33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Baseline Wander Estimation and Removal by Quadratic Variation Reduction

A. Fasano, *Member, IEEE*, V. Villani, *Student Member, IEEE*, L. Vollero, *Member, IEEE*

Abstract— **The baseline wander is a low frequency additive noise partially overlapping the band of ECG signal. This makes its removal difficult without affecting the ECG. In this work we propose a novel approach to baseline wander estimation and removal based on the notion of *quadratic variation*. The quadratic variation is a suitable index of variability for vectors and sampled functions. We derive an algorithm for baseline estimation solving a constrained convex optimization problem. The computational complexity of the algorithm is *linear* in the size of the ECG record to detrend, making it suitable for real-time applications. Simulation results confirm the effectiveness of the approach and highlight its ability to remove baseline wander. Eventually, the proposed algorithm is not limited to ECG signals, but can be effectively applied whenever baseline estimation and removal are needed, such as EEG records.**

I. Introduction

The electrocardiogram (ECG) is a non invasive measure of the electrical activity of the heart recorded by skin electrodes. During ECG measurements, electrical changes on the patient's skin are detected, amplified and registered. The changes are caused by depolarization/polarization of the heart muscle during each heart beat. The analysis of ECG signals is commonly used as a diagnostic tool for detecting cardiac diseases. Unfortunately ECG signal is contaminated by several kinds of noise such as 50 or $60\,$Hz power-line interference, electromyographic noise and baseline wander [1]. This last is caused by fluctuations of the impedance between electrodes and skin, patient's movements and respiration.

The baseline wander is modeled as a low frequency additive noise over the range $0\div0.8\,$Hz, partially overlapping the band of ECG signal [2]. The in-band nature of this type of noise makes its removal difficult without affecting the ECG. The simplest approach is to use high-pass filtering with cutoff frequency of about $0.8\,$Hz [3]. However, this is not always viable, since it introduces distortions in the ECG signal, particularly in the ST segment, thus impacting on the diagnosis of some diseases like myocardial infarction and ischemia [4].

To overcome this problem, more sophisticated approaches have been proposed in the literature. These include adaptive filtering [5], wavelets [6] and baseline estimation [7], [8]. In [5] a two-stage adaptive filter is proposed. The first stage is an adaptive transversal high-pass filter that removes frequency components below $0.3\,$Hz. The second stage removes the frequency components higher than $0.3\,$Hz that are not correlated with QRS complexes. Wavelet approaches [6] decompose the signal into levels and adaptively filter only those levels directly affected by baseline drift. Other major approaches resort to piecewise cubic splines [7] or linear interpolation [8] between consecutive pre-known isoelectric levels estimated from PR intervals [9]. However, the performance of these methods strongly depends on the reliability of the selected fiducial points and decays in case of bradycardia, tachycardia, or rapid changes in the baseline, i.e., when fiducial points are difficult to find [10].

In this work we propose a novel approach to baseline wander estimation and removal, based on the concept of *quadratic variation* reduction. The rationale behind this approach is described in Section II. The baseline wander estimator is derived in Section III as the solution to a convex optimization problem. Section IV and V follow with simulation results and conclusions.

II. Rationale

As highlighted in the previous section, baseline wander noise is an additive low[1] "variability" component affecting the measured ECG. Thus, provided that we introduce a suitable index of "variability", baseline can be estimated searching for the low variability component closest, in some sense, to the measured ECG. Then, the estimated baseline can be subtracted from the measured ECG. In the following, we make this idea precise.

The variability of a generic vector can be quantified introducing the following

Definition 1: Given a vector $\boldsymbol{x} = [x_1 \cdots x_n]^{\mathrm{T}} \in \mathbb{R}^n$, the *quadratic variation* of $\boldsymbol{x}$ is defined as

$$[\boldsymbol{x}] \doteq \sum_{k=1}^{n-1} (x_k - x_{k+1})^2 \qquad (1)$$

and is denoted by $[\boldsymbol{x}]$.

The quadratic variation is a well-known property used in the analysis of stochastic processes [11]. However, in this context we consider it as a function of deterministic or random vectors.

Introducing the $(n-1) \times n$ matrix

$$\boldsymbol{D} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -1 \end{bmatrix}, \qquad (2)$$

[1] Low with respect to ECG "variability".

Authors are with the Faculty of Engineering, Università Campus Bio-Medico di Roma, via Álvaro del Portillo 21, 00128 Rome, Italy.
E-mail: {a.fasano,v.villani,l.vollero}@unicampus.it

978-1-4244-4122-8/11/$26.00 ©2011 IEEE

the quadratic variation of $\boldsymbol{x}$ can be expressed as

$$[\boldsymbol{x}] = \|\boldsymbol{D}\boldsymbol{x}\|^2 , \tag{3}$$

where $\|\cdot\|$ denotes the Euclidean norm.

The quadratic variation is a consistent index of variability and its use is motivated by the following property. For vectors affected by additive noise, on average it does not decrease and is an increasing function of noise variances. In fact, let $\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{w}$, where $\boldsymbol{x}_0$ is a deterministic vector and $\boldsymbol{w} = [w_1 \cdots w_n]^{\mathrm{T}}$ is a zero-mean random vector with covariance matrix $\boldsymbol{K}_w = \mathbb{E}\left\{\boldsymbol{w}\boldsymbol{w}^{\mathrm{T}}\right\}$. We do not make any assumption about the distribution of $\boldsymbol{w}$, so the following considerations hold regardless of the statistics of the noise. Computing the average quadratic variation of $\boldsymbol{x}$ we get

$$\begin{aligned}\mathbb{E}\left\{\|\boldsymbol{D}\boldsymbol{x}\|^2\right\} &= \|\boldsymbol{D}\boldsymbol{x}_0\|^2 + \mathbb{E}\left\{\mathrm{tr}\left(\boldsymbol{D}\boldsymbol{w}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{D}^{\mathrm{T}}\right)\right\} \\ &= \|\boldsymbol{D}\boldsymbol{x}_0\|^2 + \mathrm{tr}\left(\boldsymbol{D}\boldsymbol{K}_w\boldsymbol{D}^{\mathrm{T}}\right)\end{aligned} \tag{4}$$

where, in the first equality, we have exploited the invariance of the trace under cyclic permutations. Note that $\mathrm{tr}\left(\boldsymbol{D}\boldsymbol{K}_w\boldsymbol{D}^{\mathrm{T}}\right) \geq 0$, since it is the trace of a positive semidefinite matrix [12], but in all practical cases the inequality is strict. In fact, we have

$$\begin{aligned}\mathrm{tr}\left(\boldsymbol{D}\boldsymbol{K}_w\boldsymbol{D}^{\mathrm{T}}\right) &= \sum_{k=1}^{n-1} \mathbb{E}\left\{(w_k - w_{k+1})^2\right\} \\ &= \sum_{k=1}^{n-1} \left(\sigma_k^2 + \sigma_{k+1}^2 - 2\sigma_{k,k+1}\right)\end{aligned} \tag{5}$$

where $\sigma_k^2 = \mathbb{E}\left\{w_k^2\right\}$ and $\sigma_{k,k+1} = \mathbb{E}\left\{w_k w_{k+1}\right\}$. From (5) follows that $\mathrm{tr}\left(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{K}_w\right) = 0$ if and only if all the components of the noise vector $\boldsymbol{w}$ are almost surely equal[2] and that $\mathbb{E}\left\{\|\boldsymbol{D}\boldsymbol{x}\|^2\right\}$ is an increasing function of noise variances.

The next section is devoted to the development of an efficient algorithm for baseline removal exploiting the concept of quadratic variation.

III. Baseline estimation and removal

In this section, we denote by $\widetilde{\boldsymbol{z}}$ the vector collecting n samples of a measured ECG record, i.e., one that is affected by baseline wander, by $\boldsymbol{x}$ the vector of estimated baseline, and by $\boldsymbol{z} = \widetilde{\boldsymbol{z}} - \boldsymbol{x}$ the corresponding ECG vector after baseline removal. Following the line of reasoning presented in the previous section, we propose to estimate the baseline $\boldsymbol{x}$ solving the following optimization problem

$$\begin{cases} \text{minimize} & \|\boldsymbol{x} - \widetilde{\boldsymbol{z}}\|^2 \\ \text{subject to} & \|\boldsymbol{D}\boldsymbol{x}\|^2 \leq \rho \end{cases} \tag{6}$$

where $\boldsymbol{D}$ is defined in (2) and ρ is a nonnegative constant that controls the quadratic variation of the estimated baseline. Its value is chosen in accordance with the peculiarity of the problem and satisfies $\rho < \|\boldsymbol{D}\widetilde{\boldsymbol{z}}\|^2$ in order to avoid trivial solutions.[3] Note that we do not need to know in advance the appropriate value for ρ in any particular problem. In fact, as it will be clear later, the solution to the optimization problem (6) can be expressed in terms of a parameter that controls the quadratic variation of the solution and that is related to the value of ρ in (6). In this way, baseline can be estimated without caring about ρ in the constraint $\|\boldsymbol{D}\boldsymbol{x}\|^2 \leq \rho$, and reducing parametrically the quadratic variation of the solution $\boldsymbol{x}$ to the desired level.

Let us consider (6) in more detail. It is a convex optimization problem, since both the objective function and the inequality constraint are convex. As a consequence, any locally optimal point is also globally optimal [13]. Moreover, since the objective function is strictly convex and the problem is feasible, the solution exists and is unique. It is possible to prove that the solution to (6) is given by

$$\boldsymbol{x} = \left(\boldsymbol{I} + \lambda\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D}\right)^{-1}\widetilde{\boldsymbol{z}} \tag{7}$$

where $\boldsymbol{I}$ denotes the identity matrix, and λ is a nonnegative parameter determined by

$$\|\boldsymbol{D}\boldsymbol{x}\|^2 = \left\|\boldsymbol{D}\left(\boldsymbol{I} + \lambda\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D}\right)^{-1}\widetilde{\boldsymbol{z}}\right\|^2 = \rho. \tag{8}$$

Note that in (7) the inverse exists for any $\lambda \geq 0$. It is interesting that the solution to (6) is a linear operator acting on $\widetilde{\boldsymbol{z}}$. Moreover, the parameter λ controls the quadratic variation of the solution $\boldsymbol{x}$, i.e., the degree of variability of the estimated baseline. In fact, it is possible to prove that $[\boldsymbol{x}]$ is a *continuous and strictly decreasing* function of λ for $\lambda \in [0, +\infty)$ regardless of $\widetilde{\boldsymbol{z}}$, provided that $\widetilde{\boldsymbol{z}}$ is not a constant vector.[4] This is equivalent to say that (8), when $\widetilde{\boldsymbol{z}}$ is not a constant vector, establishes a one-to-one correspondence between $\lambda \in [0, +\infty)$ and $\rho \in \left(0, \|\boldsymbol{D}\widetilde{\boldsymbol{z}}\|^2\right]$, with $\lambda = 0$ corresponding to $\rho = \|\boldsymbol{D}\widetilde{\boldsymbol{z}}\|^2$ and

$$\lim_{\lambda\to+\infty} \rho = \lim_{\lambda\to+\infty} \left\|\boldsymbol{D}\left(\boldsymbol{I} + \lambda\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D}\right)^{-1}\widetilde{\boldsymbol{z}}\right\|^2 = 0 , \tag{9}$$

which holds true regardless of $\widetilde{\boldsymbol{z}} \in \mathbb{R}^n$.

A consequence of this is that we *do not need to know* in advance the value of ρ in (6), since baseline can be estimated according to (7) and λ can be adapted to the particular problem or to fulfill some performance criterion. That is, λ is used in place of ρ as the controlling parameter.

In particular, since baseline is characterized by low values of the quadratic variation with respect to the ECG signal, it is estimated with (7) using *large values* of λ. In our simulations values of the order of 10^4 or even more are quite common.

Based on this fact, it is worthwhile considering the behavior of the solution (7) in the limit for $\lambda \to +\infty$. It is possible to prove the following result

$$\lim_{\lambda\to+\infty} \left(\boldsymbol{I} + \lambda\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D}\right)^{-1}\widetilde{\boldsymbol{z}} = \frac{1}{n}\boldsymbol{1}^{\mathrm{T}}\widetilde{\boldsymbol{z}} = \frac{1}{n}\sum_{k=1}^{n}\widetilde{z}_k , \tag{10}$$

[2] That is $w_1 = w_2 = \cdots = w_n$ with probability 1.

[3] When $\rho \geq \|\boldsymbol{D}\widetilde{\boldsymbol{z}}\|^2$ the solution is $\boldsymbol{x} = \widetilde{\boldsymbol{z}}$ and baseline coincides with the measured ECG.

[4] If $\widetilde{\boldsymbol{z}}$ is a constant vector $[\boldsymbol{x}] = [\widetilde{\boldsymbol{z}}] = 0$ regardless of λ. However, this is impossible for a vector representing an ECG record.

where $\mathbf{1}^{\mathrm{T}} = [1 \cdots 1]$ is the $n-$size constant unit vector and $\widetilde{z}_k$ is the $k-$th component of $\widetilde{\boldsymbol{z}}$.

The asymptotic solution in (10) corresponds to the mean value of $\widetilde{\boldsymbol{z}}$, which is the constant vector closest in the l_2 norm to the measured ECG. Consistently with (9) such a constant vector has zero quadratic variation.

As λ ranges from 0 to $+\infty$ solution (7) captures components of the measured ECG $\widetilde{\boldsymbol{z}}$ with decreasing quadratic variation. When $\lambda \to +\infty$ such components reduce to the constant in (10), whereas when λ is finite more complex trends of the measured ECG are captured.

Once the baseline has been estimated, it can be removed from the measured ECG by subtraction

$$\boldsymbol{z} = \widetilde{\boldsymbol{z}} - \boldsymbol{x} = \left[\boldsymbol{I} - \left(\boldsymbol{I} + \lambda \boldsymbol{D}^{\mathrm{T}}\boldsymbol{D}\right)^{-1}\right]\widetilde{\boldsymbol{z}} \tag{11}$$

$$= \boldsymbol{D}^{\mathrm{T}}\left(\frac{1}{\lambda}\boldsymbol{I} + \boldsymbol{D}\boldsymbol{D}^{\mathrm{T}}\right)^{-1}\boldsymbol{D}\widetilde{\boldsymbol{z}}\,, \tag{12}$$

where in the last equality the Sherman-Morrison-Woodbury formula [12] has been applied.

It is important to consider the computational aspects related to baseline estimation through the proposed algorithm, since matrix inversion is involved in (7). When the size of vector $\widetilde{\boldsymbol{z}}$ is large, as it is the case for typical ECG records, the computational burden, both in terms of time and memory, and the accuracy become serious issues, even for batch processing.

It is possible to prove that baseline estimation using (7) can be performed with complexity $O(n)$, i.e., *linear in the size of vector* $\widetilde{\boldsymbol{z}}$. This property is very important and makes the proposed algorithm suitable also for real-time applications. Just to give an idea of how fast the algorithm is, a MATLAB (ver. 7.11) implementation of (7) running over a PC equipped with 2.5 GHz Core 2 Duo processor, takes about 0.84 s to estimate the baseline from an ECG record of 10^7 double precision floating point samples.

Eventually, it is worthwhile noting that the algorithm we propose is not limited to ECG, but can be applied in very general situations, whenever baseline estimation and removal are needed. This is due to the fact that the formulation and the rationale behind it, i.e., quadratic variation reduction, have general validity. In this regard, we successfully applied it also to EEG recordings.

IV. Simulation results

The performance of the proposed algorithm has been investigated both on real and simulated ECG traces.

As for real signals, we considered ECG traces from the MIT-BIH Normal Sinus Rhythm Database [14] freely available on Physionet. This database includes 18 long-term ECG recordings of subjects with no significant arrhythmias. Signals were acquired at a sampling frequency of 128 Hz with 12−bit resolution. Figure 1 reports a 40 s segment of the record nsrdb/16272 and Figure 2 shows the same record after baseline wander removal using the proposed algorithm with $\lambda = 10^4$. A visual and qualitative comparison of the two figures highlights how quadratic variation reduction technique managed to correctly detrend the ECG signal. Undesired wandering due to baseline has been removed without any visible distortion of the original signal.

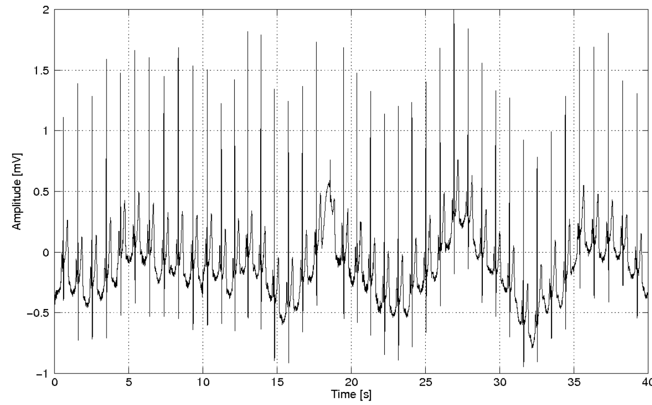

Fig. 1. ECG record from real data.

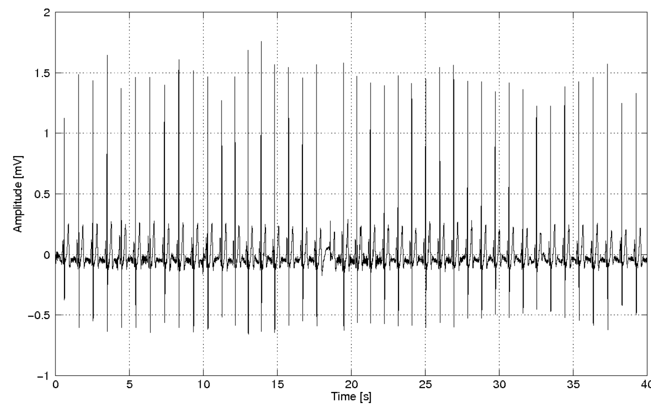

Fig. 2. ECG record of Figure 1 after baseline wander removal with the proposed algorithm.

In order to quantify the performance of the proposed method we carried out a quantitative analysis on simulated ECG signals affected by synthetic baseline wander. Moreover we compared the performance of our algorithm versus the baseline removal algorithm based on cubic spline interpolation described in [7]. As a performance metric we considered the squared Euclidean distance

$$d^2(\boldsymbol{x}, \boldsymbol{\xi}) = \|\boldsymbol{x} - \boldsymbol{\xi}\|^2 \tag{13}$$

between the synthetic baseline wander, denoted by $\boldsymbol{\xi}$, and the corresponding estimated one $\boldsymbol{x}$.

The synthetic baseline-free ECG signal, denoted in the following by $\boldsymbol{z}_0$, was generated according to the model described in [15], setting the sampling frequency to 1024 Hz, the heart rate to 60 bpm and including zero mean additive Gaussian noise with standard deviation $\sigma = 0.03$. Synthetic baseline wander $\boldsymbol{\xi}$ was rendered as Gaussian white noise with standard deviation $\sigma = 20$ low-pass filtered with bandwidth 0.8 Hz in order to include drift due to respiration and motion artifacts.

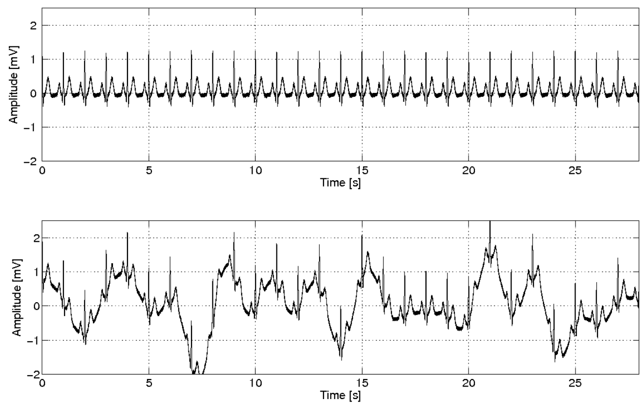

Fig. 3. Synthetic ECG signal (upper panel) with added known baseline wander (lower panel).

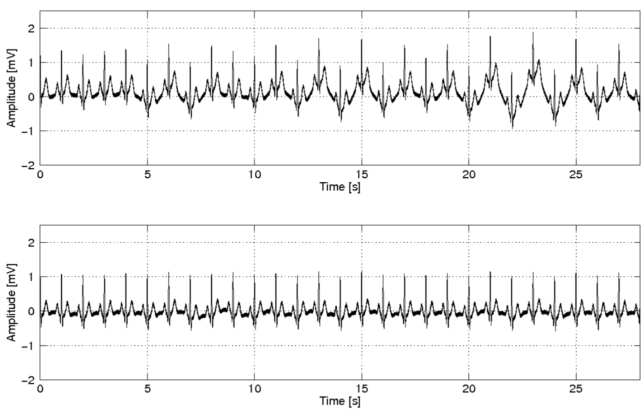

Fig. 4. ECG in lower panel of Figure 3, after baseline wander removal with cubic splines (upper panel) and with the proposed algorithm (lower panel).

Figure 3 reports the synthetic baseline-free ECG signal $\boldsymbol{z}_0$ (upper panel) and the corresponding corrupted one $\widetilde{\boldsymbol{z}} = \boldsymbol{z}_0 + \boldsymbol{\xi}$ (lower panel) with synthetic baseline added. The upper panel of Figure 4 shows the detrended ECG record $\boldsymbol{z}_{cs} = \widetilde{\boldsymbol{z}} - \boldsymbol{x}_{cs}$ after baseline estimation[5] with cubic spline interpolation following [7]. The lower panel of Figure 4 reports the detrended ECG record $\boldsymbol{z}_{qv} = \widetilde{\boldsymbol{z}} - \boldsymbol{x}_{qv}$ after baseline estimation according to the proposed algorithm. Here $\lambda = 1.3 \times 10^4$, which corresponds to the value that entails the minimum $d^2(\boldsymbol{x}_{qv}, \boldsymbol{\xi})$.

As can be seen in Figure 4, the quadratic variation reduction method is more effective in baseline wander removal. In the upper panel of Figure 4, indeed, a residual baseline drift is still present, whereas in the lower panel, where the proposed algorithm has been used, the isoelectric levels are better aligned. This is confirmed by the values assumed by the figure of merit (13) in this case: the cubic spline interpolation method gives $d^2(\boldsymbol{x}_{cs}, \boldsymbol{\xi}) = 40.51$, whereas the proposed algorithm returns a significantly lower $d^2(\boldsymbol{x}_{qv}, \boldsymbol{\xi}) = 18.66$.

[5] We denote by $\boldsymbol{x}_{cs}$ the baseline estimated with cubic spline interpolation and by $\boldsymbol{x}_{qv}$ the one estimated with our algorithm.

V. Conclusions

In this work we considered the problem of baseline wander estimation and removal, in particular for ECG signals. We proposed a novel approach based on the notion of *quadratic variation* reduction. We derived the algorithm for baseline wander estimation solving a constrained convex optimization problem. The algorithm is remarkably fast, since its computational complexity is *linear* in the size of the ECG record to detrend. This makes it perfectly suitable for real-time applications. Simulation results, both on real and simulated data, confirm the effectiveness of the approach and highlight its ability to remove baseline wander.

Eventually, it is worthwhile noting that the proposed algorithm is not limited to ECG signals, but can be effectively applied whenever baseline estimation and removal are needed. In this regard, we successfully applied it also to EEG traces.

References

[1] G. D. Clifford, F. Azuaje, and P. McSharry, *Advanced Methods and Tools for ECG Data Analysis*. Norwood, MA, USA: Artech House, Inc., 2006.

[2] L. T. Sheffield, C. A. Berson, D. B. Remschel, P. C. Gillette, R. E. Hermes, L. Hinkle, H. Kennedy, D. M. Mirvis, and C. Oliver, "Recommendations for standard of instrumentation and practice in the use of ambulatory electrocardiography," *Circulation*, pp. 626A–636A, 1985.

[3] J. A. Van Alste and T. S. Schilder, "Removal of base-line wander and power-line interference from the ECG by an efficient FIR filter with a reduced number of taps," *IEEE Transactions on Biomedical Engineering*, vol. 32, no. 12, pp. 1052–1060, Dec 1985.

[4] A. R. Houghton and D. Gary, *Making Sense of the ECG: A Hands-on Guide*. Arnold Publishing Co., 2003.

[5] P. Laguna, R. Jane, and P. Caminal, "Adaptive filtering of ECG baseline wander," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC88)*, vol. 2, pp. 508–509, 1988.

[6] Z. Donghui, "Wavelet approach for ECG baseline wander correction and noise reduction," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC05)*, pp. 1212–1215, 2005.

[7] C. R. Meyer and H. N. Keiser, "Electrocardiogram baseline noise estimation and removal using cubic splines and state-space computation techniques," *Computers and Biomedical Research*, vol. 10, no. 5, pp. 459–470, Oct 1977.

[8] V. S. Chouhan and S. S. Mehta, "Total removal of baseline drift from ECG signal," *International Conference on Computing: Theory and Applications (ICCTA07)*, 2007.

[9] P. W. MacFarlane, J. Peden, J. Lennox, M. P. Watts, and T. D. Lawrie, "The Glasgow system," *Trends in Computer-Processed Electrocardiograms: Proceedings of the IFIP Working Conference on Trends in Computer-Processed Electrocardiograms*, pp. 143–150, 1977.

[10] L. Sörnmo and P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Academic Press, June 2005.

[11] S. E. Shreve, *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer Science+Business Media, Inc, 2004.

[12] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, February 1990.

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, March 2004.

[14] The MIT-BIH Normal Sinus Rhythm Database. [Online]. Available: http://www.physionet.org/physiobank/ database/nsrdb/

[15] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 3, pp. 289–294, 2003.