}_0 and \mathcal{S}_{N-1} , which do not have any left and right neighbors, respectively, $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_{N-1}$ are defined to be empty vectors.

The knot-sequence and segment boundaries are illustrated in Fig. 1. Accordingly, the augmented vector

$$\bar{\boldsymbol{\theta}}_k^T \triangleq [\boldsymbol{\alpha}_k^T \mid \boldsymbol{\theta}_k^T \mid \boldsymbol{\beta}_k^T] \quad (7)$$

contains $(n_k + 2q)$ successive samples of the desired signal from the end of \mathcal{S}_{k-1} to the beginning of \mathcal{S}_{k+1} . In the following, we set $q = M - 1$ to guarantee smooth transitions between successive segments.

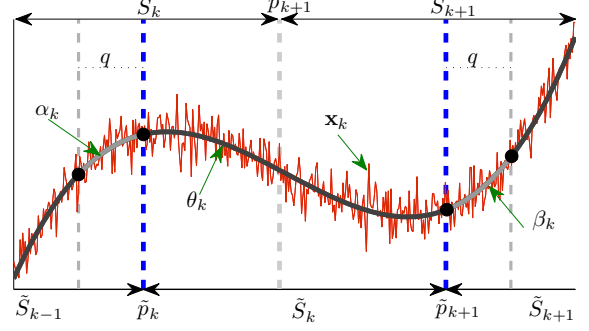


Figure 1: Illustration of the data model for each segment and its boundaries

Using these definitions, the objective is to find the least square error (LSE) estimate of $\boldsymbol{\theta}_k$ with controllable degree of smoothness over \mathcal{S}_k and its boundaries with neighboring segments. The problem can be formulated as a LSE problem with an inequality constrain:

$$\hat{\boldsymbol{\theta}}_k = \arg \min_{\boldsymbol{\theta}_k} \|\mathbf{x}_k - \boldsymbol{\theta}_k\|^2, \text{ s.t. } \|\tilde{\mathbf{D}}_d \bar{\boldsymbol{\theta}}_k\|^2 \leq \epsilon_k^2 \quad (8)$$

where $\tilde{\mathbf{D}}_d \in \mathbb{R}^{(n_k+M-1) \times (n_k+2M-2)}$ is the order d difference operator, $\|\cdot\|$ represents the L2-norm and ϵ_k^2 is a pre-defined upper bound for the signal's "steepness". The Lagrangian form of (8) is as follows:

$$\hat{\boldsymbol{\theta}}_k = \arg \min_{\boldsymbol{\theta}_k} \left\{ \|\mathbf{x}_k - \boldsymbol{\theta}_k\|^2 + \lambda_k \|\tilde{\mathbf{D}}_d \bar{\boldsymbol{\theta}}_k\|^2 \right\} \quad (9)$$

where λ_k is a positive regularization factor that is tuned to minimize the cost, while satisfying the inequality constraint in (8).

Equation (9) is in the form of Tikhonov regularization problems [3, 9], and similar to the regularized SURE cost function [6]. As noted in [9], a nice interpretation for the constrained LSE is to draw the unconstrained LSE solution toward the null space of the matrix $\tilde{\mathbf{D}}_d$, resulting in a smooth estimate of $\boldsymbol{\theta}_k$.

If $\tilde{\mathbf{D}}_d$ is partitioned into $\tilde{\mathbf{D}}_d = [\mathbf{E}_d \mid \mathbf{D}_d \mid \mathbf{F}_d]$, where $\mathbf{E}_d, \mathbf{F}_d \in \mathbb{R}^{(n_k+M-1) \times (M-1)}$ and $\mathbf{D}_d \in \mathbb{R}^{(n_k+M-1) \times n_k}$, the argument of (9) can be expanded as

$$J_k = \|\mathbf{x}_k - \boldsymbol{\theta}_k\|^2 + \lambda_k \|\mathbf{D}_d \boldsymbol{\theta}_k + \mathbf{b}_k\|^2 \quad (10)$$

where $\mathbf{b}_k \triangleq \mathbf{E}_d \boldsymbol{\alpha}_k + \mathbf{F}_d \boldsymbol{\beta}_k$ is a *boundary vector* independent of $\boldsymbol{\theta}_k$. Equation (10) is in the form of a classical LSE problem with a quadratic constraint. It is straightforward to show that the solution is as follows

$$\hat{\boldsymbol{\theta}}_k = (\mathbf{I} + \lambda_k \mathbf{D}_d^T \mathbf{D}_d)^{-1} [\mathbf{x}_k - \lambda_k \mathbf{D}_d^T \mathbf{b}_k] \quad (11)$$

In the special case that the boundary smoothness constraint is not required (the boundary conditions are not

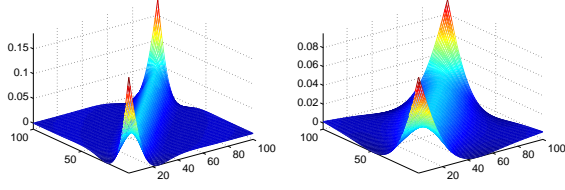


Figure 2: The *smoothing matrix* of Tikhonov regularization, for $T = 100$ and $\lambda = 50$ (left) and $\lambda = 200$ (right). This matrix has a local effect on the intermediate samples, when left-multiplied in a signal vector.

available or the data has only one segment), $\mathbf{b}_k = \mathbf{0}$ and (11) reduces to the Tikhonov regularization method used in previous studies [9]:

$$\hat{\boldsymbol{\theta}}_k = (\mathbf{I} + \lambda_k \mathbf{D}_d^T \mathbf{D}_d)^{-1} \mathbf{x}_k \quad (12)$$

We refer to $(\mathbf{I} + \lambda_k \mathbf{D}_d^T \mathbf{D}_d)^{-1}$, as the *smoothing matrix*. This matrix is shown in Fig. 2 for $T = 100$ and $\lambda = 50, 200$.

In Appendix A, it is shown that an exactly similar solution is obtained by seeking “*maximally smooth estimates under an upper bound on the noise variance of each segment*,” using a segment-wise regularization factor $\gamma_k = 1/\lambda_k$.

For the hereby proposed method (and the alternative formulation stated in Appendix A), there are practical considerations regarding the selection of the regularization factors λ_k (or γ_k), and the boundary vectors $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$, which are discussed below.

2.3. Regularization factor selection

Depending on whether the upper bound of signal roughness ϵ_k is known or not, the regularization factor λ_k can be fixed or calculated per-segment. Both cases are studies below.

2.3.1. Known (presumed) smoothness bound

When the smoothness bound is presumed or known (e.g., from prior knowledge of the signal form), an optimal λ_k can be calculated using ϵ_k . For this, the singular value decomposition (SVD) of \mathbf{D}_d is performed:

$$\mathbf{D}_d = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T \quad (13)$$

where $\mathbf{U} \in \mathbb{R}^{(n_k+M-1) \times (n_k+M-1)}$ and $\mathbf{V} \in \mathbb{R}^{n_k \times n_k}$ are orthonormal matrices and $\boldsymbol{\Sigma} \in \mathbb{R}^{(n_k+M-1) \times n_k}$ is a diagonal matrix, with diagonal entries $(\sigma_1, \sigma_2, \dots, \sigma_{n_k})$. Next, letting

$$\tilde{\boldsymbol{\theta}}_k \triangleq \mathbf{V}^T \hat{\boldsymbol{\theta}}_k, \quad \tilde{\mathbf{x}}_k \triangleq \mathbf{V}^T \mathbf{x}_k, \quad \tilde{\mathbf{b}}_k \triangleq \mathbf{U}^T \mathbf{b}_k, \quad (14)$$

it is straightforward to show that the block solution in (11) reduces to a sample-wise update:

$$\tilde{\theta}_{ik} = \frac{\tilde{x}_{ik} - \lambda_k \sigma_i \tilde{b}_{ik}}{1 + \lambda_k \sigma_i^2}, \quad (i = 1, \dots, n_k) \quad (15)$$

where $\tilde{\theta}_{ik}$, \tilde{x}_{ik} , and \tilde{b}_{ik} are the i th elements of the n_k -vectors $\tilde{\boldsymbol{\theta}}_k$, $\tilde{\mathbf{x}}_k$, and $\tilde{\mathbf{b}}_k$, respectively. Moreover, the mean square error and the smoothness bound are

$$e(\lambda_k) = \|\mathbf{x}_k - \hat{\boldsymbol{\theta}}_k\|^2 = \|\tilde{\mathbf{x}}_k - \tilde{\boldsymbol{\theta}}_k\|^2 = \sum_{i=1}^{n_k} \left(\frac{\sigma_i^2 \tilde{x}_{ik} + \sigma_i \tilde{b}_{ik}}{1/\lambda_k + \sigma_i^2} \right)^2 \quad (16)$$

$$c(\lambda_k) = \|\mathbf{D}_d \hat{\boldsymbol{\theta}}_k + \mathbf{b}_k\|^2 = \|\boldsymbol{\Sigma} \tilde{\boldsymbol{\theta}}_k + \tilde{\mathbf{b}}_k\|^2 = \sum_{i=1}^{n_k} \left(\frac{\sigma_i \tilde{x}_{ik} + \tilde{b}_{ik}}{1 + \lambda_k \sigma_i^2} \right)^2 \quad (17)$$

It is seen that $e(\lambda_k)$ and $c(\lambda_k)$ are monotonically increasing and decreasing with λ_k , respectively. Therefore, the minimum of (8) is obtained on the boundary of its constraint region.

When the smoothness bound ϵ_k is known (or assumed), the optimal regularization factor, denoted by λ_k^* , is numerically calculated by solving the following equation for λ_k^* :

$$c(\lambda_k^*) = \sum_{i=1}^{n_k} \left(\frac{\sigma_i \tilde{x}_{ik} + \tilde{b}_{ik}}{1 + \lambda_k^* \sigma_i^2} \right)^2 = \epsilon_k^2 \quad (18)$$

Due to the positive and monotonic decay of (18) toward zero by increase of λ_k^* , it has a unique solution if $\epsilon_k^2 \leq c(0)$. In order to find λ_k^* , we use a variant of the Bisection algorithm [26], detailed in Appendix B. This optimal value is next used in (11) or (15) to calculate $\boldsymbol{\theta}_k$. For $\epsilon_k^2 > c(0)$, the boundary of the constrain region is no longer achieved and $\lambda_k^* = 0$, resulting in $\hat{\boldsymbol{\theta}}_k = \mathbf{x}_k$.

2.3.2. Unknown smoothness bound

When ϵ_k is unknown, or no assumption is available on its upper bound (or equivalently on the noise variance δ_k^2 , as shown in Appendix A), methods such as the L-curve have been used for finding optimal regularization factors [27, 28, 29, 30, 4]. In [31], a maximum a posteriori (MAP) interpretation has been presented for the constrained least squares problem and its optimal regularization factor selection using the L-curve. In our case, the L-curve is the plot of $\|\mathbf{x}_k - \boldsymbol{\theta}_k\|$ versus $\|\mathbf{D}_d \boldsymbol{\theta}_k + \mathbf{b}_k\|$, as λ is swept from zero to infinity. The λ corresponding to the corner of the L-curve (the point at which $\|\mathbf{x}_k - \boldsymbol{\theta}_k\| + \|\mathbf{D}_d \boldsymbol{\theta}_k + \mathbf{b}_k\|$ is minimized) can be empirically chosen as the optimal regularization factor, which provides the best balance between the signal

smoothness and minimum mean square error. The results of this method versus the aforementioned method, which is based on prior knowledge of the signal noise variance are detailed in the results section.

Depending on the problem, one may have additional priors regarding the signal/noise structure, which can improve the regularization factor estimate. For instance, if the signal and noise are both nonstationary from one block to another (like the ECG example studied in Section 6.6), but the noise is known to have rather slow inter-block variations, the regularization factor λ_k^* can be made more consistent and its inter-segment fluctuations can be damped by using the following block-wise updates:

$$\tilde{\lambda}_k^* = \kappa \lambda_k^* + (1 - \kappa) \tilde{\lambda}_{k-1}^* \quad (19)$$

where $0 < \kappa \leq 1$ is a forgetting factor, which adds a degree of freedom for block-wise smoothing of λ_k^* . This approach is further justified in Section 4, where we present a more physical interpretation for λ_k^* (beyond a simple regularization factor).

2.4. Knot sequence selection

In theory, the knot sequence \mathcal{K} is basically arbitrary. In practice, in order to minimize the impact of smoothing at the segment boundaries, one can customize the knot sequence on a signal-dependent basis. For instance, for ECG signals, one can detect the fiducial points of the ECG and set the knots at intermediate points, which do not convey major clinical information. Alternatively, for online filtering, the knots can be equally spaced throughout the signal. Our experiments over biosignals such as the ECG, show that by introducing the segment smoothness margins α_k and β_k in (11), the boundary effects are negligible, and the fixed spaced knots is more appropriate in practice (imposing no processing overhead).

2.5. Boundary conditions

Recalling (11), it is seen that for estimating the signal in each segment, the noiseless samples (or an estimate) of the boundaries of its left (α_k) and right (β_k) segments have been assumed to be known; while in practice they are not known. For this, we use a two-run procedure for estimating the boundaries.

Starting from a given knot sequence $\mathcal{K} = \{p_0, p_1, \dots, p_N\}$ ($p_k \in [0, L]$, $p_0 = 0$ and $p_N = L$), in the first run, a boundary-less solution is found for each block according to (12). Due to the local behavior of the smoothing matrix shown in Fig. 2, the resulting estimates are smooth in their intermediate points; but not necessarily at the segment boundaries.

In the second run, a secondary knot sequence $\tilde{\mathcal{K}} = \{\tilde{p}_0, \tilde{p}_1, \dots, \tilde{p}_N\}$ is defined:

$$\begin{aligned} \tilde{p}_0 &= p_0 \\ \tilde{p}_i &= (p_i + p_{i-1})/2, \quad 0 < i < N \\ \tilde{p}_N &= L \end{aligned} \quad (20)$$

which is a 50% shifted and overlapped version of \mathcal{K} (except for the first and last points, which are fixed at 0 and L , respectively), and the signal is re-estimated from (11), using the boundary vectors α_k and β_k , from the first run.

2.6. The block-wise algorithm

Finally, by putting together the details in this section, the overall block-wise algorithm is summarized in Algorithm 1.

Algorithm 1 Block-wise smoothing algorithm

Require: Noisy signal $x[m]$ ($0 \leq m < L$)

Require: Knot sequence $\mathcal{K} = \{p_0, p_1, \dots, p_N\}$

Require: Upper bound of segment steepness ϵ_k^2 (or upper bound of segment noise variance δ_k^2)

- 1: Segmentize $x[m]$ into N segments \mathbf{x}_k using \mathcal{K}
 - 2: **for** $k = 0, \dots, N - 1$ **do** ▷ First Run
 - 3: Calculate λ_k^* for \mathbf{x}_k , using ϵ_k^2 in Algorithm 2
 - 4: Estimate $\hat{\theta}_k$ from (12), without boundary overlaps
 - 5: **end for**
 - 6: Calculate $\tilde{\mathcal{K}} = \{\tilde{p}_0, \tilde{p}_1, \dots, \tilde{p}_N\}$ from (20)
 - 7: Re-segmentize $x[m]$ into N segments $\tilde{\mathbf{x}}_k$ using $\tilde{\mathcal{K}}$
 - 8: **for** $k = 0, \dots, N - 1$ **do** ▷ Second Run
 - 9: Calculate λ_k^* for $\tilde{\mathbf{x}}_k$, using Algorithm 2
 - 10: Set the boundary vectors α_k and β_k from the output of the first run
 - 11: Re-estimate $\hat{\theta}_k$ from (11), using the boundary vectors
 - 12: **end for**
-

It should be noted that this algorithm may be simply implemented in real-time with one segment of lag. In this case, the first and second runs of the algorithm can be interlaced over successive blocks.

3. Local Lipschitz regularity

There is a close relationship between the proposed block-wise scheme and the concept of *global* versus *local* regularity. In fact, under the Fourier decaying assumption in (3), $\theta(t)$ is *uniformly Lipschitz- ν* , over \mathbb{R} [11, Ch. 6], which means that there exists a polynomial

$\phi_u(t)$ of degree $d = \lfloor \nu \rfloor$ and a constant $\ell > 0$ (independent of ν), for which

$$\forall t \in \mathbb{R} : |\theta(t) - \phi_u(t)| \leq \ell |t - u|^\nu \quad (21)$$

The parameter $\nu \geq 0$ is the *Lipschitz* (or *Hölder*) *exponent* and the smallest ℓ , which satisfies (21) is known as the *Lipschitz constant*. Equation (3) implies that $\theta(t)$ is d -times *continuously differentiable* (globally) and the d th order derivative $\theta(t)^{(d)}$ is uniformly Lipschitz- μ , where $0 \leq \mu = (\nu - d) < 1$. At each u the polynomial $\phi_u(t)$ is uniquely defined. Specifically, since $\theta(t)$ is d -times continuously differentiable in a neighborhood of u , then $\phi_u(t)$ is the Taylor expansion of $\theta(t)$ around u [11, Ch. 6, P. 206]. This result holds for all $t \in \mathbb{R}$, including uniformly sampled values $\theta[m] = \theta(mT_s)$ ($m \in \mathbb{Z}$).

On the other hand, by selecting a “sufficiently small” sampling time $T_s > 0$, the sequence obtained from applying the approximate difference operator \mathbf{d}_d to the sample sequence $\theta[m]$ has the same order as $\theta(t)^{(d)}$ and is bounded by the same upper bound in (21).

While ν and ℓ are global constants, if we confine the study to a single segment S_k , the upper bound in (21) can become tighter and segment dependent. In other words, one can have local regularity Lipschitz exponents ν_k and Lipschitz constants ℓ_k , which vary from one segment to another. The block-wise scheme proposed in the previous section, implicitly seeks the upper bound ϵ_k^2 through the segment dependent regularity factor λ_k . This is the same idea behind the zooming property of wavelet transforms [11, Ch. 6].

4. Frequency domain analysis: a block-wise Tikhonov-Wiener smoother

The block-wise matrix-form scheme developed in the previous section can also be stated in terms of a zero-phase, forward-backward linear time-invariant (LTI) smoother. Although, the transfer function of classical Tikhonov regularization has been already derived in previous studies [32, 33], our formulation additionally provides a frequency domain interpretation for the classical Tikhonov regularization in terms of a Wiener smoother. To show this, the sample-wise form of the data model in (5) is used. In each data segment k , we have

$$x_k[m] = \theta_k[m] + \eta_k[m], \quad (0 \leq m < n_k) \quad (22)$$

where $x_k[m]$ is the noisy signal, $\theta_k[m]$ is the desired signal and $\eta_k[m]$ is the zero-mean observation noise, with

$E\{\eta_k[m]^2\} = \delta_k^2$. Using the difference operator defined in Section 2.1, the d th order approximate derivative of $\theta_k[m]$ is found as follows:

$$\begin{aligned} w_k[m] &\triangleq \mathbf{d}_d[m] * \theta_k[m] \\ &= \mathbf{d}_d[0]\theta_k[m] + \dots + \mathbf{d}_d[M-1]\theta_k[m-M+1] \end{aligned} \quad (23)$$

Therefore, (9) can be rewritten as

$$\hat{\theta}_k[n] = \arg \min_{\theta_k[n]} \sum_{m=0}^{n_k-1} (x_k[m] - \theta_k[m])^2 + \lambda_k \sum_{m=M}^{n_k+M-1} w_k[m]^2 \quad (24)$$

where $\lambda_k \geq 0$ is the regularization factor¹. By setting the derivative of (24) with respect to $\theta_k[n]$ equal to zero and after some simplifications, we find:

$$(\delta[n] + \lambda_k \mathbf{d}_d[n] * \mathbf{d}_d[-n]) * \hat{\theta}_k[n] = x_k[n] \quad (25)$$

where $\delta[\cdot]$ is the Kronecker delta. If the first term in (25) is invertible, we arrive at

$$\hat{\theta}_k[n] = (\delta[n] + \lambda_k \mathbf{d}_d[n] * \mathbf{d}_d[-n])^{-1} * x_k[n] \quad (26)$$

which simply represents LTI non-causal smoothing of the noisy signal using the following transfer function:

$$\psi_k(z) \triangleq \frac{1}{1 + \lambda_k \mathbf{d}_d(z) \mathbf{d}_d(z^{-1})} \quad (27)$$

Equation (27) can be regarded as the transfer function of Tikhonov regularization (smoothing).

It is insightful to rewrite (27) in the following form

$$\psi_k(z) \triangleq \frac{\frac{\epsilon_k^2}{\mathbf{d}_d(z) \mathbf{d}_d(z^{-1})}}{\frac{\epsilon_k^2}{\mathbf{d}_d(z) \mathbf{d}_d(z^{-1})} + \delta_k^2} = \frac{S_{\theta}(z)}{S_{\theta}(z) + S_{\eta}(z)} \quad (28)$$

which is exactly in the form of an *optimal Wiener smoother* for a random process obtained from an autoregressive all-pole model with a spectra denoted by $S_{\theta}(z)$, driven by input white (innovation) noise with variance ϵ_k^2 , in presence of white observation noise with variance δ_k^2 and a white spectra denoted by $S_{\eta}(z)$. In this case, the parameter $\lambda_k = \delta_k^2 / \epsilon_k^2$ is the equivalent regularization factor defined throughout this work.

Fig. 3, illustrates the analogy between Tikhonov regularization and optimal Wiener smoothing using the

¹The lower and upper bounds of the second summation in (24) are due to the fact that in extending (9) to the sample-wise form, the first and last $M-1$ samples of $w_k[m]$ correspond with the boundary vectors and only the middle n_k correspond to $\theta_k[n]$.

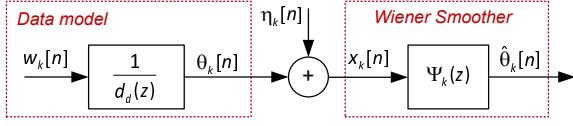


Figure 3: An equivalent Wiener smoother scheme for Tikhonov regularization

sample-wise formulation. This shows that the method proposed in the previous section can be envisaged as a piecewise-stationary Wiener smoother with smooth inter-block transitions. The frequency response of each block is controlled by the regularization factor λ_k .

Using a real-valued impulse response for \mathbf{d}_d , $\psi_k(z)$ is all-pole, lowpass and non-causal with conjugate symmetric pole pairs. Additionally, due to its invariance with respect to z and z^{-1} , its poles are reciprocal with respect to the unit circle. The only exception is the possible poles lying on the unit circle (which is studied below). In other cases, $\psi_k(z)$ can always be factorized as the cascade of a minimum-phase system $\psi_k^{min}(z)$ (using the poles inside the unit circle) applied as a causal filter, and a maximum-phase system $\psi_k^{max}(z)$ (using the poles outside the unit circle) applied as an anti-causal (reverse) filter. Therefore, a forward-backward smoothing scheme can be used to implement the smoother. Due to the reciprocity of the poles of $\psi_k(z)$ inside and outside the unit circle, only ψ_k^{min} is required for the filter implementation. The procedure is as follows: first find $v_k[n] = \psi_k^{min}[n] * x_k[n]$. Next, $v_k[n]$ is time reversed to $\tilde{v}_k[n] = v_k[-n]$, and given to the minimum-phase filter $y_k[n] = \psi_k^{min}[n] * \tilde{v}_k[n]$. The final result is $\hat{\theta}_k[n] = y_k[-n]$. Unity DC gain is achieved by normalizing the results of the forward and backward filters by $\sum_n \psi_k^{min}[n]$.

The whole procedure is rather similar to the *filtfilt* function in Matlab and Octave, using $\psi_k^{min}(z)$ as the transfer function. It should just be noted that according to the two-run scheme described in Section 2.6, during the forward and backward filtering phases, the output initial conditions of the convolutions in each segment k are taken from the outputs of the neighboring segments S_{k-1} and S_{k+1} , obtained from the first run.

4.1. Stability

$\psi_k(z)$ is only stable for odd-length \mathbf{d}_d , as it has inevitable unit-circle poles for even-length difference operators. The reasoning is as follows: supposing that \mathbf{d}_d is a real-valued FIR impulse response of length M , $\psi_k(z)$ is an IIR filter with a denominator of length $2M - 1$ and $2M - 2$ conjugate symmetric roots. Noting that these roots should also be reciprocal with respect to the unit

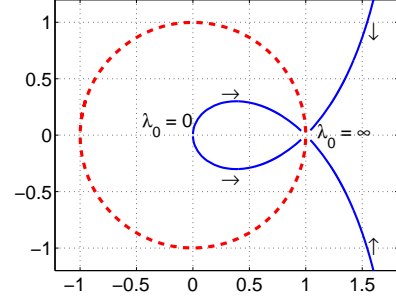


Figure 4: Root locus of the forward-backward smoother defined in (27), for $d = 2$. Each line segment is the track of one of the four conjugate symmetric and reciprocal poles, as λ_k varies from 0 to ∞ (in directions of arrows).

circle, the reciprocity of the poles is only attained when the number of poles is a multiple of four, i.e., M is odd. For even values of M , (at least) two of the conjugate symmetric poles are forced to be on the unit circle, and the remaining poles are reciprocal inside and outside the unit circle. Therefore, the factorization of $\psi_k(z)$ in terms of the cascade of a minimum-phase and a maximum-phase filter (without any unit-circle roots) is only achievable for odd values of M , resulting in a stable forward-backward smoother. The system is unstable for even values of M .

As a special case, using the frequency form of the difference operators defined in (1), the frequency response of the forward-backward smoother is found as follows

$$\psi_k(e^{j\omega}) = \frac{1}{1 + \lambda_k(2 - 2 \cos \omega)^d} = \frac{1}{1 + \lambda_k(2 \sin \frac{\omega}{2})^{2d}} \quad (29)$$

which resembles the amplitude response of a discretized Butterworth filter of order d , in low frequencies ($|\omega| = |2\pi f/f_s| \ll 1$). The filter response is monotonically decreasing over $\omega \in [0, \pi]$ and is maximally flat in its pass-band, in the sense that its first $2d - 1$ derivatives are zero at $\omega = 0$ without any pass-band or stop-band ripples. Therefore, the formulation is also attractive from the filter design perspective. The root loci and the corresponding filter frequency response of $\psi_k(z)$ are shown in Figs. 4 and 5 for different values of d and λ_k . Accordingly, the block-wise scheme presented in the previous section can be interpreted as a smoother, which adapts its bandwidth according to the noise variance and the signal steepness controlled by λ_k^* .

4.2. Regularization factor selection

The properties of the LTI model under the variation of the parameter λ_k is similar to the block-wise scheme:

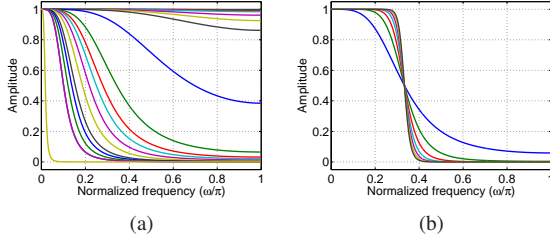


Figure 5: The frequency response of the forward backward smoother defined in (27), (a) for $d = 2$, as λ_k varies from 0 to ∞ (from top to bottom), (b) for $\lambda_k = 1$ and $d = 2, 4, 6, 8, 10, 12, 14$ (from smoothest to steepest). The maximally flat pass-band property of the filter is seen here.

the solution becomes smoother as λ_k increases from zero to infinity. From (29), it is noticed how λ_k is inversely related to the equivalent lowpass filter bandwidth. If the upper-bound of the observation noise variance δ_k^2 is known (or assumed), λ_k has an optimal value, which can be numerically found by solving:

$$\frac{1}{2\pi} \int_{2\pi} |X(e^{j\omega})(1 - \psi_k(e^{j\omega}))|^2 = L\delta_k^2 \quad (30)$$

which is equivalent to solving $e'(\gamma_k^*) = L\delta_k^2$ for γ_k^* , in the alternative formulation stated in Appendix A and then setting $\lambda_k^* = 1/\gamma_k^*$.

5. Generalization to multiple order smoothness

An interesting generalization of the proposed method is to penalize the norms of multiple derivatives of the estimated signal at the same time, but with different regularization factors. As compared to the formulation proposed in Section 2, this approach provides additional control over the estimated waveforms, as it draws the results towards a smaller subset of the null space of \mathbf{D}_d . This feature can be useful for extracting smooth signals of order d , with bounded lower order derivatives. In this case, the cost function in (10) can be rewritten as follows:

$$J_k^M = \|\mathbf{x}_k - \boldsymbol{\theta}_k\|^2 + \sum_{i=0}^d \lambda_k^{(i)} \|\mathbf{D}_i \boldsymbol{\theta}_k + \mathbf{b}_k^{(i)}\|^2 \quad (31)$$

where $\mathbf{b}_k^{(i)} \triangleq \mathbf{E}_i \boldsymbol{\alpha}_k + \mathbf{F}_i \boldsymbol{\beta}_k$ is the multiple order extension of the *boundary vector* and $\lambda_k^{(i)}$ is the i th derivative regularization factor. It is straightforward to show that in this case, the block-wise solution in (11) is generalized as follows

$$\hat{\boldsymbol{\theta}}_k = (\mathbf{I} + \sum_{i=0}^d \lambda_k^{(i)} \mathbf{D}_i^T \mathbf{D}_i)^{-1} [\mathbf{x}_k - \sum_{i=0}^d \lambda_k^{(i)} \mathbf{D}_i^T \mathbf{b}_k^{(i)}] \quad (32)$$

Similarly, the sample-wise solution in (25), its corresponding transfer function in (27), and the frequency response in (29) become

$$(\delta[n] + \sum_{i=0}^d \lambda_k^{(i)} \mathbf{d}_i[n] * \mathbf{d}_i[-n]) * \hat{\boldsymbol{\theta}}_k[n] = x_k[n] \quad (33)$$

$$\psi_k(z) = \frac{1}{1 + \sum_{i=0}^d \lambda_k^{(i)} \mathbf{d}_i(z) \mathbf{d}_i(z^{-1})} \quad (34)$$

$$\psi_k(e^{j\omega}) = \frac{1}{1 + \sum_{i=0}^d \lambda_k^{(i)} (2 \sin \frac{\omega}{2})^{2i}} \quad (35)$$

The discussion regarding the stability is as before (cf. Section 4.1). As compared to the single order model, it is apparently more difficult to find multiple optimal regularization factors $\lambda_k^{(i)}$. However, they can be simply fixed by intuition or be found by Monte Carlo simulation over desired datasets.

6. Case study: electrocardiogram denoising

6.1. Motivation

The electrocardiogram (ECG) is a good example of non-stationary signals. While the normal ECG has a pseudo-periodic behavior (due to the cardiac rhythm), abnormal and ectopic ECG beats are highly irregular.

In modern digital ECG acquisition systems, in order to enable better visual interpretations and to cover the vast inter-subject variations, the ECG is commonly oversampled far above the minimum required rate. For instance, although the major diagnostic information of the ECG is below 100 Hz (in adults), current ECG acquisition standards recommend a sampling frequency of 250-500 Hz for low and moderate quality ECG, and 1-10 kHz for high quality recordings [12]. On the other hand, apart from the baseline wander, other ECG noises and artifacts can span across the entire Nyquist band. Therefore, in moderate and high-quality digitized ECG, the desired waveform is rather smooth (the adjacent samples are highly correlated), while the noise samples are rather uncorrelated. Based on these facts, in this section the proposed denoising scheme is used for removing wide-band noise from ECG recordings.

It should be noted that in previous research, filtering techniques such as extended Kalman filters based on the pseudo-periodicity of ECG signals have resulted in very good results for normal ECG [18]. However, the performance of these methods significantly degrade for abnormal ECG, which makes them inapplicable for denoising highly irregular and ectopic beats. Therefore, LTI filters with fixed impulse responses are still considered as the gold standard of ECG acquisition system front-ends.

6.2. Database

Three standard databases from the PhysioBank are used for evaluation:

1. *The MIT-BIH normal sinus rhythm database*, which contains 18 long-term two-channel normal ECG recordings, sampled at 128 Hz [34].
2. *The MIT-BIH arrhythmia database*, which contains 48 half-hour two-channel ambulatory ECG recordings with annotated anomalies, sampled at 360 Hz [35].
3. *The PTB diagnostic database (PTB)*, which contains 549 fifteen channel ECG recordings, sampled at 1 kHz [36]. Due to the huge size of this database, only the first ninety subjects are used in this study.

These data are selected to show the performance of the proposed methods in low, moderate and high sampling frequencies for normal and abnormal ECG.

6.3. Algorithm parameters

In the following case studies, the smoothness order is set to $d = 2$, $\mathbf{d}_2 = (1, -2, 1)$, and the smoothing operator $\tilde{\mathbf{D}}_2$ is as defined in (2). The boundary vectors α_k and β_k have a length $q = M - 1$ ($=2$ in this case). The knot spacing has been fixed to 100 ms, which is a fraction of a typical ECG beat length and the ECG is roughly smooth up to its second order difference. The accuracy threshold of Algorithm 2 is set to $C = 10^{-8}$, and the forgetting factor in (19) is set to $\kappa = 1$, to show the block-wise variations of the regularization factor λ_k^* without any smoothing.

In the later presented results, several modes of the proposed algorithm are studied:

1. Single block version (Section 4), using a global optimal λ_0 found by a naive grid search.
2. A known input noise variance (δ_k^2)
3. Underestimated input noise variance ($0.8\delta_k^2$)
4. Overestimated input noise variance ($1.2\delta_k^2$).

The motivation for the two latter cases is to check the performance when the pre-assumption on (or estimation of) the input noise variance is inexact.

6.4. Benchmark algorithms

To date, numerous filters have been used for ECG denoising. In order to keep the comparisons fair and generic, we select benchmarks, which are not data dependent, do not require prior assumptions (such as pseudo-periodicity), the detection of the R-peaks or other ECG fiducial points, and have been reasonably successful for ECG denoising in previous work.

6.4.1. Linear time-invariant (LTI) lowpass filter

Although simple, linear-phase LTI filters are currently the standard choice for ECG system front-ends, as they impose little assumptions on the signals and preserve their phase contents. Therefore, the first benchmark is a naive linear-phase lowpass Butterworth filter with variable cutoff frequency. The -3dB cutoff frequency was swept linearly over 100 values ranging from 5 Hz to the Nyquist frequency (half the sampling frequency), in three Butterworth filter orders 2, 4, and 6. The designed filters were applied to the data using the *filtfilt* function in Matlab (to have zero-phase lag and provide similar comparison conditions with the proposed method). According to the simulation procedure described in the sequel, it was found that the filter order 2 slightly outperformed higher orders.

6.4.2. Wavelet denoiser (WD)

The well-known wavelet shrinkage scheme developed by Donoho et al. was originally developed using signal smoothness priors [2]. The optimal performance of this method using appropriate mother wavelets and different wavelet shrinkage schemes has been extensively studied in the past two decades. WD is offline and block-wise and highly benefits from the signal properties in various decomposition levels, which are much more sophisticated (and perhaps more effective) than the simple smoothness priors used in this work. Nevertheless, we use it for performance comparison with the proposed method. Due to the diversity of the literature on WD performance on ECG signals, an exhaustive Monte Carlo simulation was performed on the three databases using various combinations of WD parameters, listed in Table 1. Overall, 6960 different parameter combinations have been studied, which is broader than our previous WD combination set reported in [18].

Each method was tested in presence of input white noise in variable SNR ranging from -15 dB to +30 dB. The average and standard deviation of SNR improvement obtained after filtering was used as a measure of performance.

Among the different tested parameter combinations, the SURE shrinkage rule, with a single-level rescaling and soft thresholding strategy resulted in the highest performance over all three databases. The multi-level rescaling performance was rather close (but lower) than the single-level; but the no-rescaling method always had poor results, except in very high input SNR (24dB and above). Among the tested mother wavelets, *Coiflets* 2, 3, 4, 5, and *Symlets* 5, 6, 7, 8, *Daubechies* 2 commonly gave superior results, without a significant difference (considering the standard deviation of the results).

Table 1: Wavelet denoising parameter combinations tested over the databases

Parameter	Values
Mother wavelet	Daubechies 1 (Haar), 2, 3, \dots , 10, 12, 16 Coiflets 1, \dots , 5 Symlets 2, \dots , 8 Biorthogonal 1.5, 2.6, 2.8, 5.5, 6.8
Shrinkage rule	SURE, heuristic SURE, universal, minimax
Thresholding strategy	hard, soft
Rescaling approach	no-scaling, single-level, multi-level
Decomposition level	1, \dots , 10

The optimal decomposition level for WD is known to be sampling frequency dependent. Throughout the Monte Carlo simulations, the optimal level was found to be 7, 5, and 9, for the normal, arrhythmia, and PTB databases, respectively. In the results section, the WD results are reported in two cases:

1. The best WD decomposition level adapted per each input SNR (named WAVELET₀)
2. A fixed decomposition level optimized over all input SNR (named WAVELET).

The first case is only used as a corner case, as the exact input SNR is not known in practice and the optimal decomposition level is unknown.

6.4.3. Order-optimized Savitzky-Golay filter (SGF)

The Savitzky-Golay filter using the SURE shrinkage rule for optimal order selection has been recently used for ECG denoising [15]. Three variants of the SGF have been reported in [15]:

1. SGF of order 3 with variable length impulse response and using SURE for finding the optimal impulse response (SGBW).
2. SGF of order 3 with variable length impulse response and using regularized SURE for finding the optimal impulse response (SGBWR).
3. SGF of variable order from 1 to 5 using SURE for finding the optimal impulse response length (SGO).

All three schemes were applied to the studied databases², with the order parameter swept from 5 to 75 in steps of 2 (as required by the algorithm in [15]). It was noticed that the optimal filter order was inversely related to the input SNR and the performance highly relied on this parameter. However, as with all the benchmark methods, the optimal order was selected as the one

²The source codes for this benchmark were kindly provided by the authors of [15].

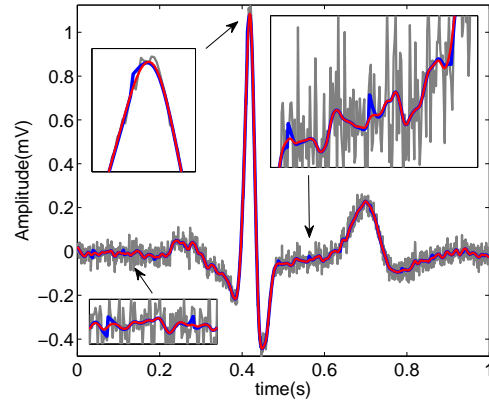


Figure 6: A sample noisy ECG segment (in gray), the smoothed signals after the first stage of Algorithm 1 without any boundary constraints (in blue), and after the second stage using boundary constraints (in red).

that maximized the average SNR improvement over the entire range of input SNR.

6.5. Stationary white noise removal

6.5.1. Visual inspection

In order to show the impact of considering the boundary conditions between segments, a short ECG segment adopted from the PTB diagnostic database and the results obtained from the first and second runs of Algorithm 1 are shown in Fig. 6. Accordingly, the transitions between blocks have become totally smooth after the second stage of the smoother (in Algorithm 1). Both runs give rather similar estimates in their intermediate points, which shows that the segment boundary constraints have local impacts and do not affect the interior samples.

In Fig. 7, a sample ECG segment adopted from the MIT-BIH arrhythmia database is shown, which contains ectopic beats. The signal was artificially contaminated by white Gaussian noise in three input SNR: 0dB, 6dB

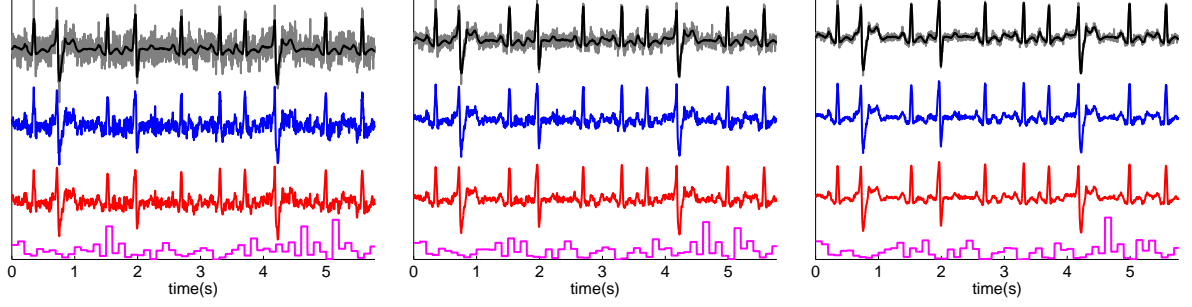


Figure 7: A sample ECG segment containing an ectopic beat in three input SNR: 0dB (left), 6dB (middle) and 12dB (right). From top to bottom, the signals correspond to: the original signal overlaid with the noisy signal, the sample-wise smoother scheme with a fixed λ , and the block-wise matrix form. The last signal is the optimal penalty factor $\gamma_k^* = 1/\lambda_k^*$ corresponding to each 100 ms data segment.

and 12dB. The denoising results using the block-wise and sample-wise schemes are shown in Fig. 7. In order to calculate the optimal λ_k^* , the alternative formulation of the algorithm described in Appendix A was used and the noise variance δ_k^2 was given to the algorithm as a prior. For the sample-wise smoother (Section 4), a single regularization factor λ was fixed to be the maximum of λ_k^* obtained from the block-wise scheme. The optimal penalty factor $\gamma_k^* = 1/\lambda_k^*$ is shown in the last row of Fig. 7. Accordingly, although the noise variance was uniform throughout the whole signal, γ_k^* has significant block-wise variations, due to the local ECG amplitude changes over each beat.

The L-curves corresponding to the data in Fig. 7 are shown in Fig. 8. Accordingly, for all the plotted L-curves, due to the fixed signal SNR, the optimal λ_k^* (shown by circles) are located at a fixed error level $\|\mathbf{x}_k - \boldsymbol{\theta}_k\|$, which by visual inspection are rather close to the corner of the L-curve.

Therefore, for real scenarios in which δ_k^2 (or ϵ_k^2) are unknown, according to our primary model in (8) an upper bound of these parameters can be used instead. Choosing worst case upper bounds for δ_k^2 (or ϵ_k^2) leads in sub-optimal results. Alternatively, a sweep of the regularization factor for obtaining the corner of the L-curve gives an empirical estimate for λ_k^* .

6.5.2. Monte Carlo simulation

In order to compare the overall performance of the proposed method, it was applied to the three standard databases described in Section 6.2. The total number of 10 s data segments were 1116, 2016, and 1150, for the Normal, Arrhythmia, and PTB databases, respectively. White noise was synthetically produced and added to each segment in different SNR ranging from -15 dB to 30 dB in 3 dB steps, and the performance

was studied for three difference operator orders $d = 2, 4$, and 6. High SNR (above 24 dB) cases are considered very good quality ECG; intermediate SNR (ranging between 18 dB to -6 dB) have been used as a typical range for *noise stress testing* studies [37], and very low input SNR (-6 dB and below) are only of interest for extremely noisy cases demanding rather coarse ECG analysis, such as R-peak detection from poor quality data (for instance for noninvasive fetal R-wave detection from highly noisy data [38]).

For this study, the proposed methods were tested using the parameters described in Section 6.3. For the block-wise scheme the noise variance was given as prior and the optimal regularization factor was calculated according to the γ -based formulation, detailed in Appendix A. For the sample-wise forward-backward smoothing scheme, an optimal global regularization factor γ_0 was found for each data segment by using a bisection search between 10^{-12} to 10^6 over 100 iterations. The optimal regularization factor γ_0^* , was selected as the one which resulted in the maximum SNR improvement ($\Delta\text{SNR}(\text{dB}) = \text{SNR}_{\text{out}}(\text{dB}) - \text{SNR}_{\text{in}}(\text{dB})$) at the filter's output. The average and standard deviation of SNR improvements are shown in Fig. 9, for the three studied databases, in comparison with the benchmark methods. Accordingly, in low input SNR, WAVELET₀ has the best average SNR improvement, which was anticipated due to the scenario used for implementing WAVELET₀, as noted in Section 6.4.2. In high input SNR, the proposed method using fixed λ_0 and underestimated input noise variance ($0.8\Delta_k^2$) outperform other methods.

For the WD, as expected, optimizing the decomposition level per input SNR (WAVELET₀) gives better results as compared with a global decomposition level (WAVELET). The simple lowpass filter and the SGF highly degrade in high input SNR. Overall, the SGF

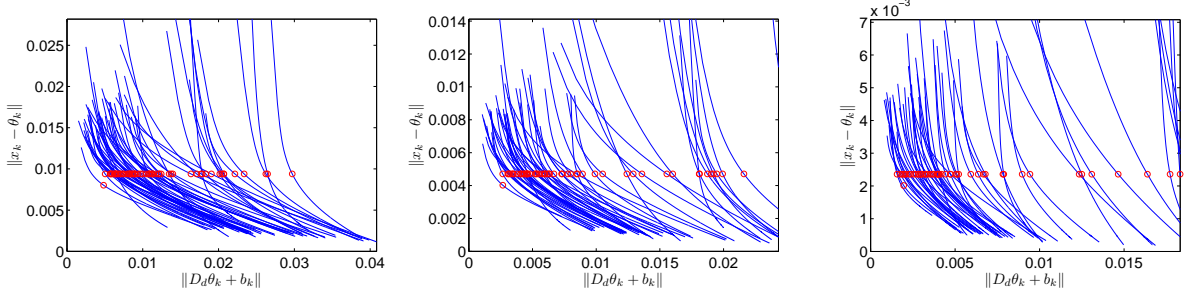


Figure 8: The L-curves corresponding to the data in Fig. 7 as λ is swept from $0.04\lambda_k^*$ to $25\lambda_k^*$, where λ_k^* is the optimal regularization factor located at the circles.

does not show any superior results (neither in terms of average nor standard deviation), except for the SGBWR applied to the PTB database in input SNR ranging from 0dB to 10dB.

For the proposed method, the difference operator order $d = 2$, outperformed higher order factors. Moreover, the SNR improvement obtained in higher sampling frequencies (specifically for the PTB database) was better than the results obtained in lower sampling frequencies. This result was already anticipated, since in higher sampling rates, the noise power, which spans over the total Nyquist band has less overlap with the desired signal. In all cases, the optimal regularization factor γ_0^* is inversely related to the SNR. Moreover, the results show that it is better to overestimate the input noise variance in low input SNR and to underestimate it for high input SNR. Using the exact input noise variance has intermediate results.

Considering the standard deviation (SD) results in Fig. 9, all variants of the proposed method are more robust (have smaller SD) than the previous techniques, including the WD. In fact, while the SD of the proposed methods remain bounded with the variation of the input SNR, the SD of other methods tend to increase in higher input SNR.

6.6. Non-stationary noise removal

The experiment described in the previous section was repeated for non-stationary white noise modulated by sinusoidal envelopes, in three input SNR: -3dB, 0dB and 3dB (Fig. 10). This noise resembles non-stationary noise due to electrode displacements and muscle artifacts, which are commonly wider in bandwidth as compared with the contaminated ECG; but overlap in the time and frequency domains. In this case, due to the variations of the noise variance, using a fixed prior noise variance δ_k^2 is inappropriate as it leads to sub-optimal performance in each block. For this case, we used the

L-curve method to find the optimal regularization parameter λ_k^* . From the last row of Fig. 10, it is noticed how the variations of $\gamma_k^* = 1/\lambda_k^*$ is inversely related to the envelope of the non-stationary noise, which is in full agreement with the interpretation of λ_k^* , in Section 4 (Fig. 3), as the ratio of the observation noise variance divided by the innovation noise variance, i.e., γ_k^* is inversely proportional to the observation noise variance.

Alternatively, for such scenarios (with non-stationary variable noise variance), a block can be added to the proposed algorithm, which estimates the block-wise noise variance using complementary priors. This extension is not covered in this study.

7. Summary and future work

In this paper, a block-wise extension of the well-known Tikhonov regularization was proposed for denoising smooth signals contaminated by wide-band noise. The proposed method can be implemented in two forms: 1) as a block-wise fixed-lag smoother with smooth inter-block transitions applied in a matrix form, 2) as a fixed-interval smoother applied as a forward-backward zero-phase filter. The filter response is maximally flat and monotonically decreasing, without any ripples. The method was also extended to smoothness of multiple derivative orders and its relationship with Lipschitz regularity and Wiener smoothing was also studied.

ECG denoising was studied as a typical case study and compared with conventional bandpass, WD and SGF filters. All Matlab source codes of the proposed method are online available in the *open-source electrophysiological toolbox* (OSET) [39].

The proposed method is very general with potential applications in various fields of signal processing. Some of the promising extensions of the hereby proposed schemes include:

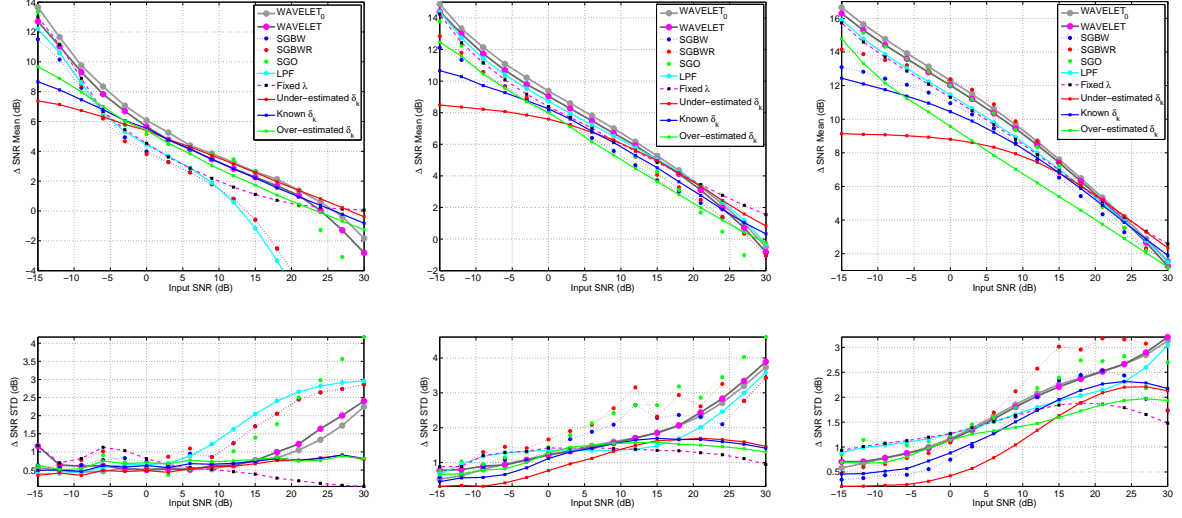


Figure 9: The results of Monte Carlo simulations over three standard databases. The average SNR improvement (first row) and SNR improvement standard deviation (second row). The left column corresponds with the Normal Sinus database, the middle column shows the Arrhythmia database and the right column is the result of the PTB database.

- The data model described in (5) was based on the assumption that the observed data \mathbf{x}_k were simply noisy observations of θ_k . For applications such as model identification and signal compression, θ_k can be further considered to have a structure, such as $\theta_k = \mathbf{H}\mathbf{s}_k$, where \mathbf{s}_k are some latent variables (hidden features) and \mathbf{H} is a mixture matrix or even some basis functions.
- The same framework can be extended to other priors, such as temporal or spectral priors. This extension is simply achieved by replacing the operator \mathbf{D}_d with other filter responses, e.g., the original Savitzky-Golay convolution coefficient set [40], or recent methods proposed for the optimal order selection of these filters [15]. The filter operator can also be time-varying, such as signal-dependent weighted averages.
- The ECG case study shows that WD using SURE shrinkage rule has very good results for offline applications, while the proposed framework is more robust (in terms of SD) and is applicable online. In future studies, the combination of the two methods may result in a piecewise online WD schemes with smooth inter-block transitions. The concept of Lipschitz regularity studied in Section 3 can serve as a common ground for relating wavelet zooming and wavelet denoising schemes with the proposed framework.
- The Wiener equivalent formulation stated in (28), suggests the possible extension of the proposed method for colored spectral noise (non-constant $S_\eta(z)$), having spectral overlap with the desired signals. In case of non-stationary colored noise, spectral factorization methods can be used to derive dynamical models for the proposed filters, which can also be used to construct a Kalman filter (smoother) implementation of the proposed schemes.
- The extension to multichannel spatially correlated data is another promising direction, which can provide a joint spatio-temporal smoothing technique, related to semi-blind source separation techniques using smoothness priors.
- In the current formulation, no assumptions were made on the stochastic distribution of the signal or noise; since the constrained least squares scheme does not rely on the signal distributions. However, if any such assumptions exist in a specific scenario (such as Gaussian assumption on the signal or noise), the formulation can be extended to a Maximum Likelihood or Maximum a Posteriori framework, which provide additional statistical interpretations for the minimized costs.

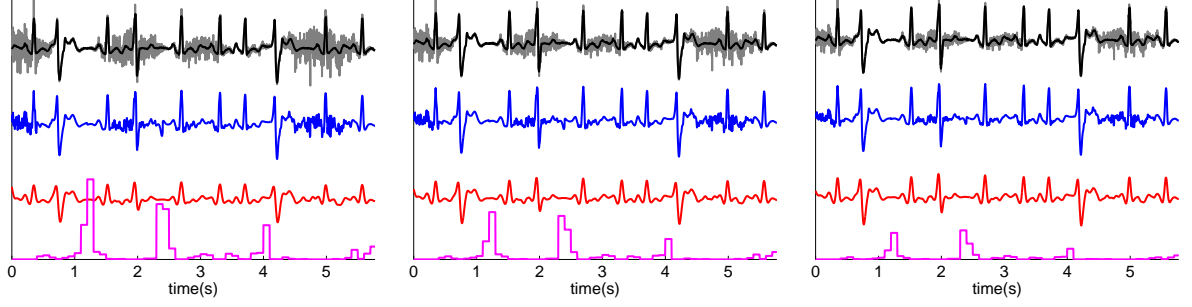


Figure 10: A sample ECG segment containing an ectopic beat contaminated by non-stationary noise resembling electrode displacements and muscle artifacts, in three input SNR: -3dB (left), 0dB (middle) and 3dB (right). From top to bottom, the signals correspond to: the original signal overlaid with the noisy signal, the fixed-lambda sample-wise smoother scheme, and the block-wise scheme with λ_k^* found by the L-curve method. The last row is the optimal penalty factor $\gamma_k^* = 1/\lambda_k^*$ corresponding to each 100 ms data segment.

Appendix A. An alternative formulation: maximal smoothness with bounded noise variance

In practice, the knowledge of the segment measurement noise variance δ_k^2 can be more accessible than the signal's steepness bound ϵ_k^2 . In this case, an alternative formulation for (8) is to seek the smoothest estimate of θ_k under an upper bound for the mean square error variance:

$$\hat{\theta}_k' = \arg \min_{\theta_k} \|\tilde{\mathbf{D}}_d \bar{\theta}_k\|^2, \text{ s.t. } \|\mathbf{x}_k - \theta_k\|^2 \leq n_k \delta_k^2 \quad (\text{A.1})$$

In this case, the Lagrangian and the matrix solutions are respectively as follows

$$\hat{\theta}_k' = \arg \min_{\theta_k} \left\{ \|\tilde{\mathbf{D}}_d \bar{\theta}_k\|^2 + \gamma_k \|\mathbf{x}_k - \theta_k\|^2 \right\} \quad (\text{A.2})$$

$$\hat{\theta}_k' = (\gamma_k \mathbf{I} + \mathbf{D}_d^T \mathbf{D}_d)^{-1} [\gamma_k \mathbf{x}_k - \mathbf{D}_d^T \mathbf{b}_k] \quad (\text{A.3})$$

Equation (A.3) is identical to (11), when the error penalty $\gamma_k = 1/\lambda_k$. Using the SVD matrix definitions from Section (2.3), and defining $\tilde{\theta}_k' \triangleq \mathbf{V}^T \hat{\theta}_k'$, the sample-wise update, the mean square error and the smoothness bound are found as follows

$$\tilde{\theta}_{ik}' = \frac{\gamma_k \tilde{x}_{ik} - \sigma_i \tilde{b}_{ik}}{\gamma_k + \sigma_i^2}, \quad (i = 1, \dots, n_k) \quad (\text{A.4})$$

$$e'(\gamma_k) = \|\mathbf{x}_k - \hat{\theta}_k'\|^2 = \sum_{i=1}^{n_k} \left(\frac{\sigma_i^2 \tilde{x}_{ik} + \sigma_i \tilde{b}_{ik}}{\gamma_k + \sigma_i^2} \right)^2 \quad (\text{A.5})$$

$$c'(\gamma_k) = \|\mathbf{D}_d \hat{\theta}_k' + \mathbf{b}_k\|^2 = \sum_{i=1}^{n_k} \left(\frac{\sigma_i \tilde{x}_{ik} + \tilde{b}_{ik}}{1 + \sigma_i^2 / \gamma_k} \right)^2 \quad (\text{A.6})$$

where it is again noticed that the optimal solution is on the boundary of the constraint region, if $n_k \delta_k^2 \leq e'(0)$.

In this case, the optimal error penalty γ_k^* is found by numerically solving the monotonically decreasing function

$$e'(\gamma_k^*) = \sum_{i=1}^{n_k} \left(\frac{\sigma_i^2 \tilde{x}_{ik} + \sigma_i \tilde{b}_{ik}}{\gamma_k^* + \sigma_i^2} \right)^2 = n_k \delta_k^2 \quad (\text{A.7})$$

using the modified bisection scheme described in Algorithm 2. As before, there is a unique solution on the constraint boundary for $n_k \delta_k^2 \leq e'(0)$. For $n_k \delta_k^2 > e'(0)$, we have $\gamma_k^* = 0$. The discussion on possible temporal smoothing of γ_k^* is similar to Section 2.3.

Appendix B. Modified bisection for optimal regularization factor selection

The modified bisection algorithm used for calculating the optimal regularization factor described in Section 2.3 is detailed in Algorithm 2.

Acknowledgment

The author would like to thank Dr. Sunder Ram Krishnan and Dr. Chandra Sekhar Seelamantula for providing the Matlab source codes of the order-optimized Savitzky-Golay filter [15], used in our benchmarks.

References

- [1] J. V. Candy, Model-Based Signal Processing, Wiley-IEEE Press, 2005.
- [2] D. L. Donoho, I. M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, Journal of the American Statistical Association 90 (432) (1995) 1200–1224.
- [3] G. H. Golub, P. C. Hansen, D. P. O'Leary, Tikhonov regularization and total least squares, SIAM Journal on Matrix Analysis and Applications 21 (1) (1999) 185–194.

Algorithm 2 Calculation of λ_k^* using Bisection

Require: Bisection stopping accuracy threshold C **Require:** Maximum number of bisection iterations

```
1:  $I_{max}$   
2: if  $c(0) < \epsilon_k^2$  then  
3:    $\lambda_k^* = 0$   
4: else  
5:    $\lambda_{min} = 0$ ;  
6:    $\lambda_{max} = 1$ ;  
7:   while  $c(\lambda_{max}) > \epsilon_k^2$  do  
8:      $\lambda_{min} = \lambda_{max}$ ;  
9:      $\lambda_{max} = 10\lambda_{max}$ ;  
10:  end while  
11:    $\triangleright$  Find  $\lambda_k^*$  using Bisection  
12:   $i = 0$ ;  
13:  while  $|\lambda_{max} - \lambda_{min}| > C$  &  $i < I_{max}$  do  
14:     $i = i + 1$ ;  
15:     $\lambda_{mid} = (\lambda_{max} + \lambda_{min})/2$ ;  
16:    if  $[c(\lambda_{min}) - \epsilon_k^2][c(\lambda_{mid}) - \epsilon_k^2] \leq 0$  then  
17:       $\lambda_{max} = \lambda_{mid}$ ;  
18:    else  
19:       $\lambda_{min} = \lambda_{mid}$ ;  
20:    end if  
21:  end while  
22:   $\lambda_k^* = (\lambda_{max} + \lambda_{min})/2$ ;  
23: end if
```

- [4] S. Boyd, L. Vandenberghe, Convex optimization, Cambridge university press, 2004.
- [5] Ç. Candan, H. Inan, A unified framework for derivation and implementation of Savitzky–Golay filters, Signal Processing 104 (2014) 203–211.
- [6] Y. Eldar, Generalized SURE for Exponential Families: Applications to Regularization, Signal Processing, IEEE Transactions on 57 (2) (2009) 471–481. doi:10.1109/TSP.2008.2008212.
- [7] P. Comon, C. Jutten, Handbook of Blind Source Separation: Independent Component Analysis and Applications, Independent Component Analysis and Applications Series, Elsevier Science, 2010.
- [8] A.-O. Boudraa, J.-C. Cexus, EMD-Based Signal Filtering, Instrumentation and Measurement, IEEE Transactions on 56 (6) (2007) 2196–2202. doi:10.1109/TIM.2007.907967.
- [9] M. P. Tarvainen, P. O. Ranta-Aho, P. A. Karjalainen, An advanced detrending method with application to HRV analysis, IEEE Transactions on Biomedical Engineering 49 (2) (2002) 172–175.
- [10] F. Zhang, S. Chen, H. Zhang, X. Zhang, G. Li, Bioelectric signal detrending using smoothness prior approach, Medical engineering & physics 36 (8) (2014) 1007–1013.
- [11] S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, 3rd Edition, Elsevier Science, 2009.
- [12] P. Kligfield, L. S. Gettes, J. J. Bailey, R. Childers, B. J. Deal, E. W. Hancock, G. van Herpen, J. A. Kors, P. Macfarlane, D. M. Mirvis, et al., Recommendations for the standardization and interpretation of the electrocardiogram: part I: the electrocardiogram and its technology, Journal of the American College of Cardiology 49 (10) (2007) 1109–1127.
- [13] M. A. Kabir, C. Shahnaz, Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains, Biomedical Signal Processing and Control 7 (5) (2012) 481–489.
- [14] B. N. Singh, A. K. Tiwari, Optimal selection of wavelet basis function applied to ECG signal denoising, Digital Signal Processing 16 (3) (2006) 275 – 287. doi:http://dx.doi.org/10.1016/j.dsp.2005.12.003. URL <http://www.sciencedirect.com/science/article/pii/S1051200405001703>
- [15] S. Krishnan, C. Seelamantula, On the selection of optimum savitzky-golay filters, Signal Processing, IEEE Transactions on 61 (2) (2013) 380–391. doi:10.1109/TSP.2012.2225055.
- [16] J. Sadasivan, S. Mukherjee, C. Seelamantula, An optimum shrinkage estimator based on minimum-probability-of-error criterion and application to signal denoising, in: Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on, 2014, pp. 4249–4253. doi:10.1109/ICASSP.2014.6854403.
- [17] R. Sameni, M. B. Shamsollahi, C. Jutten, Filtering Electrocardiogram Signals Using the Extended Kalman Filter, in: Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), Shanghai, China, 2005, pp. 5639–5642.
- [18] R. Sameni, M. B. Shamsollahi, C. Jutten, G. D. Clifford, A non-linear bayesian filtering framework for ECG denoising, Biomedical Engineering, IEEE Transactions on 54 (12) (2007) 2172–2185.
- [19] R. Sameni, M. B. Shamsollahi, C. Jutten, Model-based Bayesian filtering of cardiac contaminants from biomedical recordings, Physiological Measurement 29 (5) (2008) 595–613. doi:10.1088/0967-3334/29/5/006.
- [20] O. Sayadi, M. B. Shamsollahi, ECG denoising and compression using a modified extended Kalman filter structure, Biomedical Engineering, IEEE Transactions on 55 (9) (2008) 2240–2248.

- [21] E. Kheirati Roonizi, R. Sassi, A signal decomposition model-based bayesian framework for ecg components separation, *Signal Processing, IEEE Transactions on* 64 (3) (2016) 665–674. doi:10.1109/TSP.2015.2489598.
- [22] O. Sayadi, M. B. Shamsollahi, G. D. Clifford, Robust detection of premature ventricular contractions using a wave-based bayesian framework, *Biomedical Engineering, IEEE Transactions on* 57 (2) (2010) 353–362.
- [23] M. Akhbari, M. B. Shamsollahi, C. Jutten, A. A. Armoundas, O. Sayadi, Ecg denoising and fiducial point extraction using an extended kalman filtering framework with linear and nonlinear phase observations, *Physiological measurement* 37 (2) (2016) 203.
- [24] R. Sameni, Extraction of Fetal Cardiac Signals from an Array of Maternal Abdominal Recordings, Ph.D. thesis, Sharif University of Technology – Institut National Polytechnique de Grenoble (July 2008).
URL <http://www.sameni.info/Publications/Thesis/PhDThesis.pdf>
- [25] M. Niknazar, B. Rivet, C. Jutten, Fetal ECG extraction by extended state Kalman filtering based on single-channel recordings, *Biomedical Engineering, IEEE Transactions on* 60 (5) (2013) 1345–1352.
- [26] W. H. Press, S. A. Teukolsky, W. T. Vetterling, Numerical Recipes in C: The Art of Scientific Computing, 2nd Edition, Cambridge University Press, 1992.
- [27] W. Gander, Least squares with a quadratic constraint, *Numerische Mathematik* 36 (3) (1980) 291–307.
- [28] G. Golub, C. van Loan, Matrix Computations, 3rd Edition, The Johns Hopkins University Press, 1996.
- [29] P. C. Hansen, The L-curve and its use in the numerical treatment of inverse problems, IMM, Department of Mathematical Modelling, Technical University of Denmark, 1999.
- [30] D. Calvetti, S. Morigi, L. Reichel, F. Sgallari, Tikhonov regularization and the l-curve for large discrete ill-posed problems, *Journal of Computational and Applied Mathematics* 123 (12) (2000) 423 – 446, numerical Analysis 2000. Vol. III: Linear Algebra. doi:[http://dx.doi.org/10.1016/S0377-0427\(00\)00414-3](http://dx.doi.org/10.1016/S0377-0427(00)00414-3).
URL <http://www.sciencedirect.com/science/article/pii/S0377042700004143>
- [31] S. Oraintara, W. Karl, D. Castanon, T. Nguyen, A method for choosing the regularization parameter in generalized tikhonov regularized linear inverse problems, in: *Image Processing, 2000. Proceedings. 2000 International Conference on*, Vol. 1, 2000, pp. 93–96 vol.1. doi:10.1109/ICIP.2000.900900.
- [32] O. Kirkeby, P. A. Nelson, H. Hamada, F. Orduna-Bustamante, Fast deconvolution of multichannel systems using regularization, *IEEE Transactions on Speech and Audio Processing* 6 (2) (1998) 189–194.
- [33] P. M. Dumuid, B. S. Cazzolato, A. C. Zander, A comparison of filter design structures for multi-channel acoustic communication systems, *The Journal of the Acoustical Society of America* 123 (1) (2008) 174–185.
- [34] The MIT-BIH Normal Sinus Rhythm Database.
URL <http://www.physionet.org/physiobank/database/nsrdb/>
- [35] The MIT-BIH Arrhythmia Database.
URL <http://www.physionet.org/physiobank/database/mitdb/>
- [36] The MIT-BIH PTB Diagnosis Database.
URL <http://www.physionet.org/physiobank/database/ptbdb/>
- [37] G. Moody, W. Muldrow, R. Mark, A noise stress test for arrhythmia detectors, in: *Computers in Cardiology, 1984*, pp. 381–384.
- [38] R. Sameni, G. D. Clifford, A Review of Fetal ECG Signal Processing; Issues and Promising Directions, *The Open Pacing, Electrophysiology & Therapy Journal (TOPETJ)* 3 (2010) 4–20. doi:10.2174/1876536X01003010004.
- [39] R. Sameni, The Open-Source Electrophysiological Toolbox (OSET), version 3.1 (2014).
URL <http://www.oaset.ir>
- [40] A. Savitzky, M. J. Golay, Smoothing and differentiation of data by simplified least squares procedures., *Analytical chemistry* 36 (8) (1964) 1627–1639.