



ECG Enhancement and QRS Detection Based on Sparse Derivatives[☆]



Xiaoran Ning*, Ivan W. Selesnick

Polytechnic Institute of New York University, 6 Metrotech Center, Brooklyn, NY 11201, United States

ARTICLE INFO

Article history:

Received 23 January 2013

Received in revised form 12 June 2013

Accepted 14 June 2013

Keywords:

ECG enhancement
QRS detection
 ℓ_1 norm optimization
Sparse derivative
Denosing

ABSTRACT

Electrocardiography (ECG) signals are often contaminated by various kinds of noise or artifacts, for example, morphological changes due to motion artifact, non-stationary noise due to muscular contraction (EMG), etc. Some of these contaminations severely affect the usefulness of ECG signals, especially when computer aided algorithms are utilized. In this paper, a novel ECG enhancement algorithm is proposed based on sparse derivatives. By solving a convex ℓ_1 optimization problem, artifacts are reduced by modeling the clean ECG signal as a sum of two signals whose second and third-order derivatives (differences) are sparse respectively. The algorithm is applied to a QRS detection system and validated using the MIT-BIH Arrhythmia database (109,452 annotations), resulting a sensitivity of $Se = 99.87\%$ and a positive prediction of $+P = 99.88\%$.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

In real world scenarios, an ECG recording is often corrupted by various kinds of noise and artifacts [1], e.g., electrode contact noise, motion artifacts, muscle contraction (EMG), etc., which severely limit its usefulness. Simple frequency-selective filtering and some general denoising technics [2–5] aiming mainly for stationary noise usually underperform because of (1) the partial overlap of signal and noise bandwidths. (2) The presence of the non-stationary noise or artifacts. Thus in recent years, more sophisticated algorithms have been devised to enhance ECG signals, for example [6–8], etc. Some of them are based on projecting the ECG signal onto different signal subspaces, some are based on statistical or nonlinear dynamical models. Among all these models, the polynomial approximation based methods [9,3,4] are shown suitable for applications where preserving features of local maximas and minimas of an ECG are important [10,11].

Since there is no absolute clean ECG signal, a question should be clarified is, in what sense or by what means the performance of an algorithm can be evaluated. One approach is to simulate both synthetic ECG signal and noise and then measure the difference, for example, the mean square error (MSE) between the recovered signal and the original signal. However, the quality of the simulation can sometimes affect the evaluation. Another approach is to test on

real ECG signals and verify the result in applications, for example, ECG delineation, compression, QRS complex detection etc.

For the past decade, sparse signal representations for signal processing has been an active research field. Many algorithms based on sparsity have been developed for signal denoising, reconstruction, deconvolution, etc. For example, the lasso [12] or basis pursuit denoising (BPD) problem [13] solve the following optimization:

$$\operatorname{argmin}_{\mathbf{x}} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda R(\mathbf{x}) \quad (1)$$

where \mathbf{x} is a signal of length N and \mathbf{y} is the noisy observation of \mathbf{x} . The regularizer

$$R(\mathbf{x}) = \sum_{i=0}^{N-1} |x_i|$$

is chosen to promote the sparsity of the solution of \mathbf{x} . The positive constant λ is the regularization parameter. Increasing λ has the effect of making the solution \mathbf{x} more sparse.

When $R(\mathbf{x})$ is modified as:

$$R(\mathbf{x}) = \sum_{i=0}^{N-1} |x_i - x_{i-1}|$$

instead of sparsity of the solution of \mathbf{x} , it promotes the sparsity of the first-order difference of the solution of \mathbf{x} . The problem is then known as total variation (TV) denoising [14]. Some other examples are given in [15–20], which discussed nonlinear signal decomposition based on sparse representations.

This paper proposes the enhancement of ECG signals by modeling the ECG signal $y(t)$ as the sum of two functions, i.e.,

[☆] This research is supported by NSF under grant CCF-1018020.

* Corresponding author. Tel.: +1 347 949 0860.

E-mail addresses: xning01@gmail.com (X. Ning), selesi@poly.edu (I.W. Selesnick).

$y(t) = x_1(t) + x_2(t)$, where $x_1(t)$ is constrained such that its second-order derivative is sparse, and $x_2(t)$ is constrained such that its third-order derivative is sparse. As such, $x_1(t)$ is approximately a linear spline (piecewise linear), while $x_2(t)$ is approximately a quadratic spline (piecewise quadratic). In this approach, the function $x_1(t)$ (i.e., a linear spline-like function) is understood to be continuous but not differentiable, while the function $x_2(t)$ (i.e., a quadratic spline-like function) is continuously differentiable. The simultaneous use of two splines, of differing orders, gives the proposed model the ability to accurately reproduce the behavior of the ECG signal: the ECG waveform typically contains peaks (the QRS segment) where a piecewise linear model is suitable. Yet the ECG waveform also contains components that are not well approximated by a linear function; for these segments, a second order polynomial is more suitable.

The proposed approach for ECG enhancement models the ECG signal as (approximately) the sum of two splines, of different orders, but it does not explicitly constrain the two functions to splines exactly. Exact spline models have two disadvantages: They are overly simplified for real ECG waveforms, and they require a suitable set of knots being identified. For ECG signals contaminated by noise and artifacts, most parametric models will likely have these deficiencies. Hence, the proposed approach reflects this multi-spline model (linear spline plus quadratic splines) by jointly optimizing the sparsity of the 2nd- and 3rd-order derivatives of the functions $x_1(t)$ and $x_2(t)$ respectively.

For the joint optimization of the sparsity of the derivatives of the model components, $x_1(t)$ and $x_2(t)$, we describe a suitable convex non-differentiable optimization problem. Using the Alternating Direction Method of Multipliers (ADMM) [21–23], we derive a computationally efficient algorithm that is guaranteed to converge to the global minimizer. The proposed approach is evaluated for the purpose of ECG noise/artifact attenuation and for QRS detection.

The paper is organized as follows. In Section 2, some notations are defined and the total variation denoising problem is re-visited with more detail. In Section 3, the proposed ECG enhancement algorithm, namely Sparse Derivative Denoising (SDD), is discussed with examples. In Section 4, we describe a complete automatic QRS detection system to show how the proposed ECG enhancement method helps increase the overall detection accuracy. As the QRS detection problem is itself of value, the system is applied to the whole MIT-BIH Arrhythmia database and the performance is compared to some other state-of-art methods.

2. Preliminaries

2.1. Notation

In this paper, lower case bold and upper case bold are used to denote vector and matrix, e.g., vector \mathbf{g} and matrix \mathbf{G} . The N point signal \mathbf{x} is denoted as

$$\mathbf{x} = [x_0, x_1, \dots, x_{N-1}]^t \quad (2)$$

Since the work is illustrated in discrete time domain, all derivatives will be expressed by differences and the words ‘derivative’ and ‘difference’ are exchangeable in the following.

The first-order difference matrix of size $(N-1) \times N$ is defined as

$$\mathbf{D}_1 = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & & \ddots & \ddots \\ & & & & -1 & 1 \end{bmatrix}$$

Similarly, the second-order difference matrix of size $(N-2) \times N$ is defined as

$$\mathbf{D}_2 = \begin{bmatrix} -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & & \ddots & \ddots & \ddots \\ & & & & -1 & 2 & -1 \end{bmatrix}$$

Generally, the difference operator of order k , of size $(N-k) \times N$, is denoted as \mathbf{D}_k .

The ℓ_p norm, when $p \geq 1$ is defined as

$$\|\mathbf{x}\|_p = (|x_1|^p + |x_2|^p + \dots + |x_N|^p)^{\frac{1}{p}} \quad (3)$$

Specifically, when $p=1$, (3) becomes

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \dots + |x_N|. \quad (4)$$

When $p=2$, (3) becomes

$$\|\mathbf{x}\|_2 = \sqrt{|x_1|^2 + |x_2|^2 + \dots + |x_N|^2}. \quad (5)$$

The notation $\|\mathbf{x}\|_0$ is used to denote the total number of non-zero entries in \mathbf{x} .

2.2. Total variation denoising

Total variation denoising refers to the problem

$$\operatorname{argmin}_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \|\mathbf{D}_1 \mathbf{x}\|_1 \quad (6)$$

or

$$\operatorname{argmin}_{\mathbf{x}} \|\mathbf{D}_1 \mathbf{x}\|_1 \quad (7a)$$

$$\text{subject to } \|\mathbf{y} - \mathbf{x}\|_2 \leq r \quad (7b)$$

Problems (6) and (7) solve the same problem given that a suitable λ is chosen and the Euclidean distance between \mathbf{y} and \mathbf{x} (value of r) is known. They seek to find an estimate of the original signal \mathbf{x} from noisy data \mathbf{y} , where \mathbf{x} is modeled to have a sparse first-order difference.

There is no explicit solution to (6), but a direct algorithm to compute the exact solution is recently available [24]. We will denote the solution to (6) as

$$\operatorname{TVD}(\mathbf{y}, \lambda)$$

The method of total variation denoising can be extended by solving the following cost function, namely, compound TV denoising [25]:

$$\operatorname{argmin}_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda_1 \|\mathbf{D}_1 \mathbf{x}\|_1 + \lambda_2 \|\mathbf{D}_2 \mathbf{x}\|_1 \quad (8)$$

It simultaneously penalizes the first-order and second-order difference of \mathbf{x} , implying that \mathbf{x} has both sparse first-order and second-order differences.

3. Sparse derivative decomposition and denoising

3.1. First-order and second-order sparse derivatives

Consider the test signal \mathbf{y} in Fig. 1(a) which is composed of piecewise constant and piecewise linear components. The first and second-order difference of \mathbf{y} are shown in Fig. 1(b) and (c). It is clear that $\mathbf{D}_1 \mathbf{y}$ and $\mathbf{D}_2 \mathbf{y}$ are more sparse than \mathbf{y} , however, two more things can be observed: (1) $\mathbf{D}_1 \mathbf{y}$ is sparse only where \mathbf{y} is piecewise constant. (2) $\mathbf{D}_2 \mathbf{y}$ is less sparse compared to $\mathbf{D}_1 \mathbf{y}$ at the place where $\mathbf{D}_1 \mathbf{y}$

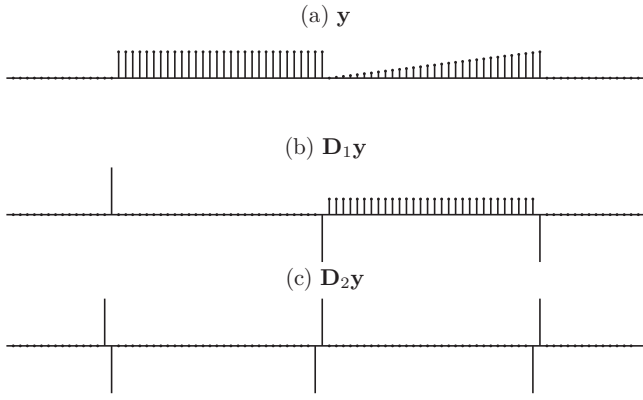


Fig. 1. (a) The test signal \mathbf{y} . (b) The first-order difference of \mathbf{y} . (c) The second-order difference of \mathbf{y} . Amplitudes of the signals are scaled to emphasize the sparsity.

is sparse. Next, we let $\mathbf{y} = \mathbf{x}_1 + \mathbf{x}_2$ as shown in Fig. 2(a) and (b) where \mathbf{x}_1 is the piecewise constant component of \mathbf{y} and \mathbf{x}_2 is the piecewise linear component. The first-order difference of \mathbf{x}_1 and second-order difference of \mathbf{x}_2 are shown in Fig. 2(c) and (d). Now comparing Figs. 1 and 2, the total number of spikes in Fig. 2(c) and (d), which equals $\|\mathbf{D}_1 \mathbf{x}_1\|_0 + \|\mathbf{D}_2 \mathbf{x}_2\|_0$ is less than the number of spikes in either $\mathbf{D}_1 \mathbf{y}$ or $\mathbf{D}_2 \mathbf{y}$ alone. The reason is that a piecewise constant signal becomes most sparse after taking first-order difference while a piecewise linear signal becomes most sparse after taking second-order difference. In another word, if somehow we are able to decompose a signal \mathbf{y} that contains both piecewise constant component and piecewise linear component into one piecewise constant signal \mathbf{x}_1 and one piecewise linear signal \mathbf{x}_2 , then $\|\mathbf{D}_1 \mathbf{x}_1\|_0 + \|\mathbf{D}_2 \mathbf{x}_2\|_0$ is smaller than $\|\mathbf{y}\|_0$ and $\|\mathbf{D}_k \mathbf{y}\|_0$ for any k . To achieve this decomposition, we propose the following optimization problem:

$$\operatorname{argmin}_{\mathbf{x}_1, \mathbf{x}_2} \|\mathbf{D}_1 \mathbf{x}_1\|_0 + \|\mathbf{D}_2 \mathbf{x}_2\|_0 \quad (9a)$$

$$\text{subject to } \mathbf{y} = \mathbf{x}_1 + \mathbf{x}_2 \quad (9b)$$

However, (9) is not a convex optimization problem and its solution is not easily accessed. As ℓ_1 is known as a convex proxy of sparsity, we modify (9) and propose a convex optimization problem. More specifically, we use $\|\mathbf{D}_1 \mathbf{x}_1\|_1$ and $\|\mathbf{D}_2 \mathbf{x}_2\|_1$ instead

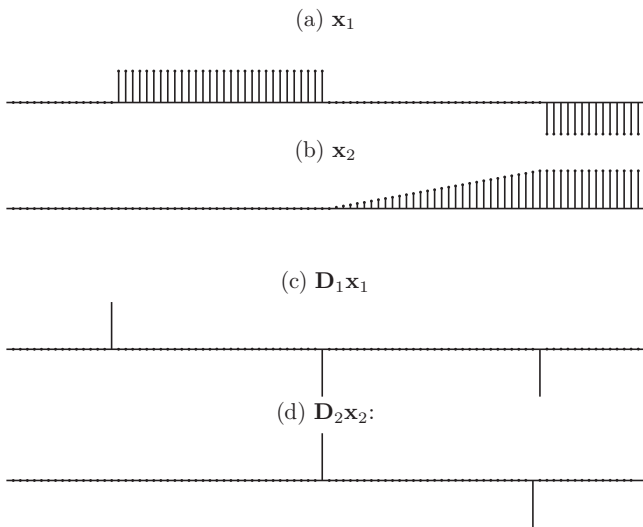


Fig. 2. (a) \mathbf{x}_1 , the piecewise constant component of \mathbf{y} . (b) \mathbf{x}_2 , the piecewise linear component of \mathbf{y} . (c) The first-order difference of \mathbf{x}_1 . (d) The second-order difference of \mathbf{x}_2 . Amplitudes of the signals are scaled to emphasize the sparsity.

of $\|\mathbf{D}_1 \mathbf{x}_1\|_0$ and $\|\mathbf{D}_2 \mathbf{x}_2\|_0$ in (9) and introduce two regularization parameters λ_1 and λ_2 :

$$\operatorname{argmin}_{\mathbf{x}_1, \mathbf{x}_2} \lambda_1 \|\mathbf{D}_1 \mathbf{x}_1\|_1 + \lambda_2 \|\mathbf{D}_2 \mathbf{x}_2\|_1 \quad (10a)$$

$$\text{subject to } \mathbf{y} = \mathbf{x}_1 + \mathbf{x}_2 \quad (10b)$$

Eq. (10) is the sparse derivative decomposition problem.

Next, we consider another model wherein the signal $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ is observed in additive noise,

$$\mathbf{y} = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{n} \quad (11)$$

We seek to find an estimates of \mathbf{x}_1 and \mathbf{x}_2 , denoted as $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$, such that $\mathbf{y} - \hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2$ resembles the noise \mathbf{n} . Moreover, $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ are modeled to have sparse first- and second-order differences respectively. This can be formulated as following:

$$\operatorname{argmin}_{\mathbf{x}_1, \mathbf{x}_2} \|\mathbf{y} - \mathbf{x}_1 - \mathbf{x}_2\|_2 + \lambda_1 \|\mathbf{D}_1 \mathbf{x}_1\|_1 + \lambda_2 \|\mathbf{D}_2 \mathbf{x}_2\|_1 \quad (12)$$

or

$$\operatorname{argmin}_{\mathbf{x}_1, \mathbf{x}_2} \lambda_1 \|\mathbf{D}_1 \mathbf{x}_1\|_1 + \lambda_2 \|\mathbf{D}_2 \mathbf{x}_2\|_1 \quad (13a)$$

$$\text{subject to } \|\mathbf{y} - \mathbf{x}_1 - \mathbf{x}_2\|_2 \leq r \quad (13b)$$

Problems (12) and (13) are referred as Sparse Derivative Denoising (SDD). A solution to problem (12) has been given in [26] and the algorithm to (13) will be given in Section 3.2.

Example 3.1. The test signal from Fig. 1(a) is corrupted by white Gaussian noise with signal to noise ratio (SNR) 20 dB as shown in Fig. 3(a). The denoised signals after applying SDD and some other denoising technics (total variation denoising [14], compound total variation denoising [25], EMD based denoising [5] and a wavelet method) are shown in Fig. 3(c) and Fig. 4. Compared to other methods, the SDD achieves better recovery for both constant component and linear component and the SNR also indicates that SDD outperforms the others.

3.2. Second-order and third-order sparse derivatives

In this section, the above concept is extended to higher order derivatives so that the features of ECG signal can be properly captured.

Consider the problem

$$\operatorname{argmin}_{\mathbf{x}_1, \mathbf{x}_2} \lambda_1 \|\mathbf{D}_2 \mathbf{x}_1\|_1 + \lambda_2 \|\mathbf{D}_3 \mathbf{x}_2\|_1 \quad (14a)$$

$$\text{subject to } \|\mathbf{y} - \mathbf{x}_1 - \mathbf{x}_2\|_2 \leq r \quad (14b)$$

The cost function models the observed signal \mathbf{y} as $\mathbf{y} = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{n}$, where \mathbf{x}_1 is a quasi piecewise linear signal which has sparse second-order difference, \mathbf{x}_2 is a quasi piecewise quadratic signal which has sparse third-order difference and \mathbf{n} is the noise or artifacts in ECG signal.

Problem (14) can be solved using variable splitting with Alternating Direction Method of Multipliers (ADMM). After variable splitting, (14) is rewritten as

$$\operatorname{argmin}_{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{x}_1, \mathbf{x}_2} \lambda_1 \|\mathbf{D}_1 \mathbf{v}_1\|_1 + \lambda_2 \|\mathbf{D}_1 \mathbf{v}_2\|_1 \quad (15a)$$

$$\text{subject to } \|\mathbf{y} - \mathbf{v}_4\|_2 \leq r \quad (15b)$$

$$\mathbf{v}_1 = \mathbf{D}_1 \mathbf{x}_1 \quad (15c)$$

$$\mathbf{v}_2 = \mathbf{D}_2 \mathbf{x}_2 \quad (15d)$$

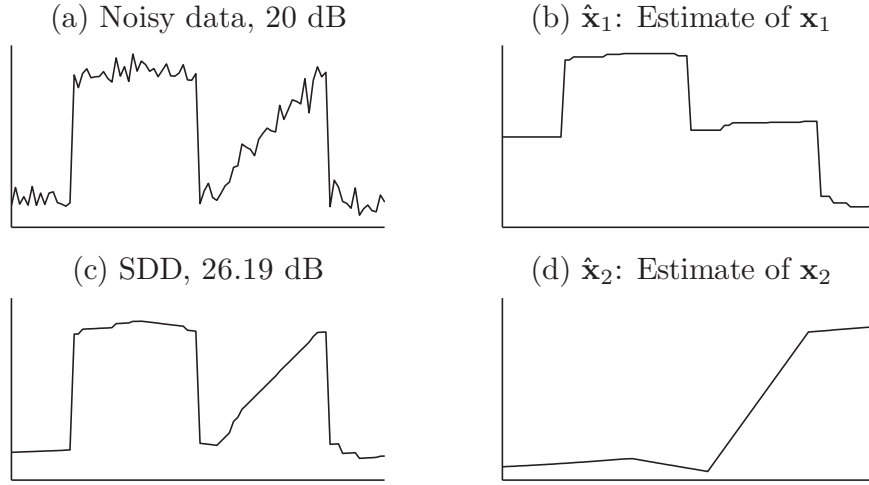


Fig. 3. Example 3.1. Illustration of sparse derivative denoising.

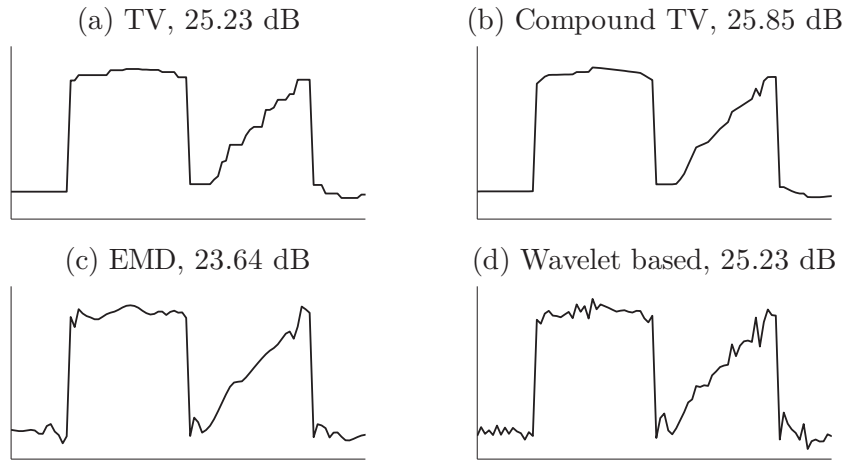


Fig. 4. Example 3.1. (cont'd). Denoised signal after applying (a) total variation denoising [14], (b) compound total variation denoising [25], (c) EMD based denoising [5], (d) wavelet based denoising (symmlets wavelet with vanishing moment 4).

$$\mathbf{v}_3 = \mathbf{x}_1 \quad (15e)$$

$$\mathbf{v}_4 = \mathbf{x}_1 + \mathbf{x}_2 \quad (15f)$$

Note that the constraints can be expressed in a matrix form,

$$\mathbf{v} = \mathbf{G}\mathbf{x} \quad (16)$$

where

$$\mathbf{G} = \begin{bmatrix} \mathbf{D}_1 & 0 \\ 0 & \mathbf{D}_2 \\ \mathbf{I} & 0 \\ \mathbf{I} & \mathbf{I} \end{bmatrix} \quad (17)$$

$$\mathbf{v} = [\mathbf{v}_1^t, \mathbf{v}_2^t, \mathbf{v}_3^t, \mathbf{v}_4^t]^t; \quad (18)$$

$$\mathbf{x} = [\mathbf{x}_1^t, \mathbf{x}_2^t]^t; \quad (19)$$

The augmented Lagrangian for (15) is

$$\begin{aligned} L = & \lambda_1 \|\mathbf{D}_1 \mathbf{v}_1\|_1 + \lambda_2 \|\mathbf{D}_1 \mathbf{v}_2\|_1 + \mu \|\mathbf{v}_1 - \mathbf{D}_1 \mathbf{x}_1 + \mathbf{d}_1\|_2^2 + \mu \|\mathbf{v}_2 \\ & - \mathbf{D}_2 \mathbf{x}_2 + \mathbf{d}_2\|_2^2 + \mu \|\mathbf{v}_3 - \mathbf{x}_1 + \mathbf{d}_3\|_2^2 + \mu \|\mathbf{v}_4 - \mathbf{x}_1 - \mathbf{x}_2 + \mathbf{d}_4\|_2^2 \\ & \text{subject to } \|\mathbf{y} - \mathbf{v}_4\|_2 \leq r \end{aligned} \quad (20)$$

Next, L is minimized by ADMM as following:

$$\begin{aligned} \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4 \leftarrow & \underset{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4}{\operatorname{argmin}} \lambda_1 \|\mathbf{D}_1 \mathbf{v}_1\|_1 + \lambda_2 \|\mathbf{D}_1 \mathbf{v}_2\|_1 \\ & + \mu \|\mathbf{v}_1 - \mathbf{D}_1 \mathbf{x}_1 + \mathbf{d}_1\|_2^2 + \mu \|\mathbf{v}_2 - \mathbf{D}_2 \mathbf{x}_2 + \mathbf{d}_2\|_2^2 \\ & + \mu \|\mathbf{v}_3 - \mathbf{x}_1 + \mathbf{d}_3\|_2^2 + \mu \|\mathbf{v}_4 - \mathbf{x}_1 - \mathbf{x}_2 + \mathbf{d}_4\|_2^2 \\ & \text{subject to } \|\mathbf{y} - \mathbf{v}_4\|_2 \leq r \end{aligned} \quad (21a)$$

$$\begin{aligned} \mathbf{x}_1, \mathbf{x}_2 \leftarrow & \underset{\mathbf{x}_1, \mathbf{x}_2}{\operatorname{argmin}} \mu \|\mathbf{v}_1 - \mathbf{D}_1 \mathbf{x}_1 + \mathbf{d}_1\|_2^2 + \mu \|\mathbf{v}_3 - \mathbf{x}_1 + \mathbf{d}_3\|_2^2 \\ & + \mu \|\mathbf{v}_4 - \mathbf{x}_1 - \mathbf{x}_2 + \mathbf{d}_4\|_2^2 \end{aligned} \quad (21b)$$

$$\mathbf{d}_1 \leftarrow \mathbf{d}_1 + \mathbf{v}_1 - \mathbf{D}_1 \mathbf{x}_1 \quad (21c)$$

$$\mathbf{d}_2 \leftarrow \mathbf{d}_2 + \mathbf{v}_2 - \mathbf{D}_2 \mathbf{x}_2 \quad (21d)$$

$$\mathbf{d}_3 \leftarrow \mathbf{d}_3 + \mathbf{v}_3 - \mathbf{x}_1 \quad (21e)$$

$$\mathbf{d}_4 \leftarrow \mathbf{d}_4 + \mathbf{v}_4 - \mathbf{x}_1 - \mathbf{x}_2 \quad (21f)$$

$$\text{go back to (21a)} \quad (21g)$$

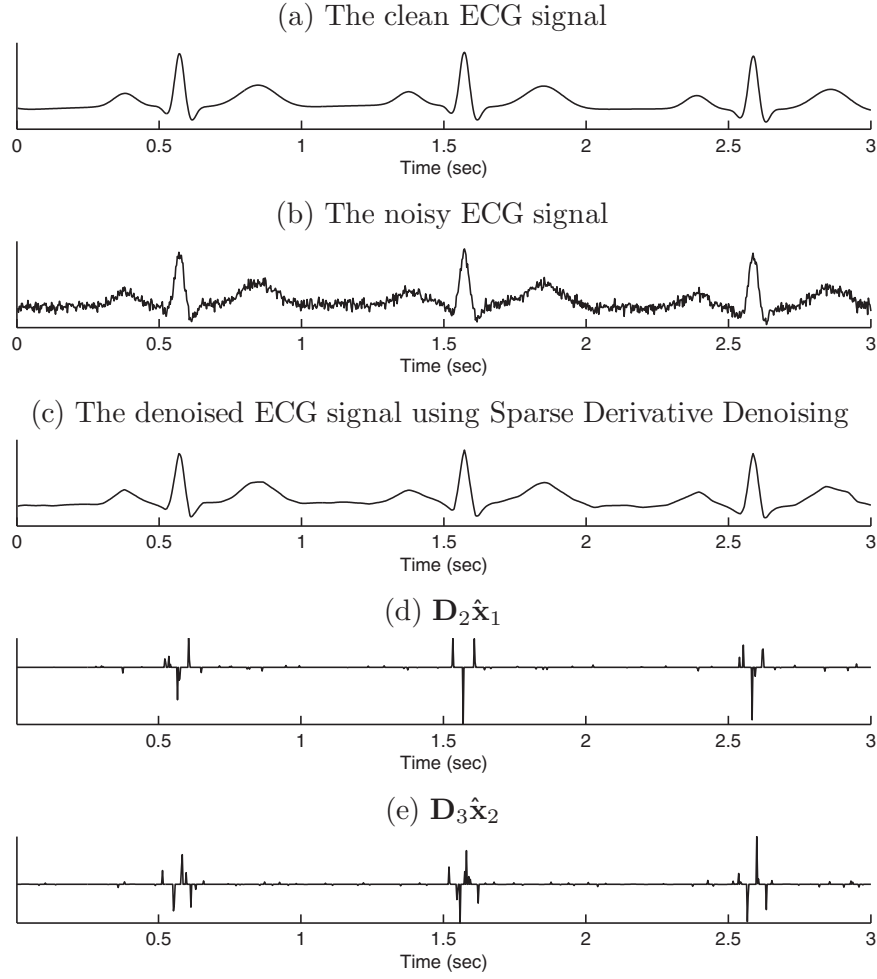


Fig. 5. Example 3.2. Illustration of SDD for ECG denoising.

The minimization of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ are separable, thus can be solved independently. However, $\mathbf{x}_1, \mathbf{x}_2$ are not separable, thus have to be solved jointly. Letting $\mathbf{d} = [\mathbf{d}_1^t, \mathbf{d}_2^t, \mathbf{d}_3^t, \mathbf{d}_4^t]^t$, problem (21) can be written as (22):

$$\mathbf{v}_1 \leftarrow \underset{\mathbf{v}_1}{\operatorname{argmin}} \quad \lambda_1 \|\mathbf{D}_1 \mathbf{v}_1\|_1 + \mu \|\mathbf{v}_1 - \mathbf{D}_1 \mathbf{x}_1 + \mathbf{d}_1\|_2^2 \quad (22a)$$

$$\mathbf{v}_2 \leftarrow \underset{\mathbf{v}_2}{\operatorname{argmin}} \quad \lambda_2 \|\mathbf{D}_2 \mathbf{v}_2\|_1 + \mu \|\mathbf{v}_2 - \mathbf{D}_2 \mathbf{x}_2 + \mathbf{d}_2\|_2^2 \quad (22b)$$

$$\mathbf{v}_3 \leftarrow \underset{\mathbf{v}_3}{\operatorname{argmin}} \quad \|\mathbf{v}_3 - \mathbf{x}_1 + \mathbf{d}_3\|_2^2 \quad (22c)$$

$$\mathbf{v}_4 \leftarrow \underset{\mathbf{v}_4}{\operatorname{argmin}} \quad \|\mathbf{v}_4 - \mathbf{x}_1 - \mathbf{x}_2 + \mathbf{d}_4\|_2^2 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{v}_4\|_2 \leq r \quad (22d)$$

$$\mathbf{x} \leftarrow \underset{\mathbf{x}}{\operatorname{argmin}} \quad \|\mathbf{v} - \mathbf{G}\mathbf{x} + \mathbf{d}\|_2^2 \quad (22e)$$

$$\mathbf{d} \leftarrow \mathbf{d} + \mathbf{v} - \mathbf{G}\mathbf{x} \quad (22f)$$

$$\text{go back to (22a)} \quad (22g)$$

Problems (22a) and (22b) are TV denoising problems as described in Section 2.2 and the solutions are given by

$$\mathbf{v}_1 \leftarrow \operatorname{TVD} \left(\mathbf{D}_1 \mathbf{x}_1 - \mathbf{d}_1, \frac{\lambda_1}{\mu_1} \right) \quad (23)$$

$$\mathbf{v}_2 \leftarrow \operatorname{TVD} \left(\mathbf{D}_2 \mathbf{x}_2 - \mathbf{d}_2, \frac{\lambda_2}{\mu_2} \right) \quad (24)$$

Problem (22c) is trivial and the solution is the simple update:

$$\mathbf{v}_3 \leftarrow \mathbf{x}_1 - \mathbf{d}_3 \quad (25)$$

The solution to (22d) can be expressed as

$$\mathbf{v}_4 \leftarrow \operatorname{projball}(\mathbf{x}_1 + \mathbf{x}_2 - \mathbf{d}_4, \mathbf{y}, r) \quad (26)$$

where $\operatorname{projball}(\mathbf{a}, \mathbf{b}, r)$ denotes the projection of point \mathbf{a} onto the ball center at \mathbf{b} with radius r . It is given explicitly by

$$\operatorname{projball}(\mathbf{a}, \mathbf{b}, r) := \begin{cases} \mathbf{b} + \frac{r}{\|\mathbf{a} - \mathbf{b}\|_2} (\mathbf{a} - \mathbf{b}), & \|\mathbf{a} - \mathbf{b}\|_2 > r \\ \mathbf{a}, & \|\mathbf{a} - \mathbf{b}\|_2 \leq r \end{cases} \quad (27)$$

Finally, (22e) is an ℓ_2 minimization problem and has a closed form solution

$$\mathbf{x} \leftarrow (\mathbf{G}^t \mathbf{G})^{-1} [\mathbf{G}^t (\mathbf{v} + \mathbf{d})] \quad (28)$$

Note that

$$\mathbf{G}^t \mathbf{G} = \begin{bmatrix} \mathbf{D}_1^t \mathbf{D}_1 + 2\mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{D}_2^t \mathbf{D}_2 + 2\mathbf{I} \end{bmatrix} \quad (29)$$

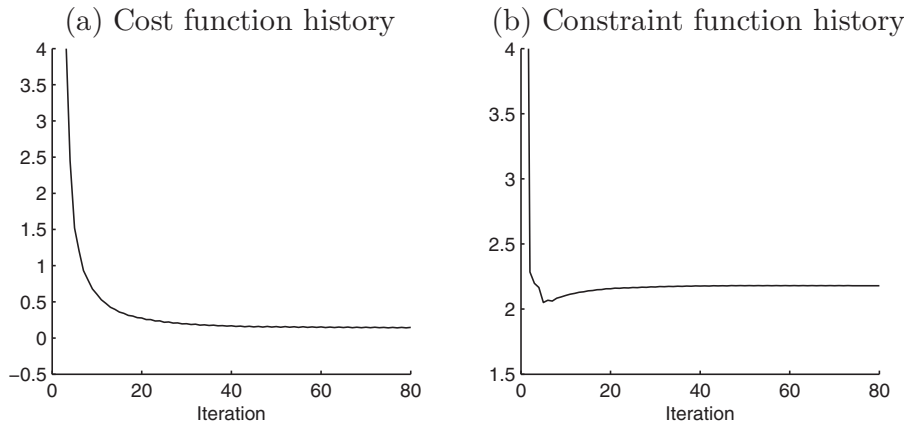


Fig. 6. Example 3.2 (cont'd). (a) The cost function $\lambda_1 \|\mathbf{D}_2 \hat{\mathbf{x}}_1\|_1 + \lambda_2 \|\mathbf{D}_3 \hat{\mathbf{x}}_2\|_1$ versus iteration. (b) constraint $\|\mathbf{y} - \hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2\|_2$ versus iteration.

is a full rank matrix and since all block matrices in $\mathbf{G}^t \mathbf{G}$ are sparse matrices, $\mathbf{G}^t \mathbf{G}$ is sparse. Therefore, existing efficient algorithms for solving sparse system of linear equations can be applied to (28). Complete algorithm for problem (14) is listed as Algorithm 3.1.

Convergence. Since (14) is a convex optimization problem, the local minimum is also the global minimum. In addition, the matrix \mathbf{G} in Eq. (17) is of full column rank. With these two conditions, the convergence of the cost function to its global minimum is guaranteed by Theorem 1 in [22].

The following examples illustrate the use of SDD for ECG signal denoising:

Example 3.2. A 3 s synthetic ECG signal with sampling rate 360 Hz (Fig. 5(a)) is generated using the algorithm from [27]. The signal is contaminated by additive white Gaussian noise as in Fig. 5(b). After applying SDD, the denoised signal $\hat{\mathbf{x}}_1 + \hat{\mathbf{x}}_2$, the second-order derivative of $\hat{\mathbf{x}}_1$ and the third-order derivative of $\hat{\mathbf{x}}_2$ are shown in Fig. 5(c)–(e). From Fig. 5(d) and (e), it is clear that the second and third-order derivative of $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ are sparse and spikes almost only exist where the morphology behaviors of the ECG change. The cost function history and constraint function history are plotted in Fig. 6.

Example 3.3. This example shows SDD's ability to deal with non-stationary noise which exists in real ECG recordings. The test signal (Fig. 7(a)) is obtained from record 231 of the MIT-BIH Arrhythmia database [28]. As is described in Section 3.C of [29], signals of this type pose major difficulty to most QRS detection methods because of the low SNR, or more specifically, the non-stationary noise (EMG for example). By carefully examining the test signal, noise from 4 to 5.5 s has significantly higher power than average. After applying SDD, Fig. 7(b) shows that the artifact is almost completely removed while the heights of P, R, T peaks of the ECG signal are well-preserved. The results of low-pass filtering (Kaiser window with cut-off frequency 20Hz, 60dB roll-off), EMD-based denoising [5] and LASIP [3,4] are also displayed in Fig. 7(c)–(e) for comparison. While the artifacts are removed to a certain degree after low-pass filtering and EMD-based denoising, the QRS peaks are not well preserved. The LASIP method is capable of well preserving the feature points since it is a set of methods based on Savitzky-Golay filtering and adaptive selection of sliding windows, however, comparing with SDD, after applying LASIP, considerable amount of noise remains.

Algorithm 3.1. Constrained sparse derivative denoising
Initialize $\mu, \mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3, \mathbf{d}_4$

$$\mathbf{v}_1 \leftarrow \text{TVD} \left(\mathbf{D}_1 \mathbf{x}_1 - \mathbf{d}_1, \frac{\lambda_1}{\mu} \right) \quad (30a)$$

$$\mathbf{v}_2 \leftarrow \text{TVD} \left(\mathbf{D}_2 \mathbf{x}_2 - \mathbf{d}_2, \frac{\lambda_2}{\mu} \right) \quad (30b)$$

$$\mathbf{v}_3 \leftarrow \mathbf{x}_1 - \mathbf{d}_3 \quad (30c) \quad \mathbf{v}_4 \leftarrow \text{projball}(\mathbf{x}_1 + \mathbf{x}_2 - \mathbf{d}_4, \mathbf{y}, r) \quad (30d)$$

$$\mathbf{x} \leftarrow (\mathbf{G}^t \mathbf{G})^{-1} [\mathbf{G}^t (\mathbf{v} + \mathbf{d})] \quad (30e)$$

$$\mathbf{d}_1 \leftarrow \mathbf{d}_1 + \mathbf{v}_1 - \mathbf{D}_1 \mathbf{x}_1 \quad (30f)$$

$$\mathbf{d}_2 \leftarrow \mathbf{d}_2 + \mathbf{v}_2 - \mathbf{D}_2 \mathbf{x}_2 \quad (30g)$$

$$\mathbf{d}_3 \leftarrow \mathbf{d}_3 + \mathbf{v}_3 - \mathbf{x}_1 \quad (30h)$$

$$\mathbf{d}_4 \leftarrow \mathbf{d}_4 + \mathbf{v}_4 - \mathbf{x}_1 - \mathbf{x}_2 \quad (30i)$$

Repeat until convergence

3.3. SNR enhancement: re-weighted ℓ_1

The sparsity of SDD can be further enhanced by exploiting the re-weighted ℓ_1 process proposed in [30]. Consider the cost function

$$\underset{\mathbf{x}_1, \mathbf{x}_2}{\text{argmin}} \quad \lambda_1 R_1(\mathbf{x}_1) + \lambda_2 R_2(\mathbf{x}_2) \quad (31a)$$

$$\text{subject to} \quad \|\mathbf{y} - \mathbf{x}_1 - \mathbf{x}_2\|_2 \leq r \quad (31b)$$

where

$$R_1(\mathbf{v}) = \|\mathbf{W}_1 \mathbf{D}_2 \mathbf{v}\|_1 \quad (31c)$$

$$R_2(\mathbf{v}) = \|\mathbf{W}_2 \mathbf{D}_3 \mathbf{v}\|_1 \quad (31d)$$

This is a more general form of problem (14) where \mathbf{W}_1 and \mathbf{W}_2 are two diagonal matrices with diagonal $\mathbf{w}_1 = [w_{11}, w_{12}, \dots, w_{1(N-2)}]^t$ and $\mathbf{w}_2 = [w_{21}, w_{22}, \dots, w_{2(N-3)}]^t$.

In the re-weighted process, (31) is solved in each iteration with updated \mathbf{W}_1 and \mathbf{W}_2 . These two diagonal matrices are initially set to identity matrices and updated in the following manner:

$$w_{1i} = \frac{1}{|\mathbf{a}_i| + \epsilon_1}, \quad i = 1, \dots, N-2 \quad (32)$$

$$w_{2j} = \frac{1}{|\mathbf{b}_j| + \epsilon_2}, \quad j = 1, \dots, N-3 \quad (33)$$

where

$$\mathbf{a} = \mathbf{D}_2 \hat{\mathbf{x}}_1 \quad (34)$$

$$\mathbf{b} = \mathbf{D}_3 \hat{\mathbf{x}}_2 \quad (35)$$

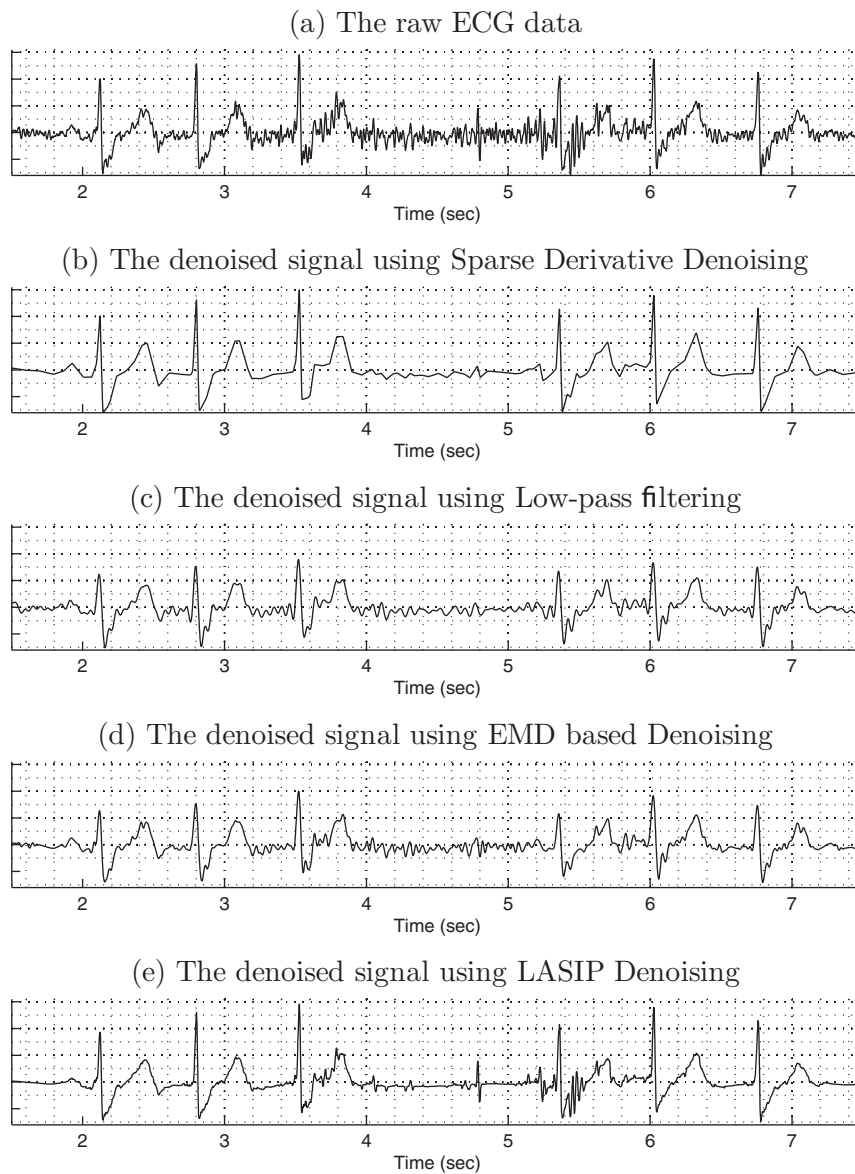


Fig. 7. Example 3.3. A compare of Sparse derivative denoising (SDD) with some other methods. (a) The raw ECG data which is contaminated by non-stationary noise. (b) The result of SDD. (c) The result of low-pass filtering using Kaiser window with cut-off frequency 20Hz. (d) The result of EMD based denoising [5]. (e) The result of LASIP [3,4].

Note that ϵ_1 and ϵ_2 are pre-defined fixed numbers. They are used to provide stability and to ensure that a zero-valued component in $\mathbf{D}_2\hat{\mathbf{x}}_1$ or $\mathbf{D}_3\hat{\mathbf{x}}_2$ in a certain iteration does not prohibit a nonzero estimate at the next iteration. According to Ref. [30], ϵ_1 and ϵ_2 should be set slightly smaller than the expected nonzero magnitudes of $\mathbf{D}_2\mathbf{x}_1$ and $\mathbf{D}_3\mathbf{x}_2$ respectively. In Fig. 8, the results of SDD and re-weighted SDD for one ECG beat are shown together where it is evident that $\mathbf{D}_2\hat{\mathbf{x}}_1$ and $\mathbf{D}_3\hat{\mathbf{x}}_2$ are more sparse after the re-weighting process.

Example 3.4. A 2048 sample (sampling rate 360 Hz) synthetic ECG test signal is generated using the algorithm given in [27], noisy observations of 0 dB, 5 dB and 10 dB are generated by adding white Gaussian noise. For each case, experiments are performed 100 times for consistency and the empirical mean is computed. Results are again compared to EMD [5], LASIP [3] and symmlets wavelet (with vanishing moment 6). The analysis is detailed in Table 1. While SDD has uniformly better performance than the other methods, after re-weighting, the SNR is further enhanced.

4. ECG QRS Detection

This section addresses ECG QRS detection. We will demonstrate and evaluate SDD for this application.

4.1. Terminology and criterion of performance evaluation

In the ECG QRS detection literature, true positive (TP), false positive (FP), false negative (FN) refer to the number of true beat detection, false beat detection and true beats that are failed to be detected. The common criterions used to evaluated the

Table 1
Example 3.4. Signal-to-noise ratio after denoising.

	0 dB	5 dB	10 dB
Re-weighted SDD	12.02 dB	16.74 dB	21.03 dB
SDD	11.88 dB	16.42 dB	20.69 dB
EMD[5]	11.85 dB	16.29 dB	20.64 dB
LASIP[3]	11.13 dB	15.04 dB	19.31 dB
Symmlets wavelet	11.05 dB	14.98 dB	19.09 dB

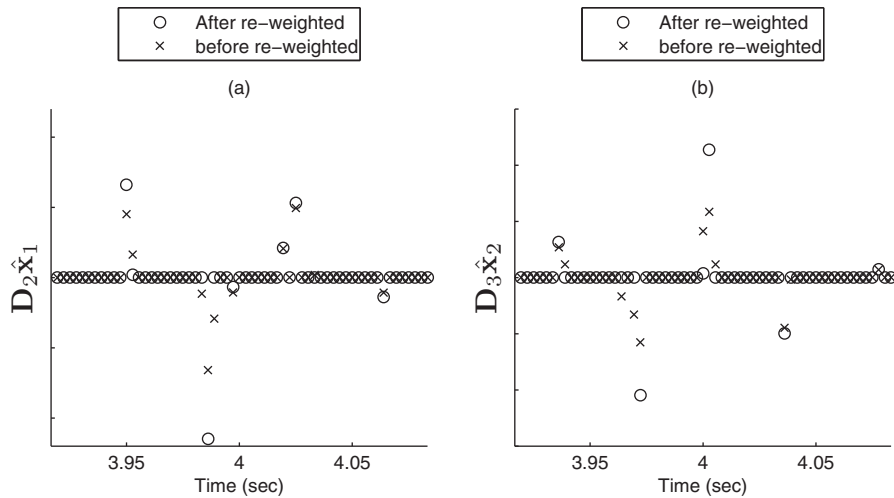


Fig. 8. One beat of ECG signal from Example 3.2. Derivatives of signal components using SDD and re-weighted SDD are plotted together. After re-weighting, the derivatives become more sparse (2 iterations, $\epsilon_1 = \epsilon_2 = 0.001$.)

overall performance include sensitivity (Se), positive prediction (+P), detection error rate and average time error that are defined as following:

$$\text{Se}(\%) = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (36)$$

$$+P(\%) = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (37)$$

$$\text{Detection Error Rate}(\%) = \frac{\text{TP} + \text{FN}}{\text{Total number of QRS complex}} \quad (38)$$

$$\begin{aligned} \text{Average Time Error (ms)} \\ = \frac{\sum |\text{detected QRS time} - \text{actual QRS time}|}{\text{TP}} \end{aligned} \quad (39)$$

4.2. Method and algorithm

4.2.1. Stage 1, pre-processing

Conventionally, the ECG recording is filtered by a band-pass filter with passband 8–20 Hz to eliminate baseline wander and high-frequency noise. Though simple and efficient, the band-pass filtering is of limited use for two reasons. First, the low-pass filtering is not necessary for any first-derivative base detection algorithms (including the detector of this work) since the first-order difference for finite length signal is by itself a one pole low-pass filter. After taking the first-order difference, the baseline wander will not affect the detection process with high probability. Second, in most cases, the artifact in real ECG signal not only exists in the frequency range higher than 20 Hz but also the range below. As has been shown in Fig. 7(c), the signal after filtering still exhibits considerable amount of noise. Thus in the pre-process stage, the band-pass filtering is replaced by SDD.

Since we prefer the detection system to be free of human intervention, the parameters λ_1 , λ_2 and r in (13) must be specified before hand or calculated adaptively based on the data. This is achieved by a simple process as following. (a) There is usually not an analytical solution for how to choose λ_1 and λ_2 since they depend on the features or morphology of the signal. However, for this problem, λ_1 and λ_2 can be pre-decided by experiment and once specified, they

are fixed and can be used for any ECG signal ($\lambda_1 = \lambda_2 = 1$ for ECG signal). (b) The value of r can be specified adaptively by employing the observed fact that the power of the noise to be removed is approximately proportional to the energy of the highpassed subband signal with passband approximately 25 Hz. That is,

$$r \approx \|\mathbf{H}_{hp}\mathbf{y}\|_2 \quad (40)$$

where \mathbf{y} is the raw ECG signal and \mathbf{H}_{hp} denotes a high-pass filter with passband 25 Hz. This confirms the prior knowledge that most of the artifacts in ECG is in the frequency band higher than 20 Hz. Note that to increase the accuracy, the raw ECG data is divided into segments (4000 samples for each segment with sampling rate 360 Hz for example) and the SDD is applied to each of the segments with the noise power of that segment independently estimated.

4.2.2. Stage 2, transformation and detection

The procedure is based on [31,32] with modification. First, the pre-processed ECG $\hat{\mathbf{y}}$ is divided into segments: \mathbf{y}_i , $i = 1, \dots, M$. For each \mathbf{y}_i , the Hilbert transform is applied to the first-order derivative of \mathbf{y}_i [31] to obtain the analytical signal \mathbf{a}_i , $i = 1, \dots, M$. Next, the variable threshold $\text{thr}(i)$ for each segment \mathbf{a}_i is calculated according to the root mean squared value $\text{RMS}(i)$ of \mathbf{a}_i or more specifically, according to (41) (in (41), $\max(i)$ denotes the maximal value of segment i). Note that the adaptive threshold described is slightly different from the one given in [31]. By introducing the small value c (in our example $c = 1$), the new detector prevents a very small amplitude fluctuation from being detected as a peak when a true peak is absent in the current segment and the RMS of the current segment is close to zero. Once the peak candidates are found by applying this first threshold, the location of the true peak can be determined according to the fact that the true peak has the largest magnitude within its 200 ms time window. Next, we examine all the R peak to R peak (R–R) intervals that have been detected. Whenever there is a R–R interval exceeding 1.5 times the previous, a second threshold

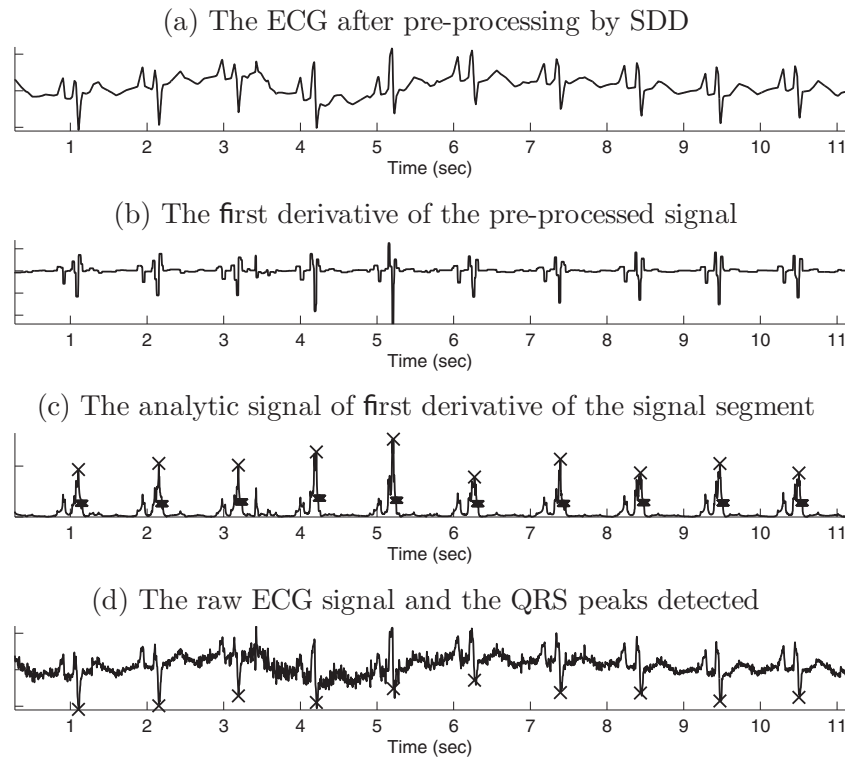


Fig. 9. Example 4.1. QRS detection in record 108. (a) The ECG signal after pre-processing by SDD. (b) The first derivative of the pre-processed signal, baseline wander does not affect the algorithm. (c) The analytic signal, the location of the peaks and the threshold for each peak. (d) The QRS peak in the original raw ECG test signal.

with the threshold value set to be 0.9 times the previous threshold is applied to this interval.

$$\text{thr}(i) = \begin{cases} 0.39 \max(i), & \text{RMS}(i) > 0.18 \max(i) \\ & \& \max(i) \leq 2 \max(i - 1) \\ 0.39 \max(i - 1), & \text{RMS}(i) > 0.18 \max(i) \\ & \& \max(i) \geq \max(i - 1) \\ 1.6 \text{RMS}(i), & c \leq \text{RMS}(i) \leq 0.18 \max(i) \\ c, & \text{otherwise} \end{cases} \quad (41)$$

In [29], the authors point out some instances of failure indicating two types of difficulty for first-derivative based methods: low SNR and irregular morphology change. Next we demonstrate how SDD and the proposed detector combination increases the detection accuracy in the presence of these difficulties.

Example 4.1. The test signal in Fig. 9 is from record 108 of MIT-BIH Arrhythmia database. This signal brings two difficulties to QRS detection. First, low SNR with abrupt noise tends to increase the number of FP. Second, the QRS complexes with reversed polarity

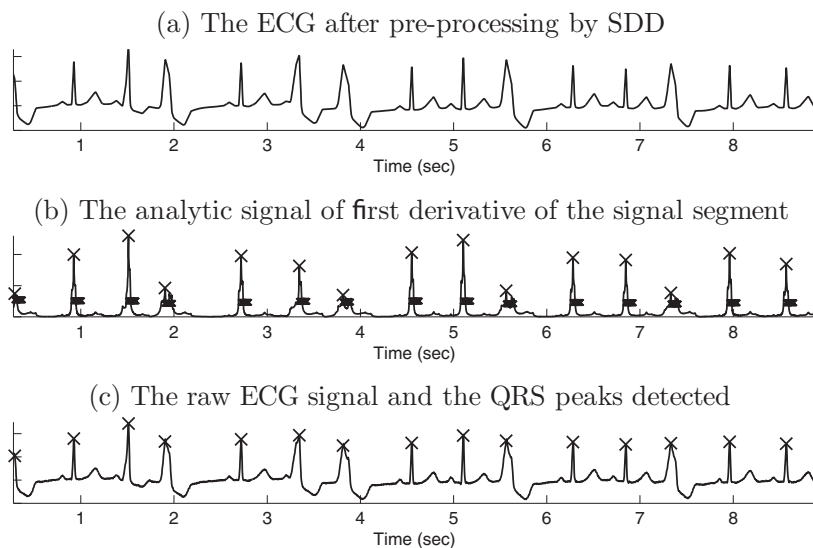


Fig. 10. Example 4.2. QRS detection in record 208. (a) The ECG signal after pre-processing by SDD. (b) The analytic signal, the location of the peaks and the threshold for each peak. (c) The QRS peak in the original raw ECG test signal.

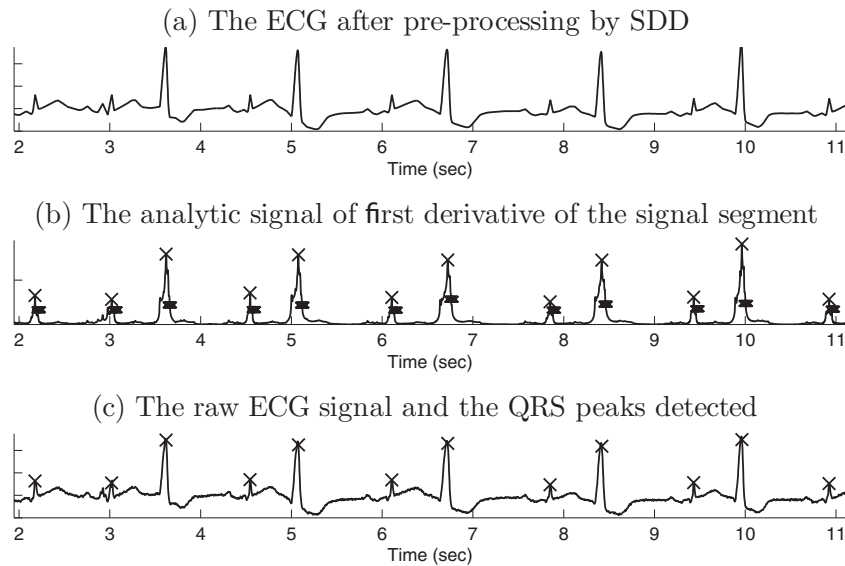


Fig. 11. Example 4.2 (cont'd). QRS detection in record 228. (a) The ECG signal after pre-processing by SDD. (b) The analytic signal, the location of the peaks and the threshold for each peak. (c) The QRS peak in the original raw ECG test signal.

Table 2
Cumulative results for QRS peak detection (N/R: not reported)

	Anotations	TP	FP	FN	Se%	+P%	Error%	Time error	Reference
Proposed	109,452	109,314	127	138	99.87	99.88	0.24	21.39 ms	
Method 1	109,456	108,499	758	957	99.13	99.31	1.50	7.00 ms	Method 1 in [29]
Method 2	109,456	108,681	836	775	99.29	99.24	1.47	7.08 ms	Method 2 in [29]
Method 3	109,456	109,099	405	354	99.68	99.63	1.47	55.82 ms	Method 3 in [29]
Method 4	109,456	108,989	447	467	99.57	99.59	0.69	7.90 ms	Method 4 in [29]
Method 5	109,456	107,344	884	2112	98.07	99.18	2.73	6.50 ms	Method 5 in [29]
Method 6	109,428	109,208	153	220	99.80	99.86	0.34	N/R	[33]
Method 7	109,819	109,635	135	184	99.83	99.88	0.29	N/R	[34]
Method 8	104,182	104,070	65	112	99.89	99.94	0.17	N/R	[35]
Method 9	109,481	109,146	137	335	99.69	99.88	0.43	N/R	[36]
Method 10	109,267	108,927	248	340	99.69	99.77	0.54	N/R	[32]

tend to increase the risk of FN. When it comes to the proposed method, the first difficulty is alleviated by SDD as shown in Fig. 9(a). Note that since the proposed method makes use of the analytic signal, which as shown in Fig. 9(c), is a positive waveform, the second difficulty is automatically solved. Fig. 9(b) verifies that baseline wander does not affect the detection procedure.

Example 4.2. The test signal in Fig. 10 is from record 208 and the test signal in Fig. 11 is from record 228. They both have potential to increase FN due to wide premature ventricular contractions (PVCs) and low-amplitude QRS complex respectively. The problems usually become more severe after frequency filtering because the QRS height is attenuated. Note that these two test signals are not 'noisy'. Due to the relatively accurate estimate of the parameter r , SDD did not remove any 'false artifact' and all the QRS peaks are well preserved. As a result, all peaks are successfully detected as shown in Fig. 10(c) and Fig. 11(c).

4.3. Overall performance evaluation

The proposed method (without re-weighting) is evaluated on the MIT-BIH Arrhythmia database for the first channel of all 48 records. Each signal is of length 30 min and sampled at 360 Hz. The first and last 10 s are truncated to avoid the end effect of the Hilbert transform. Besides, episodes of ventricular flutter in record 207 (approximately 3 min) are excluded from the analysis. The overall result of the proposed method and some other benchmark methods are listed in Table 2. In the table, Methods 1–5 [29] are

first-derivative methods with different details and detectors. Methods 6 [33], 7 [34] and 8 [35] are methods that used wavelet transform but different detection algorithms. Method 9 [36] described an algorithm which maps one dimensional ECG signals to two dimensional vectors, obtains a measure and finally locates QRS peaks based on the measure. Method 10 [32] invented a detection rule which used time-averaged signal to help better locate the true QRS peaks. Note that results of different methods cannot be compared directly because the total number of annotations of each experiment are not exactly equal and some methods did not provide information of average time error. However, from Table 2, it can still be inferred that the proposed method performs fairly well because of the relatively high Se% and +P% (only lower than method 8 [35], which is tested on a significant smaller dataset compared to others). Note that usually if we increase the average time error which means we tolerate more time difference between the true peak and the peak detected, the error rate of detection goes down correspondingly, vice versa. For example, if we allow the average time error to be up to 50 ms, the detection rate of error further goes down to 0.23%, on the other hand, if we restrict the average time error to be within 10ms, the detection rate of error increases to 0.25% or higher.

5. Discussion and Conclusion

In this paper, an ECG enhancement method, namely SDD, is proposed. First, a raw ECG signal is modeled as sum of three

components where the first two components have sparse second- and third-order derivative respectively and the third component contains only noise or artifact. Second, according to the model, a convex optimization problem is formulated. Third, to solve this problem, we derive an algorithm with guaranteed convergence by using variable splitting and ADMM.

The performance of the propose method is evaluated in two aspects. First, it is tested with synthetic ECG signals in white noise environment as in many other ECG literatures. In this case, SDD (re-weighted SDD) provides a significant SNR improvement in comparison to benchmark denoising schemes, EMD, LASIP and Symmlets wavelet. However, when real ECG recordings are considered, white Gaussian noise is usually an over-simplified model. Thus, in the second part of the evaluation, SDD is tested with real ECG recordings. We show that, by design, SDD is capable of effectively removing non-stationary noise (artifact). These advantages make SDD, being a pre-processing method, suitable for many ECG analysis and applications. To this end, an automatic QRS detection system which uses SDD in the pre-processing stage is proposed and validated on the MIT-BIH Arrhythmia database. Again, the detection result is promising compared to other state-of-art methods.

One issue for ECG QRS detection is the running time. The SDD algorithm is not designed specifically in the purpose of real-time application, however, the processing time for the test signal (4000 sample or approximately 11 s) in Example 4.1 is around 3.3 s (Matlab simulation on a i5 3.0GHz PC). This reveals the potential of SDD as a candidate for real-time application.

References

- [1] G.M. Friesen, T.C. Jannett, M.A. Jadallah, S.L. Yates, S.R. Quint, H.T. Nagle, A comparison of the noise sensitivity of nine QRS detection algorithms, *IEEE Transactions on Biomedical Engineering* 37 (1) (1990) 85–98.
- [2] K. Ashfanoo, S. Celia, Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains, *Biomedical Signal Processing and Control* 7 (5) (2012) 481–489.
- [3] K. Vladimir, F. Alessandro, E. Karen, A. Jaakko, Directional varying scale approximations for anisotropic signal processing, in: *Proceedings: XII European Signal Processing Conference, EUSIPCO 2004, Vienna, 2004*, pp. 101–104.
- [4] V. Katkovnik, K.O. Egiazarian, J. Astola, *Local Approximation Techniques in Signal and Image Processing*, SPIE Press, SPIE Field Guides, Bellingham WA, 2006.
- [5] Y. Kopsinis, S. McLaughlin, Development of EMD-based denoising methods inspired by wavelet thresholding, *IEEE Transactions on Signal Processing* 57 (April) (2009) 1351–1362.
- [6] X. Hu, V. Nenov, A single-lead ECG enhancement algorithm using a regularized data-driven filter, *IEEE Transactions on Biomedical Engineering* 53 (2) (2006) 347–351.
- [7] O. Sayadi, M.B. Shamsollahi, ECG denoising and compression using a modified extended Kalman filter structure, *IEEE Transactions on Biomedical Engineering* 55 (September) (2008) 2240–2248.
- [8] R. Vullings, B. De Vries, J.W.M. Bergmans, An adaptive Kalman filter for ECG signal enhancement, *IEEE Transactions on Biomedical Engineering* 58 (April) (2011) 1094–1103.
- [9] A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Analytical Chemistry* 36 (1964) 1627–1639.
- [10] S. Hargittai, Savitzky–Golay least-squares polynomial filters in ECG signal processing, *Computers in Cardiology* (September) (2005) 763–766.
- [11] K. Pandia, S. Ravindran, R. Cole, G. Kovacs, L. Giovangrandi, Motion artifact cancellation to obtain heart sounds from a single chest-worn accelerometer, in: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, March, 2010, pp. 590–593.
- [12] R. Tibshirani, Regression shrinkage and selection via the LASSO, *Journal of the Royal Statistical Society, Series B* 58 (1994) 267–288.
- [13] S.S. Chen, D.L. Donoho, A. Michael, Saunders, Atomic decomposition by basis pursuit, *SIAM Journal on Scientific Computing* 20 (1998) 33–61.
- [14] T. Chan, S. Osher, J. Shen, The digital TV filter and nonlinear denoising, *IEEE Transactions on Image Processing* 10 (February) (2001) 231–241.
- [15] M. Elad, J.L. Starck, P. Querre, D.L. Donoho, Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA), *Applied and Computational Harmonic Analysis* 19 (3) (2005) 340–358 (Computational Harmonic Analysis – Part 1).
- [16] M.J. Fadili, J.L. Starck, J. Bobin, Y. Moudden, Image decomposition and separation using sparse representations: an overview, *Proceedings of the IEEE* 98 (June) (2010) 983–994.
- [17] L. Granai, P. Vandergheynst, Sparse decomposition over multi-component redundant dictionaries., in: *IEEE Workshop on Multimedia Signal Processing (MMSp)*, September–October, 2004, pp. 494–497.
- [18] J.L. Starck, M. Elad, D.L. Donoho, Image decomposition via the combination of sparse representations and a variational approach, *IEEE Transactions on Image Processing* 14 (October) (2005) 1570–1582.
- [19] I.W. Selesnick, Resonance-based signal decomposition: a new sparsity-enabled signal analysis method, *Signal Processing* 91 (12) (2011) 2793–2809.
- [20] I.W. Selesnick, S. Arnold, V.R. Dantham, Polynomial smoothing of time series with additive step discontinuities, *IEEE Transactions on Signal Processing* 60 (December) (2012) 6305–6318.
- [21] M.V. Afonso, J.M. Biucas-Dias, M.A.T. Figueiredo, An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems, *IEEE Transactions on Image Processing* 20 (March) (2011) 681–695.
- [22] M.V. Afonso, J.M. Biucas-Dias, M.A.T. Figueiredo, Fast image recovery using variable splitting and constrained optimization, *IEEE Transactions on Biomedical Engineering* 19 (September) (2010) 2345–2356.
- [23] S.P. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Foundation and Trends in Machine Learning* 3 (January) (2011) 1–122.
- [24] L. Condat, A direct algorithm for 1d total variation denoising, *Technical Report, Hal-00675043*, 2012, http://hal.archives-ouvertes.fr/docs/00/79/44/37/PDF/Condat.fast_TV.pdf
- [25] J.M. Biucas-Dias, M.A.T. Figueiredo, An iterative algorithm for linear inverse problems with compound regularizers., in: *Proc. IEEE Int. Conf. Image Processing*, October, 2008, pp. 685–688.
- [26] S. Setzer, G. Steidl, T. Teuber, Infimal convolution regularizations with discrete ℓ_1 -type functionals, *Communications in Mathematical Sciences* 9 (3) (2011) 797–827.
- [27] P.E. McSharry, G.D. Clifford, L. Tarassenko, L.A. Smith, A dynamical model for generating synthetic electrocardiogram signals, *IEEE Transactions on Biomedical Engineering* 50 (3) (2003) 289–294.
- [28] 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. Peng, H.E. Stanley, Physiobank, physiotoolkit, and physionet, *Circulation* 101 (23) (2000) e215–e220.
- [29] M.A. Natalia, D. Zhi-De, P. Chi-Sang, Analysis of first-derivative based QRS detection algorithms, *IEEE Transactions on Biomedical Engineering* 55 (2) (2008) 478–484.
- [30] E.J. Candes, M.B. Wakin, S.P. Boyd, Enhancing sparsity by reweighted ℓ_1 Minimization, *Journal of Fourier Analysis and Applications* 14 (2008) 877–905.
- [31] D. Benitez, P.A. Gaydecki, A. Zaidi, A.P. Fitzpatrick, The use of the Hilbert transform in ECG signal analysis, *Computer Biology Medicine* 31 (September) (2001) 399–406.
- [32] P.S. Hamilton, W.J. Tompkins, Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database, *IEEE Transactions on Biomedical Engineering* 33 (December) (1986) 1157–1165.
- [33] P.M. Juan, A. Rute, O. Salvador, P.R. Ana, L. Pablo, A wavelet-based delineator: evaluation on standard databases, *IEEE Transactions on Biomedical Engineering* 51 (4) (2004) 570–581.
- [34] M. Bahoura, M. Hassani, M. Hubin, DSP implementation of wavelet transform for real time ECG wave forms detection and heart rate analysis, *Computer Methods and Programs in Biomedicine* 52 (1) (1997) 35–44.
- [35] L. Cuiwei, Z. Chongxun, T. Changfeng, Detection of ECG characteristic points using wavelet transforms, *IEEE Transactions on Biomedical Engineering* 42 (January) (1995) 21–28.
- [36] L. Jeongwhan, J. Keesam, Y. Jiyoung, L. Myoung, A simple real-time QRS detection algorithm, in: *Proc. IEEE. Eng. Med. Biol. Mag.*, vol. 4, October–3 November, 1996, pp. 1396–1398.