ECG signal compression and noise filtering with truncated orthogonal expansions

Running head: Truncated expansions of noisy ECG

Keywords. ECG compression, truncated orthogonal expansions, noise filtering.

April 19, 1999

by

Salvador Olmos (#), José García (#), Raimon Jané (&), Pablo Laguna (#)

(#) Departamento de Ingeniería Electrónica y Comunicaciones.

Centro Politécnico Superior

Universidad de Zaragoza (Spain)

(&) Centre de Recerca d'Enginyeria Biomèdica

Dep. Eng. Sistemes, Automàtica i Informàtica Industrial

Universitat Politècnica de Catalunya (Spain)

Number of manuscript pages: XX

Number of tables:

Number of figures: 18

Address for correspondence

(#) Salvador Olmos

Dep. Ingeniería Electrónica y Comunicaciones

María de Luna, 3

50015-ZARAGOZA (Spain)

Phone: +34-976-761932

FAX: +34-976-762111

e-mail: olmos@posta.unizar.es

ACKNOWLEDGMENT

This work was supported in part by TIC97-0945-CO2-01:02 from CICYT and P40/98 from CONSID-DGA (Spain).

Abstract

With the increasing use of the electrocardiographic signal (ECG) as a diagnostic tool in cardiology, there exists a requirement for effective ECG compression techniques. The goal of any data compression system is to maximize compression while minimizing distortion. Orthogonal expansions is a tool widely used because its compression capacity in recurrent signals. In this paper we analyze the effect of noise in orthogonal expansions of ECG signals. When the observed signal is embedded in additive noise, distortion measurements, such as the mean square error, are not a monotonic decreasing function of the number of transform coefficients, due to the noise presence. We analyze and compare two different ways to estimate the transform coefficients: inner product and adaptive estimation with the LMS algorithm. For stationary signals, we demonstrate and quantify the superior performance obtained by the adaptive system when low values of the step-size are used $\mu < \mu_{lim}$. For non-stationary signals, we propose, based on experimental results, values of the LMS step-size μ depending on the noise characteristics and the signal to noise ratio. Theoretical results are contrasted with a simulation study with actual ECG signals from MIT-BIH Arrythmia database and three kinds of noise: simulated Gaussian white noise, and two records of physiological noise that essentially contains electrode motion artifacts and muscular activity.

1 Introduction

The great amount of data obtained when recording ECG signals leads to the need of data compression techniques for storing, transmitting and analyzing the data, without loss of clinical information. For example, a typical Holter recording (two leads, 24 hours long, 500 Hz of sampling rate and 12 bits of precision) needs more than 123 MBytes of memory for storing the data. Therefore, data compression systems will be very useful if they can reduce this volume of data removing redundancies from the signal.

Jalaleddine presented in 1990 [1] a very good review of data compression techniques for ECG signals and he classified them into three major groups: a) direct methods, b) transformation using orthogonal functions and c) parameter extraction. The most used techniques for ECG signals concern with the two first ones, because they are reversible processes that permit a subsequent reconstruction of the signal for later analysis. Many algorithms based on direct methods were proposed in the first years of ECG data compression [2–6] because they are quite simple and can be implemented easily on real time systems. With the increasing calculation power of computers many algorithms for ECG data compression using orthogonal transforms have been designed during the last decade [7–10] showing its superior performance respect to direct methods.

In this work we analyze the effect of additive noise on orthogonal transform based compression of ECG signals. An increase of the number of basis functions in the orthogonal transform representation reduces distortion, or equivalently, improves the signal quality. When the observed signal is corrupted by noise, not only the reconstructed signal energy increase for larger values of the number of basis functions but also the noise energy increase [11]. If all the signal-space basis functions are used by the transform coder, the distortion, evaluated as the difference between the reconstructed signal and the original clean signal, will be equal to the noise energy. In consequence, there will be an optimum number of coefficients (and then of basis functions) that will minimize the distortion. Quadratic error indexes [12], such as the mean square error (MSE), have become standard in order to quantify distortion. When we compress noisy signals we want to extract only the information from the original clean signal. If we use the MSE index between the reconstruction and the observed noisy signal we can get high values of MSE due to the presence of the unwanted noise, that do not represent the actual distortion between the original clean signal and its reconstruction. In this work we propose a simulation study where we generate noisy signals from actual ECG records from MIT-BIH Arrythmia database [13]. Three different noise sources are considered: simulated Gaussian white noise and two records of physiological noise that essentially contains electrode motion artifacts and muscular activity. The MSE index will be measured between the reconstruction and the original clean signal.

In section 2 it is presented a brief review of orthogonal transform compression systems. In next sections we describe and compare the performance of two classical ways to estimate the transform coefficients: the inner product and an adaptive estimation with the LMS algorithm. Expressions for the MSE for clean and noisy ECG signals are derived for both estimation methods. The LMS algorithm steady-state performance for stationary signals is analyzed and compared with the classical inner product in section 5. For the selection of the LMS step-size μ we give an expression of the limit value μ_{lim} that gets same steady-state performance than inner product. The operation of the LMS algorithm in a non-stationary environment is also analyzed in section 6 with a new criteria for the optimum step-size μ selection. Finally, theoretical results are contrasted in a simulation study with actual ECG signals from MIT-BIH Arrythmia database.

2 Truncated orthogonal expansions of ECG signals

In order to apply an orthogonal transform to the ECG signal it is necessary to make some preprocessing steps in order to segment the signal in vectors. Each heartbeat is treated as a separated vector. This requires that the position of each QRS complex be determined prior to the compression phase. QRS complexes were detected and labeled using ARISTOTLE software [14]. All beat vectors are aligned with respect to the fiducial point of the QRS complex, and the beginning of the vector is established with different criteria for different detected morphologies. For normal beats, the beginning of the window was defined 250 ms prior to the QRS fiducial point. All signal vectors are zero-padded to the maximum heartbeat length (N samples). Figure 1 shows one example of the ECG signal segmentation. More details of the segmentation for other morphologies can be found in [15].

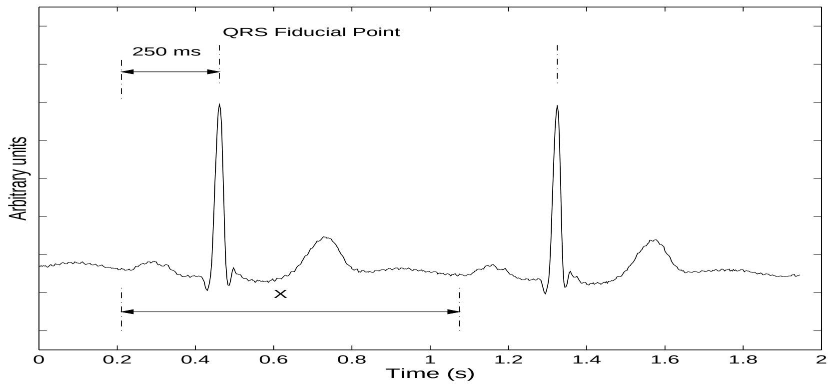


Figure 1: *Beat segmentation for ECG data compression.*

The operation of an orthogonal transform data compression system is illustrated in Fig. 2. The ECG signal vector $\mathbf{X} = [x_0, x_1 \ldots, x_{N-1}]^T$ is operated on by the orthogonal transform $\mathbf{T}$, to produce the transform vector $\mathbf{C} = [c_0, c_1 \ldots, c_{N-1}]^T$. The elements of $\mathbf{C}$ are the magnitude of the projections of $\mathbf{X}$ vector onto the basis formed by the rows of $\mathbf{T}$. The purpose of the transformation is to convert the data vector $\mathbf{X}$ into a transform coefficient vector which can be optimally quantized. Typically the components of $\mathbf{X}$ are correlated and the transformation $\mathbf{T}$ tries to decorrelate the signal samples and also to pack the signal energy in a few transform coefficients. The reconstructed signal $\mathbf{X_R}$ can be obtained with the inverse transform $\mathbf{T}^{-1}$ applied to the quantized p coefficients. In this work we will not consider the quantization step $\mathbf{Q}$, and we will only analyze the effect of noise on truncated orthogonal expansions of ECG signals. The mean square error between the rank p approximation $\mathbf{X_R}$ and $\mathbf{X}$ is the energy represented in the discarded coefficients

$$MSE_p = \frac{1}{N} \sum_{i=p}^{N-1} c_i^2 \,. \tag{1}$$

In order to outperform the compression ratio the transform coefficient series corresponding to the subsequent heartbeats can be differentially quantized using some algorithms like DPCM or LPC [16].

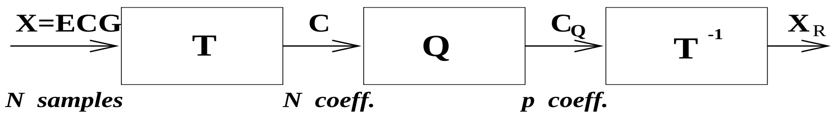


Figure 2: *Block diagram of an orthogonal transform data compression system.*

To reduce the transients in the series of coefficients originated by ECG morphology changes we independently apply the transform coder to each heartbeat series with the same morphology. The heartbeat morphologies (normal beats, ventricular beats and so on) are labeled with the software ARISTOTLE [14].

Data compression systems have a general trade-off between distortion and compression ratio. When transform coding schemes are applied it is very important to select the orthogonal transform $\mathbf{T}$ that can represent the maximum amount of signal energy with the minimum number of coefficients. Several orthogonal transforms $\mathbf{T}$ have been applied to ECG signals: Discrete Cosine Transform (DCT) [17], Legendre Transform (LT) [18], Hermite Transform (HT) [19], Karhunen-Loève Transform (KLT) [10, 20]. In this work we have selected the KLT, but equivalent results, in terms of noise behavior, can be obtained for the other orthogonal transforms. The KLT minimizes the cost function

$$\xi = E\{|e[k]|^2\} = E\{|x - \hat{x}[k]|^2\} \tag{2}$$

where $\hat{x}[k]$ is the approximation of $x[k]$ with a linear combination of p orthogonal functions. The KLT is optimal in the sense that it needs the minimum number of coefficients for a given MSE [21, 22]. The KLT is a signal dependent transform and its basis functions are calculated as the eigenvectors of the covariance matrix of a training set of signals. In the KL domain the transform coefficients are uncorrelated (the covariance matrix is diagonal), thus redundancies are removed. The eigenvalues of the covariance matrix are the expected values of the squared KL coefficients at the training set, giving a measure of the importance of each function in the linear combination for representing the signals at the training set. Sorting the basis functions (eigenvectors) in decreasing order of eigenvalues and selecting the more significant ones we can get a good representation of the signal with a reduced number of coefficients. In consequence, the KLT defines the domain where the signal energy is more concentrated.

The KL basis functions for the ECG signal were calculated from a large training set with over 110.000 beats from MIT and European ST-T databases, and some records from people non cardiac diagnosed. If we consider independent signal vector series from individual beat morphologies in order to estimate the covariance matrix and calculate the basis functions of the KLT, we get a better performance than only considering one basis for all kinds of morphologies. This result is because we need lower number of functions for representing an homogeneous pattern space than a non-homogeneous one. The morphologies of the first KL basis functions for normal beats are quite similar to the most frequent morphologies of normal beats (see Fig. 3). The sampling rate was 360 Hz and the maximum heartbeat length was 1194 ms (N=430 samples).

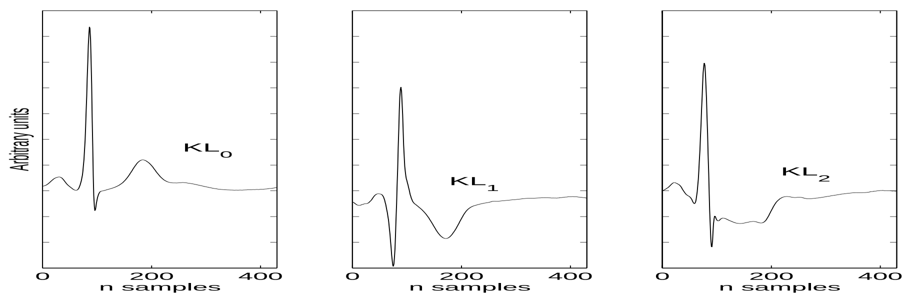


Figure 3: *The first three KL basis functions for normal beats.*

We have recently demonstrated that truncated orthogonal expansions of recurrent signals, like ECG, are equivalent to apply a linear time-variant periodic filter to the input signal [23]. The time variant transfer function of the system can be easily calculated from the basis functions used in the expansion. In next sections we analyze the performance of two classical methods to estimate the transform coefficients: inner product and adaptive estimation with the LMS algorithm.

3 Inner product estimation

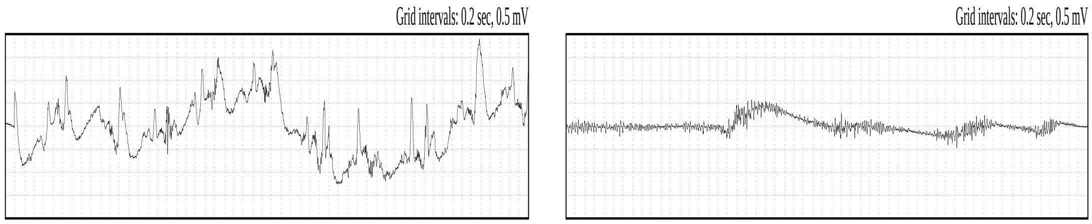
If noise $\mathbf{N} = [n_0, n_1, \ldots, n_{N-1}]^T$ is added to the input ECG signal $\mathbf{X}$, the MSE between the original clean signal $\mathbf{X}$ and the reconstructed signal $\mathbf{X_R}$ will have now two components [11]:

$$MSE_p^{direct} = \frac{1}{N} \sum_{i=p}^{N-1} c_i^2 + \frac{1}{N} \sum_{i=0}^{p-1} \alpha_i^2 \, , \tag{3}$$

where α_i are the coefficients of the noise in the transformed domain. The first component is the truncation signal error (the same than in equation (1)), and the second one is due to the noise represented in the approximation. If we are interested in the reconstruction with p basis functions we obtain a contaminating noise energy

$$NOISE_p^{direct} = \frac{1}{N} \sum_{i=0}^{p-1} \alpha_i^2 \, . \tag{4}$$

Three different kinds of noise have been considered in this work: *simulated Gaussian white noise*, *muscular noise* and *motion artifact noise*. White noise is artificially generated, but more realistic sources of noise present in ECG signals are also considered as electrode motion and muscular noise. Noise data was obtained from two records of MIT-BIH database (Noise Stress Test database) [13]. These records were obtained using a Holter recorder on an active subject, with leads placed so that the subject's ECG is not visible. Record '*em*' contains electrode motion artifact (usually the result of intermittent mechanical forces acting on the electrodes), with significant amounts of baseline wander and muscle noise as well. Record '*ma*' contains primarily muscle noise (EMG), with a spectrum that overlaps that of the ECG, but extends to higher frequencies. In Fig. 4 there is an excerpt of these two noise records.



(a) Record em (electrode motion) *(b) Record ma (muscular activity)*

Figure 4: *Excerpts of noise records.*

In order to represent the noise energy distribution in the KL domain, we averaged the square KL coefficients representation from 500 contiguous windows (1194 ms long) of noise (equivalent to 597 seconds). The noise windows were not aligned with any criteria, because we considered that noise was uncorrelated with the ECG signal.

The MSE_p^{direct} component due to the presence of noise in the KL domain is represented in Fig. 5 (values of MSE are normalized to the noise energy). *White noise* presents a linear behavior as it was expected since its contribution is equally distributed at any domain. In contrast, the energy of *em* and *ma* noise is more concentrated in the first KL functions, so their representations in the KL-domain will be more overlapped with the ECG signal than white noise.

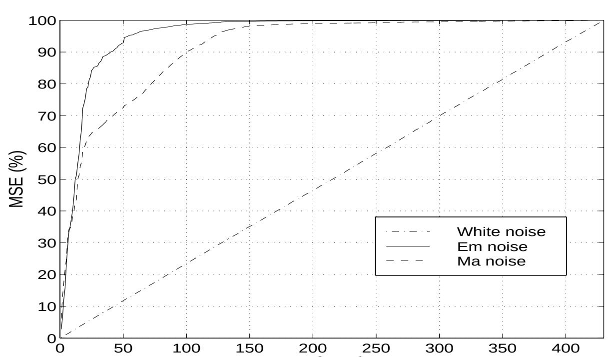
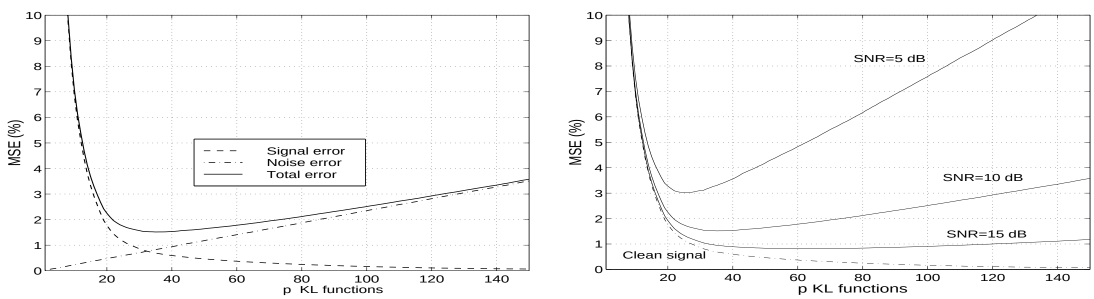


Figure 5: MSE_p^{direct} *noise component in the KL domain for three kinds of noise.*

In Fig. 6(a) it is shown the total MSE_p^{direct} and its two components (signal error and noise error) when Gaussian white noise is added to the original signal with a signal to noise ratio of SNR=10 dB. The SNR was measured as the classical ratio between signal and noise power. The values of MSE_p^{direct} were obtained using the equation (3), taking the eigenvalues of the covariance matrix which represent the ECG signal at the training set (they are the squared expected values of the coefficients) and the noise representation in Fig. 5. Due to the presence of white noise in the input signal, reconstructions with p=35 KL functions have lower values of MSE than those ones with p=80 functions. Thus the number of coefficients must be selected carefully and accordingly to the amount of noise present in the signal. In Fig. 6(b) we represent the values of the total MSE_p^{direct} for white noise with several values of SNR. It can be seen that there is an optimum value of $p\,(p^*)$ that minimizes the MSE_p^{direct} of noisy signals. The optimum value depends on the transformed representation of both signal and noise, and the SNR of the input signal. It is clearly seen that the optimum number of coefficients (p^*) decreases as the SNR also decreases. Moreover, the selection of the optimum number of coefficients has different sensitivity because of the error curve slope.



(a) The two components and total MSE_p^{direct} with SNR=10 dB.

(b) MSE_p^{direct} for noisy signals with several values of SNR of input signal.

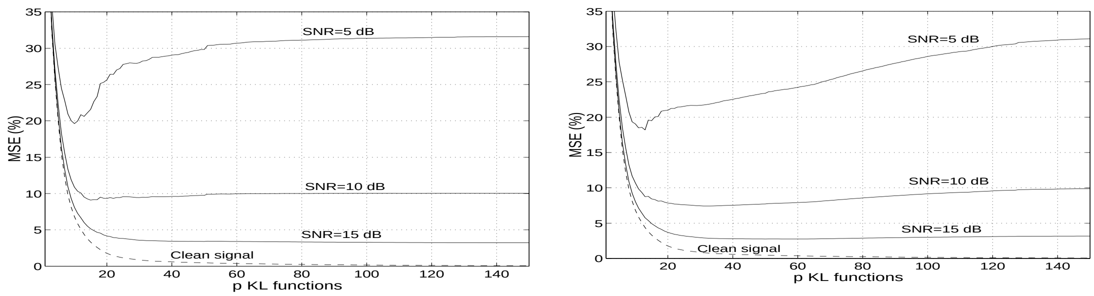
Figure 6: *MSE_p^{direct} for ECG signals contaminated with simulated Gaussian white noise.*

The SNR of the reconstructed signal will be

$$SNR_p^{direct} = \frac{\sum_{i=0}^{p-1} c_i^2}{\sum_{i=0}^{p-1} \alpha_i^2} = \frac{\sum_{i=0}^{p-1} c_i^2}{p\,\sigma^2} \tag{5}$$

where the last equality holds for the case of white noise with variance $\alpha_i^2 = \sigma^2$.

Similar results of MSE_p^{direct} can be also obtained for physiological colored noise. The MSE_p^{direct} values for physiological noise are higher than for white noise because now signal and noise representations are more overlapped in the KL domain (see Fig 7). Also an optimum p^* value can be obtained that minimizes MSE_p^{direct}.



(a) MSE_p^{direct} *for 'em'-noisy signals.* *(b)* MSE_p^{direct} *for 'ma'-noisy signals.*

Figure 7: MSE_p^{direct} *for ECG signals contaminated with 'em' and 'ma' noise with several values of SNR.*

4 Adaptive coefficient estimation with the LMS algorithm

Adaptive estimation of quasi-periodic signals, such as ECG, is a wide spread technique for estimating signals embedded in uncorrelated additive noise [24, 25]. This technique has been applied to the analysis of ECG signals [26, 27] and evoked potentials [28]. It makes use of the recurrent behavior of the signal and it is based on the adaptive linear combiner (ALC) [25]. Figure 8 shows this process in schematic form. The adaptive filter input signal (the *primary input*, $d[k]$) consists of subsequent concatenated noisy observed heartbeats. Short beats are lengthened by appending zeroes as necessary, so that a new beat begins every N samples. The adaptive system dynamically estimates the amount of each reference input $\widetilde{\Phi}_i[k]$ present in the input signal. The reference inputs $\{\widetilde{\Phi}_i[k]\ (i = 0, \ldots, p-1 \leq N-1)\}$ are the periodic extension of the basis functions used to represent the ECG signal. In [26] the reference inputs were the orthonormal Hermite functions, in [27, 29] unit impulses, and in [28, 30] sine, cosine and Walsh functions. In the present study, the reference inputs are the KL basis functions of the ECG signal. The output of the adaptive filter, $y[k]$, is the signal that we want to be an estimate of $s[k]$, and $e[k]$ is the error signal $e[k] = s[k] + n[k] - y[k]$ with

$$y[k] = \sum_{i=0}^{p-1} W_i[k]\, \widetilde{\Phi}_i[k]\,. \tag{6}$$

When any adaptive algorithm is used to minimize the mean square error $\xi = E\{|e[k]|^2\}$ and the input signal is stationary, the weight vector $\mathbf{W}$ converges to the Wiener optimal solution $\mathbf{W}^* = \mathbf{R}^{-1}\mathbf{P}$ [25], where

$$\mathbf{R} = E\{\widetilde{\mathbf{\Phi}}[k]\widetilde{\mathbf{\Phi}}^T[k]\} \qquad \text{and} \qquad \mathbf{P} = E\{d[k]\,\widetilde{\mathbf{\Phi}}[k]\} \tag{7}$$

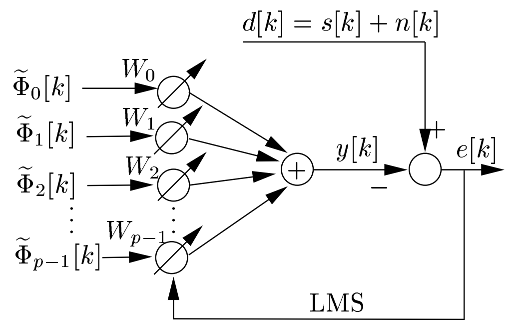


Figure 8: *Adaptive linear combiner for estimating the KL coefficients.*

and $\widetilde{\mathbf{\Phi}}[k]$ denotes the vector of reference signals at instant k

$$\widetilde{\mathbf{\Phi}}[k] = [\widetilde{\Phi}_0[k], \widetilde{\Phi}_1[k], \ldots, \widetilde{\Phi}_{p-1}[k]]^T \,. \tag{8}$$

In this case, given the orthonormality conditions of the basis functions and the lack of correlation assumed between the noise $n[k]$ and the basis functions $\widetilde{\Phi}_i[k]$, the mean value over a signal occurrence of $\mathbf{R}$ and $\mathbf{P}$ reduce[1] to

$$\mathbf{R} = \frac{1}{N}\mathbf{I} \qquad \text{and} \qquad \mathbf{P} = \frac{1}{N}[c_0, c_1, \ldots, c_{N-1}]^T \,, \tag{9}$$

being c_i the transform coefficients of the $s[k]$ signal. The optimal weight vector, $\mathbf{W}^*$, that minimizes the mean square error is $\mathbf{W}^* = [c_0, c_1, \ldots, c_{N-1}]^T$. This result means that the steady-state value of each weight W_i^* is an estimation of the i-th transform coefficient of $s[k]$. Thus the steady-state weight vector is a characterization of the deterministic signal component in the transformed domain, and the output signal $y[k]$, in the optimum case, takes the value

$$y^*[k] = \sum_{i=0}^{p-1} W_i^* \, \widetilde{\Phi}_i[k] = \sum_{i=0}^{p-1} c_i \, \widetilde{\Phi}_i[k] \,, \tag{10}$$

i.e., the projection of $s[k]$ onto the subspace spanned by $\{\widetilde{\Phi}_i[k];\, i = 0, \ldots, p-1\}$ with $p \leq N$. Therefore, $y^*[k]$ is the rank p transformed domain representation of $s[k]$, and $y^*[k] = s[k]$ when $p = N$ (i.e., if all of the basis functions are used in the expansion).

The minimum mean square error, ξ_{min}, will be

$$\xi_{min} = E\{d^2[k]\} - \mathbf{P}^T\mathbf{W}^* = \frac{1}{N}\sum_{i=p}^{N-1} c_i^2 + E\{n^2[k]\} \,. \tag{11}$$

[1] A more detailed analysis with the actual time-variant behavior of $\mathbf{R}[n]$ can be found in [31]. In consequence, all the results obtained (steady-state misadjustment, convergence time, etc.) with this approximation must be interpreted as mean values over a signal occurrence.

The remaining noise due to the misadjustment (M) depends upon the adaptive algorithm used to adjust the weight vector [25]. In this study, we use the LMS algorithm for updating the coefficients, $\mathbf{W}[k+1] = \mathbf{W}[k] + 2\,\mu\, e[k]\, \widetilde{\mathbf{\Phi}}[k]$, because it gets the best trade-off between simplicity and convergence time. The condition that assures the convergence of the LMS algorithm [32] is $0 < \mu < \frac{N}{p+2}$. The time constant for the convergence of the MSE is $\tau_{mse} = \frac{1}{4\,\mu\,\lambda} = \frac{N}{4\,\mu}$ samples where $\lambda = 1/N$ is the eigenvalue of the matrix $\mathbf{R}$ (all the eigenvalues are identical). Thus, the gain constant μ controls the stability and the speed of convergence.

To measure the steady-state excess of mean square error ξ_{ex} we calculate the misadjustment, which for the LMS algorithm can be approximated by [25]

$$M = \frac{\xi_{ex}}{\xi_{min}} \simeq \frac{\mu\, tr[\mathbf{R}]}{1 - \mu\, tr[\mathbf{R}]} = \frac{\mu\, p}{N - \mu\, p}\,. \tag{12}$$

The steady-state mean square error ξ is

$$\xi = \xi_{min} + \xi_{ex} = \xi_{min}(1+M) \simeq \left(\frac{1}{N} \sum_{i=p}^{N-1} c_i^2 + E\{n^2[k]\} \right) \left(1 + \frac{\mu\, p}{N - \mu\, p} \right) \tag{13}$$

But in this application we are interested in evaluating the energy of the difference signal between the original clean signal and the reconstruction $e'[k] = s[k] - y[k]$. From the expression of $e[k]$ we get that $e[k] = e'[k] + n[k] = s[k] - y[k] + n[k]$ and taking square expected values we get

$$\begin{aligned} \xi &= E\{(s[k]-y[k])^2\} + E\{n^2[k]\} + 2\,E\{(s[k]-y[k])n[k]\} \\ &= MSE_p^{LMS} + E\{n^2[k]\} + 2\,E\{y[k]\,n[k]\}\,. \end{aligned} \tag{14}$$

The ALC with deterministic reference inputs is equivalent to a linear system [23], and therefore the output signal, $y[n]$, can be decomposed as the sum of the outputs $y_s[k]$ and $y_n[k]$ corresponding to the inputs $s[k]$ and $n[k]$ respectively. The output $y_s[k]$ is deterministic because in these case both inputs ($s[k]$ and $\widetilde{\mathbf{\Phi}}[k]$) are deterministic. Thus the last term in equation (14) can be evaluated as

$$\begin{aligned} E\{y[k]\,n[k]\} &= E\{(y_s[k] + y_n[k])\,n[k]\} \\ &= E\{y_n[k]\,n[k]\}\,. \end{aligned} \tag{15}$$

In the particular case of white noise, this term is null as it was demonstrated in [31, 33]. Therefore, the mean square error between the original clean signal and the reconstructed signal with the LMS algorithm for white noise will be

$$MSE_p^{LMS} = \xi - E\{n^2[k]\} \simeq \frac{1}{N-\mu\, p} \sum_{i=p}^{N-1} c_i^2 + \frac{\mu\, p}{N-\mu\, p}\sigma^2 \,. \tag{16}$$

Two different terms can be considered in (16): the first one is due to the truncation signal error and the second one is due to the presence of noise in the input signal and misadjustment of the adaptive algorithm. There is a clear a trade-off in the selection of the number of functions p, in a similar way than for inner product in equation (3): high values of p reduces the first term, but also increases the second one. The noise in the reconstruction will be, analogous to equation (4),

$$NOISE_p^{LMS} = \frac{\mu\, p}{N-\mu\, p}\frac{1}{N} \sum_{i=p}^{N-1} c_i^2 + \frac{\mu\, p}{N-\mu\, p}\sigma^2 \,. \tag{17}$$

The SNR of the estimated signal $y[k]$ after convergence can be calculated as

$$SNR_p^{LMS} = \frac{\frac{1}{N}\sum_{i=0}^{p-1} c_i^2}{\left(\frac{\mu p}{N-\mu p}\right)\left(\frac{1}{N}\sum_{i=p}^{N-1} c_i^2 + \sigma^2\right)} \,. \tag{18}$$

Equations (16-18) are derived after the convergence of the weights with stationary signals and white noise. Experimental results with non-stationary signals will present some differences.

5 Comparison of inner product *versus* the LMS algorithm

In this section we compare the performance of the two methods for estimating the coefficients: inner product and adaptive estimation. Both techniques have some advantages and some drawbacks. The inner product follows the dynamic changes of the signal because it is the beat-to-beat projection of the signal vector onto the subspace of analysis. The dynamic changes of the ECG signal are directly shown in the evolution of the coefficients in the transformed domain. The main drawback of the inner product is that both signal and noise components are projected in the same way, so the reconstructed signals will be noisy. In contrast, the adaptive estimation of the coefficients can attenuate the noise uncorrelated with the signal achieving an improvement of the signal to noise ratio. But in this case the adaptive algorithm needs a period of time for the convergence. This is the well-known adaptive algorithms trade-off: signal to noise ratio improvement at steady state (related to μ value) *versus* time of convergence.

In order to compare both techniques we calculate the value of the step-size parameter of the LMS algorithm $\mu = \mu_{lim}$ that gets the same value of SNR at the output signal than the inner product. Using equations (5) and (18) the improvement of SNR of the LMS algorithm *versus* inner product at the reconstructed signal in the case of white noise will be

$$\Delta SNR_p^{LMS/direct} = \frac{SNR_p^{LMS}}{SNR_p^{direct}} = \frac{\frac{p}{N}\sigma^2}{\left(\frac{\mu p}{N-\mu p}\right)\left(\sigma^2 + \frac{1}{N}\sum_{i=p}^{N-1} c_i^2\right)}\,, \tag{19}$$

and doing $\Delta SNR_p^{LMS/direct} = 1$ we get the value of the step-size

$$\mu_{lim} = \frac{N\,\sigma^2}{(N+p)\,\sigma^2 + \frac{1}{N}\sum_{i=p}^{N-1} c_i^2}\,. \tag{20}$$

It can be seen from equation (19) that when the complete expansion is used ($p = N$) then $\Delta SNR_N = \frac{1-\mu}{\mu} = \frac{1}{M}$, and this factor is equal to one for $\mu_{lim} = 0.5$. This result can be corroborated in equation (20). When the noise energy is much more important than truncation signal error, the LMS algorithm has more advantages because it attenuates more efficiently the noise energy than direct estimation. As a consequence of that, the value of μ_{lim} for noisy signals (low values of SNR) is higher than for cleaner signals. The convergence condition must be accomplished for convergence of the algorithm. For low values of p, signal truncation error becomes more important, and the value of μ_{lim} decreases. If a value of $\mu < \mu_{lim}$ is selected the adaptive estimation of the coefficients gets cleaner reconstructed signals (18) than inner product (5) for stationary signals. If the step-size is selected as $\mu = \mu_{lim}$, it can be demonstrated that after some manipulation on equations (3), (16) and (20) the reconstruction error (MSE_p) for inner product and the LMS algorithm will be the same

$$MSE_p^{LMS}\Big\rfloor_{\mu=\mu_{lim}} = MSE_p^{direct}\,. \tag{21}$$

The values of μ_{lim} for the KLT of ECG training set with various levels of white noise are shown in Fig. 9. These values have been calculated from equation (20) and the eigenvalues of the covariance matrix.

The performance of both estimation techniques can be compared in Fig. 10. We consider two different values of SNR (15 and 10 dB) and three different values of step-size μ (0.1, 0.3 and 0.75) for the LMS algorithm. We can see in Fig. 9 that for SNR=15 dB, μ_{lim}=0.75 at p=31 and p=131. Moreover, using equation (21) for these values of p we should obtain the same value of MSE for inner product and the LMS, as it can be seen in Fig. 10(a). Similarly, for SNR=10 dB, μ_{lim}=0.75 at p=16 and 140, and

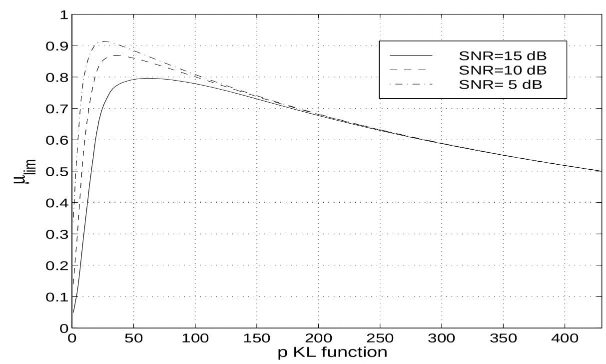
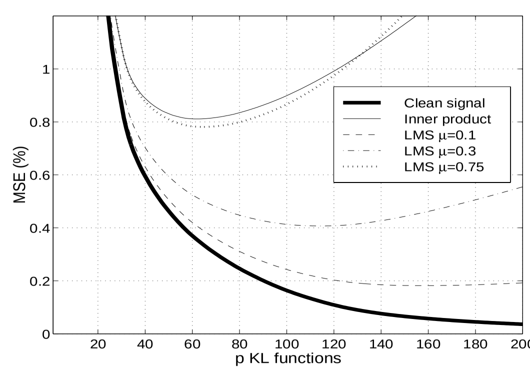
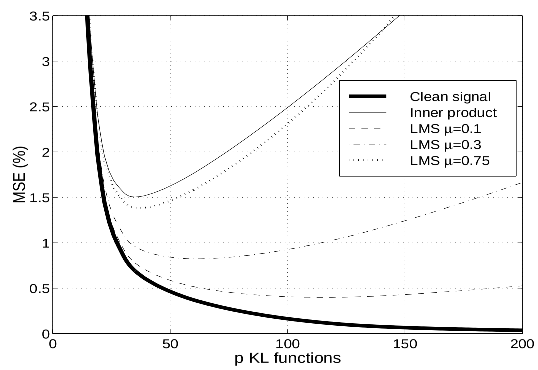


Figure 9: *Value of μ_{lim} in LMS for white noise.*

these values get the same MSE for both techniques with μ=0.75 (see Fig. 10(b)). In addition, the LMS algorithm gets lower values of MSE than inner product always that $\mu < \mu_{lim}$ for any value of p.



(a) SNR=15 dB



(b) SNR=10 dB

Figure 10: *MSE_p for inner product and LMS of signals in white noise.*

In order to compare the performance of both estimation methods, we calculate the ratio $MSE_p^{LMS/direct}$ that will be

$$MSE_p^{LMS/direct} = \frac{MSE_p^{LMS}}{MSE_p^{direct}} = \frac{N}{N-\mu p} \frac{\mu\, p\, \sigma^2 + \sum_{i=p}^{N-1} c_i^2}{p\, \sigma^2 + \sum_{i=p}^{N-1} c_i^2} \,. \tag{22}$$

Figure 11 shows this ratio for two values of SNR (10 and 15 dB) and for μ=0.1, 0.3, 0.5 and 0.75. When the whole basis is used (p=N) the MSE improvement is equal to the misadjustment

$$MSE_N^{LMS/direct} = \frac{\mu}{1-\mu} = M\,. \tag{23}$$

The maximum MSE difference between the LMS and direct estimation will occur at relative low values of p. This behavior is appropriate for data compression systems, where low values of functions are desired. When the SNR is lower (more noisy signals) the optimum number of functions p is lower too and the improvement of LMS versus direct estimation is larger.

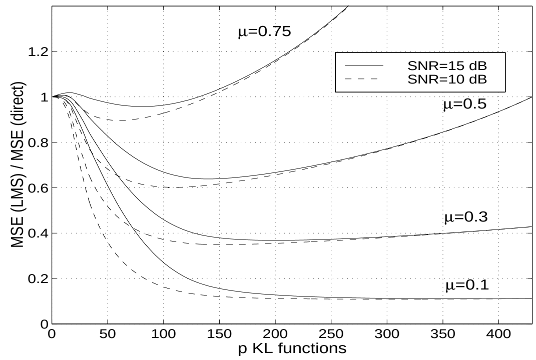


Figure 11: *Values of* $MSE_p^{LMS/direct}$ *for white noise.*

In summary, the analytic results show that the LMS algorithm with a value of $\mu < \mu_{lim}$ gets lower values of steady-state MSE than inner product for *stationary signals*. It can be seen that when the whole basis is used (p=N) the MSE_N improvement ratio of LMS *versus* direct estimation is $\frac{\mu}{1-\mu}$, but greater improvement can be achieved at lower values of p. For signals with low values of SNR, the performance of the LMS algorithm can get a great improvement over the inner product estimation if a low value of μ is selected.

From the stationary analytic results it is concluded that the best selection of the step-size μ will correspond to as low as possible values. However, the choice of a very small μ could have problems with the dynamic ECG changes and would increase the value of MSE. Then a study with real non-stationary signals is required.

6 The LMS algorithm with non-stationary signals

The non-stationary behavior of the ECG signal can be understood as beat-to-beat morphology changes. In this situation, the LMS algorithm has the task of not only seeking the minimum point of the error performance surface but also tracking the beat to beat changing position of the minimum $\mathbf{W}*$. The optimum weight vector will be fixed during the N samples of every heartbeat (ECG signal occurrence), and it will suddenly change with every new heartbeat.

The selection of the step-size μ will have now a trade-off between noise reduction capability (requiring low values of μ) and speed of adaption to track the time variant optimum weight vector (requiring high values of μ). In this situation we do not have infinite time for updating the weights as it was in sections 4 and 5. We update the weight vector during N samples (heartbeat duration) and then we will reconstruct the signal with the value of the weight vector at the end occurrence time $\mathbf{W}[i\,N]$.

The convergence analysis of the LMS algorithm have been previously analyzed by many authors, being an active research area. Some of the papers deal with stationary signals (infinite adaptation time is available) with either random inputs [24, 32, 34–36], or deterministic inputs [30, 37]. Other papers deal about slow time-varying signals where the signal dynamics is modeled with random walk [38–42] or Markov chains [43]. However, very few authors have analyzed the MSE of the LMS algorithm after a finite number of iterations [44], as it is the actual situation in many applications.

In our application we are interested in evaluating the mean square error between every occurrence i of the original signal and the reconstruction using the $p \times 1$ weight vector at the end of the occurrence $\mathbf{W}[i\,N]$, i.e.,

$$J_p = \sum_{k=0}^{N-1} \left(s[(i-1)N+k] - y'[(i-1)N+k]\right)^2 \qquad 0 < k < N-1 \tag{24}$$

where $y'[(i-1)N+k] = \mathbf{W}^T[iN]\widetilde{\mathbf{\Phi}}[k]$. This cost function differs from standard $\xi[k] = E\{e^2[k]\}$, because J_p is a global distortion measure over the whole i-th signal occurrence when the signal is reconstructed with the weight vector at the end occurrence time $\mathbf{W}[iN]$, while $\xi[k]$ is an instantaneous distortion measure using the instantaneous output signal $y[k]$. We use vectors (bold letters) to denote signals, like $\mathbf{s} = [s_0 s_1 \cdots s_{N-1}]^T$. Then the cost function can be written as

$$J_p = (\mathbf{s} - \mathbf{y}')^T(\mathbf{s} - \mathbf{y}') = (\mathbf{s} - \mathbf{M}^T\mathbf{W}[iN])^T(\mathbf{s} - \mathbf{M}^T\mathbf{W}[iN]), \tag{25}$$

where $\mathbf{M}$ is the $p \times N$ matrix of orthogonal basis functions $\mathbf{M} = [\widetilde{\mathbf{\Phi}}[0]\,\widetilde{\mathbf{\Phi}}[1]\cdots\widetilde{\mathbf{\Phi}}[N-1]]$.

The weight vector at the end of the occurrence can be calculated applying N times the weight vector update equation of the LMS algorithm $\mathbf{W}[k+1] = \left(\mathbf{I} - 2\,\mu\,\widetilde{\mathbf{\Phi}}[k]\widetilde{\mathbf{\Phi}}^T[k]\right)\mathbf{W}[k] + 2\,\mu\,d[k]\widetilde{\mathbf{\Phi}}[k]$, giving

$$\mathbf{W}[N] = \left(\prod_{k=0}^{N-1}(\mathbf{I} - 2\mu\widetilde{\mathbf{\Phi}}[k]\widetilde{\mathbf{\Phi}}^T[k])\right)\mathbf{W}[0] + 2\mu\sum_{k=0}^{N-1} d[k]\left(\prod_{i=k+1}^{N-1}(\mathbf{I} - 2\mu\widetilde{\mathbf{\Phi}}[i]\widetilde{\mathbf{\Phi}}^T[i])\right)\widetilde{\mathbf{\Phi}}[k], \tag{26}$$

where the time origin has been selected at the beginning of the signal occurrence for simplicity. $\mathbf{W}[N]$ only depends on the step-size μ, the primary input signal $\mathbf{d}$ and the initial weight vector $\mathbf{W}[0]$. The terms $\prod_k(\mathbf{I} - 2\mu\widetilde{\mathbf{\Phi}}[k]\widetilde{\mathbf{\Phi}}^T[k])$ of equation (26) can be calculated *a priori* because they only depend on the basis functions. Given an initial weight vector $\mathbf{W}[0]$ and the primary input signal $\mathbf{d}$, the weight vector at the end occurrence time $\mathbf{W}[N]$ is an N-degree polynomial of μ (see equation (26)) where the coefficients are matrices and vectors respectively

$$\mathbf{W}[N] = \left(\mathbf{I} - 2\mu\mathbf{A}_1 + (2\mu)^2\mathbf{A}_2 - (2\mu)^3\mathbf{A}_3 + \cdots\right)\mathbf{W}[0] + 2\mu\mathbf{B}_1 - (2\mu)^2\mathbf{B}_2 + (2\mu)^3\mathbf{B}_3 - \cdots \tag{27}$$

The matrices $\mathbf{A}_i$ depend only on the basis functions while the vectors $\mathbf{B}_i$ also depend on the primary input signal $\mathbf{d}$.

Looking at one signal occurrence, we will have an initial weight vector $\mathbf{W}[0]$ (result of the previous occurrence adaptation) that is far from the optimum weight vector $\mathbf{W}^*$ of the current occurrence, may be due to abrupt signal changes. In this situation the adaption time is finite (N samples) and we would like to select the optimum value of the LMS step-size μ that minimizes the cost function J_p.

Firstly, we consider the special case of complete expansions ($p{=}N$) as an introduction, and afterwards the more general case of non-complete expansions case is discussed.

6.1 Complete expansions

This particular case is not interesting for data compression because there is no rank reduction, but several authors have studied it for filtering applications using impulses as basis functions [27, 29, 45, 46]. When all basis functions are used in the expansion ($p{=}N$) equation (27) can be greatly simplified because $\mathbf{A}_1 = \sum_{i=0}^{N-1} \widetilde{\mathbf{\Phi}}[k]\widetilde{\mathbf{\Phi}}^T[k] = \mathbf{I}$, $\mathbf{A}_2 = \mathbf{A}_3 = \cdots = \mathbf{A}_N = \mathbf{0}$ because of the orthogonality property of the basis functions $\widetilde{\mathbf{\Phi}}^T[i]\widetilde{\mathbf{\Phi}}[j] = \delta_{ij}$ being δ_{ij} the Kronecker delta function) and $\mathbf{B}_1 = \mathbf{M}\mathbf{d}$, $\mathbf{B}_2 = \mathbf{B}_3 = \cdots = \mathbf{B}_N = \mathbf{0}$ (due to the same orthogonality property), giving the equation

$$\mathbf{W}[N] = (1 - 2\,\mu)\,\mathbf{W}[0] + 2\,\mu\,\mathbf{M}\,\mathbf{d}\,. \tag{28}$$

Now we can study the selection of the optimum value of μ that minimizes J_N. In this case the cost function will be

$$J_N = (\mathbf{s} - \mathbf{y}')^T(\mathbf{s} - \mathbf{y}') = (\mathbf{s} - \mathbf{M}^T\mathbf{W}[N])^T(\mathbf{s} - \mathbf{M}^T\mathbf{W}[N])\,, \tag{29}$$

where $\mathbf{W}[N]$ is given in (28). The value of the step-size μ that minimizes J_N is found forcing zero at the first derivative $\frac{\partial J_N}{\partial \mu}$ obtaining

$$\mu_{opt} = \frac{1}{2}\frac{(\mathbf{W}[0] - \mathbf{M}\mathbf{s})^T(\mathbf{W}[0] - \mathbf{M}\mathbf{s}) + (\mathbf{W}[0] - \mathbf{M}\mathbf{s})^T\mathbf{M}\mathbf{n}}{(\mathbf{W}[0] - \mathbf{M}\mathbf{d})^T(\mathbf{W}[0] - \mathbf{M}\mathbf{d})}\,. \tag{30}$$

If we calculate the second derivative we obtain

$$\frac{\partial^2 J_N}{\partial \mu^2} = 8\,(\mathbf{W}[0] - \mathbf{M}\mathbf{d})^T(\mathbf{W}[0] - \mathbf{M}\mathbf{d})\,, \tag{31}$$

that is always positive, thus μ_{opt} is the minimum of the cost function J_N. Equation (30) can be geometrically interpreted as the ratio of inner products of the transform domain vectors shown in Fig. 12.

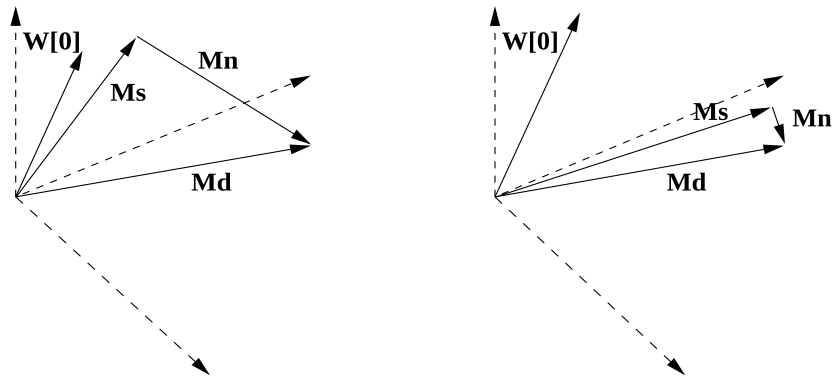
When the observed signal $\mathbf{d}$ is equal to the desired signal $\mathbf{s}$ (we do not need adaptive filtering at all because there is no noise) the optimum value is $\mu_{opt} = \frac{1}{2}$. This value of the step-size makes the LMS algorithm equivalent to the inner product because $\mathbf{W}[N] = \mathbf{M}\mathbf{d}$ according to equation (28). Moreover, the mean square error J_N is zero, because all the basis functions are used in the expansion.

When noise is present $\mathbf{n}=\mathbf{d}-\mathbf{s}$, we have a trade-off between tracking capability and misadjustment. When the noise vector norm (noise energy) is much larger than the beat-to-beat morphology changes (Fig. 12(a)), μ_{opt} will be much lower than 0.5 in order to reduce the noise energy. In this case a good initialization of the weight vector $\mathbf{W}[0]$ is the weight vector from the last signal occurrence (like in the classical LMS) that contains information of previous signal occurrences. The opposite case will be when the noise energy is low compared with the morphology signal change (Fig. 12(b)), and the optimum value will be near to μ_{opt}=0.5. In this case a good initialization of the weight vector of every signal occurrence is the inner product of the noisy observed signal occurrence $\mathbf{W}[0] = \mathbf{M}\mathbf{d}$. The LMS algorithm will try to reduce the noise energy during the adaptation time (N iterations). If all basis functions are used for updating the coefficients with the LMS algorithm the weight vector at the end occurrence time will be

$$\mathbf{W}[N] = (1 - 2\,\mu)\,\mathbf{M}\,\mathbf{d} + 2\,\mu\,\mathbf{M}\,\mathbf{d} = \mathbf{M}\,\mathbf{d}\,, \tag{32}$$

that is the same result than inner product coefficient vector, independently of the value of the step-size μ. When both, noise and non-stationarities coexist and have similar magnitudes nothing can be said *a priori*, and the equation (30) should be evaluated with the actual vectors (signal occurrences).

In section 2 was announced that the orthogonal compression system is independently applied to each morphology in order to improve compression applying differentially coders like DPCM or ADPCM to the coefficient time series. So we have independent weight vectors for normal beats, ventricular beats, and so on. In this situation each weight vector time series will have smaller changes (only due to variations from beats of the same morphology that may be non-contiguous in time).



(a) High noise and small non-stationarities. *(b) Low noise and large non-stationarities.*

Figure 12: *Geometric interpretation of* μ_{opt}.

6.2 Non-complete expansions

When non-complete expansions ($p<N$) are used, we will obtain a similar conceptual behavior but with a more complex description. Equation (27) can not be simplified as it was on the complete expansions case. The linear terms on μ ($\mathbf{A}_1$ and $\mathbf{B}_1$) are the same than for complete expansions case ($\mathbf{A}_1 = \mathbf{I}$, $\mathbf{B}_1 = \mathbf{Md}$), but now the other terms are different to null vector ($\mathbf{A}_2$, $\mathbf{A}_3, \cdots, \mathbf{A}_N \neq \mathbf{0}$ and $\mathbf{B}_2$, $\mathbf{B}_3, \cdots, \mathbf{B}_N \neq \mathbf{0}$). For example, the matrix $\mathbf{A}_2$ will be a sum of terms of the form

$$\left(\widetilde{\mathbf{\Phi}}[i]\widetilde{\mathbf{\Phi}}^T[i]\right)\left(\widetilde{\mathbf{\Phi}}[j]\widetilde{\mathbf{\Phi}}^T[j]\right) = r_{ij}\,\widetilde{\mathbf{\Phi}}[i]\widetilde{\mathbf{\Phi}}^T[j] \tag{33}$$

where the scalar value $r_{ij} = (\widetilde{\mathbf{\Phi}}^T[i]\widetilde{\mathbf{\Phi}}[j]) \neq \delta_{ij}$ for non-complete expansions. In the same way, the matrix $\mathbf{A}_3$ will be a sum of terms of the form

$$\left(\widetilde{\mathbf{\Phi}}[i]\widetilde{\mathbf{\Phi}}^T[i]\right)\left(\widetilde{\mathbf{\Phi}}[j]\widetilde{\mathbf{\Phi}}^T[j]\right)\left(\widetilde{\mathbf{\Phi}}[k]\widetilde{\mathbf{\Phi}}^T[k]\right) = r_{ij}\,r_{jk}\widetilde{\mathbf{\Phi}}[i]\widetilde{\mathbf{\Phi}}^T[k]\,. \tag{34}$$

The terms $\mathbf{B}_2, \mathbf{B}_3, \cdots$ also have a similar behavior.

The analysis of the optimum value of the step-size μ_{opt} for non-complete expansions is now troublesome due to the non-null interaction between basis functions. Alternatively, we propose a experimental study to find the values of μ_{opt} in a training set of simulated noisy ECG signals.

7 Experimental study

In order to study the effect of noise in estimating the coefficients of orthogonal transforms it is proposed the following simulation study represented in Fig. 13. Three different kinds of noise have been added to

ECG records from the MIT-BIH Arrythmia Database: simulated Gaussian white noise, and two records of physiological noise: electrode motion ('*em*') and muscular activity noise ('*ma*'). The level of noise added is much higher than the unavoidable noise present in original records. A data compression system based on truncated orthogonal expansions is applied to the simulated noisy ECG signals. The MSE index is evaluated between the reconstruction $\mathbf{X}_R$ and the clean original signal $\mathbf{S}$. We propose this simulation because in actual applications we can not get access to the clean signal $\mathbf{S}$.

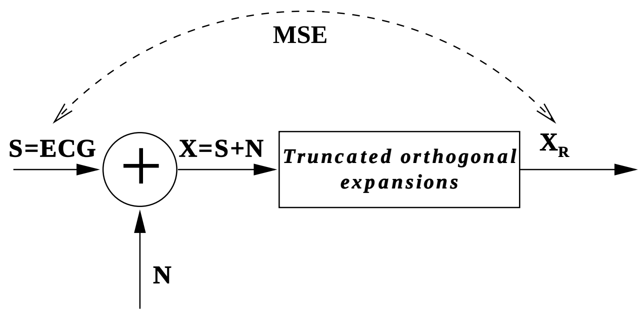
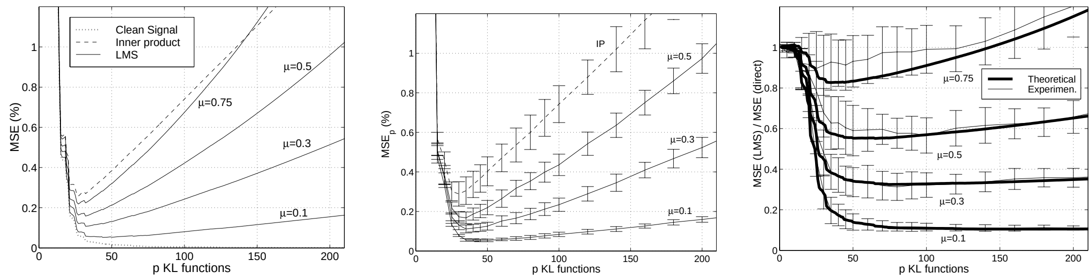


Figure 13: *Simulation of additive noisy signals.*

7.1 Simulated stationary ECG signals in white noise

In the first step, a simulation is proposed to evaluate the steady-state performance of inner product and the LMS algorithm only considering stationary ECG signals and white noise. We simulated a 100 heartbeats ECG segment repeating an average beat of record 103 from MIT-BIH Arrythmia Database. This signal is perfectly periodic and also deterministic. The length of 100 heartbeats is long enough to the convergence of the LMS algorithm and the steady-state MSE analysis. We added to it white noise with several values of SNR (0, 5, 10 and 15 dB). Both transform coefficient estimation methods are applied with a variable number of basis functions. The results obtained for SNR=15 dB are shown in Fig. 14. Theoretical values (subfigure (a)) are calculated using equations (3) and (16) for the selected beat. There is a very small difference between the selected beat and the mean performance for the whole ECG training set. For the experimental analysis we make 20 trials. Mean values and standard deviation of the steady-state MSE_p are shown in Fig. 14(b). The experimental results for white noise are very close to predicted values. The reason is that the hypothesis made in the derivation of theoretical expressions (stationarity of signal and noise, and mutually uncorrelated) were true in this simulation. Experimental results of the ratio $MSE_p^{LMS/direct}$ are compared in Fig. 14(c) with theoretical values derived in equation (22). It can be seen that experimental results are well predicted, especially for low values of the step-size μ.

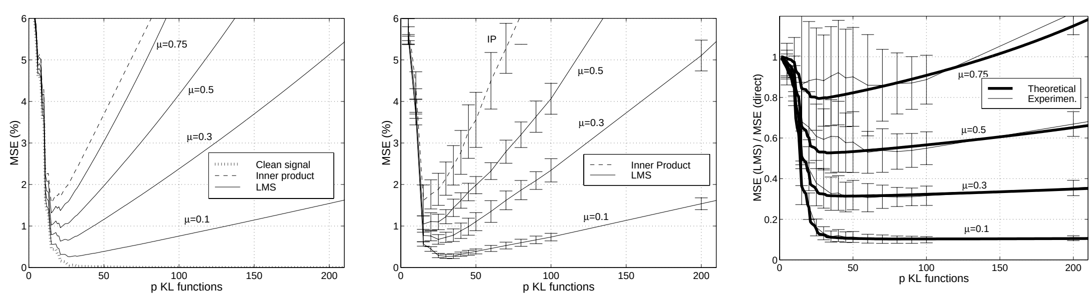
If the noise energy is larger, the value of MSE_p will be also higher for both estimation methods,



(a) Theoretical results *(b) Experimental results* *(c) MSE ratio*

Figure 14: *Theoretical/experimental values of MSE_p for white noise with SNR=15 dB.*

inner product and the LMS algorithm. However, both methods do not increase with the same law. The equivalent results are illustrated in Fig. 15. We can observe in Fig.15 (c) that the maximum improvement of the LMS algorithm *versus* inner product is a bit more higher than in the case of SNR=15 dB (Fig. 14(c)), and also this maximum difference is found at lower values of the number of functions. When complete expansions are used (p=N=430), it is corroborated that the improvement of the LMS algorithm *versus* inner product is independent of the SNR and only depends on the value of the step-size μ (see equation (23)).



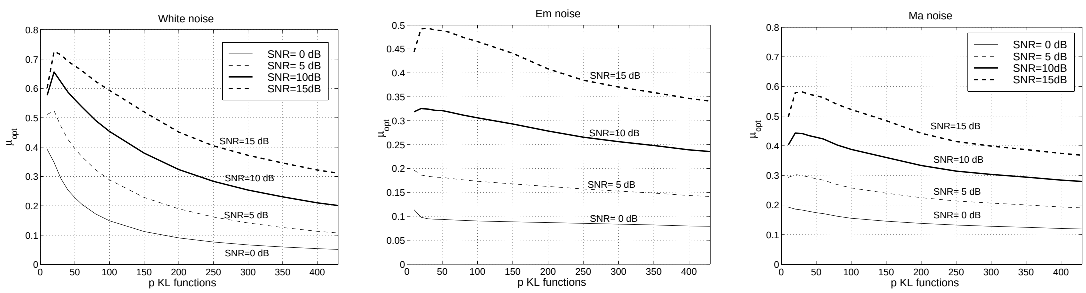
(a) Theoretical results *(b) Experimental results* *(c) MSE ratio*

Figure 15: *Theoretical/experimental values of MSE_p for white noise with SNR=5 dB.*

The selection of the step-size μ in the LMS algorithm should be selected according to the SNR of the original signal and the number of functions p. For white noise and stationary ECG signals, the best way for estimating the transform coefficients is the LMS algorithm with very low values of μ because it can attenuate the uncorrelated noise with a low value of steady-state misadjustment. The limitation is that the convergence time is high. In actual applications with time-varying ECG signals, the LMS algorithm will have to track the dynamic signal.

7.2 Actual ECG signals contaminated by noise

In order to calculate the optimum value of the LMS step-size μ_{opt} for non-complete expansions with non-stationary noisy ECG signals we select the first 5 min. of 20 records from MIT-BIH Arrythmia database. For every heartbeat and for all values of the number of functions p, we find the value of the step-size μ that minimized the cost function (29) using a numerical minimum search. We show in Fig. 16 the mean values of μ_{opt} obtained for the three kinds of noise with several values of SNR.



(a) *White noise.* (b) *'Em' noise.* (c) *'Ma' noise.*

Figure 16: *Mean values of μ_{opt} with actual ECG noisy signals.*

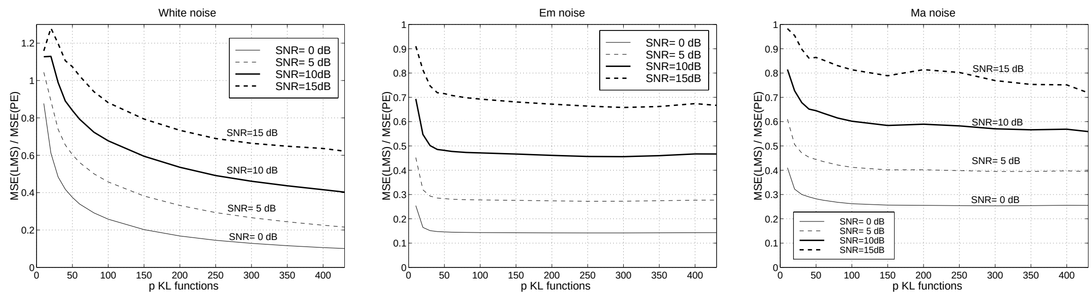
It can be seen that the mean value of μ_{opt} is lower when the noise energy is higher for all kinds of noise. This behavior is reasonable because when the impact of the noise energy is higher than beat-to-beat morphology changes, the most efficient choice of the step-size are low values of μ to attenuate the uncorrelated noise with low values of misadjustment. In contrast, when the effect of the noise energy is lower than beat-to-beat morphology changes, higher values of the step-size should be used to increase the convergence speed to track the dynamic changes.

Once the optimum value of the step-size μ_{opt} are determined, we can calculate the MSE obtained with the LMS algorithm and compare it with the inner product. We show in Fig. 17 the MSE improvement obtained with the LMS algorithm respect to inner product $MSE_p^{LMS/direct}$. We observe that now the improvement is less important than for stationary signals. Even in some conditions (high values of SNR and white noise) the inner product obtains lower values of distortion.

We can conclude that the adaptive estimation of the transform coefficients with the LMS algorithm is more appropriate than inner product for low values of SNR. Moreover, the improvement of the LMS algorithm is higher for the case of physiological noise than for white noise for $p<50$. Using typical values of SNR 10-15 dB in actual ECG records and number of basis functions between 30-50 (compression ratio[2] of 8.6-14.3), we obtain that the LMS algorithm is more or less equivalent to the inner product for white

[2] The value of compression ratio is only an approximation without considering quantization

noise, but when physiological noise is considered an improvement in MSE_p of about 40% for μ values between 0.3–0.5.



(a) White noise. *(b) 'Em' noise.* *(c) 'Ma' noise.*

Figure 17: Mean value of the ratio $MSE_p^{LMS/direct}$.

We show in Fig. 18 an example of original and reconstructed signals with p=50 basis functions. The original signal is taken from record 103 of MIT-BIH Arrythmia database. It is contaminated with *em* noise with a value of SNR=10 dB. The first line represents the actual ECG signal, the second line is the *em*-noise added to it. Noise morphology is similar to some ECG waveforms. The third line is the simulated noisy ECG signal (addition of the previous signals). The last two lines are the outputs of the reconstructed signals with inner product and the LMS algorithm with μ=0.1 respectively. It can be clearly seen that inner product can not attenuate the *em*-noise (for example at the ST-T complex of the first two beats) because the signal and noise KL representations are overlapped. However, the LMS algorithm obtains a more clean reconstruction.

8 Conclusions

In this work we have analyzed the performance of truncated orthogonal expansions for compressing ECG signals when the input signals are contaminated by additive noise. Due to the presence of noise, the distortion (evaluated as mean square error between the original clean signal and the reconstruction) has two different sources: signal truncation error and noise error. We have quantified the relative importance of both terms when a variable number of functions p are used in the expansion. We distinguish two classical methods for estimating the transform coefficients: inner product and adaptive estimation with the LMS algorithm.

A simulation study has been proposed where noisy ECG signals are generated from actual ECG records from MIT-BIH Arrythmia database and adding three different noise sources: simulated Gaussian white noise, and two records of physiological noise. All derived results are contrasted with simulated

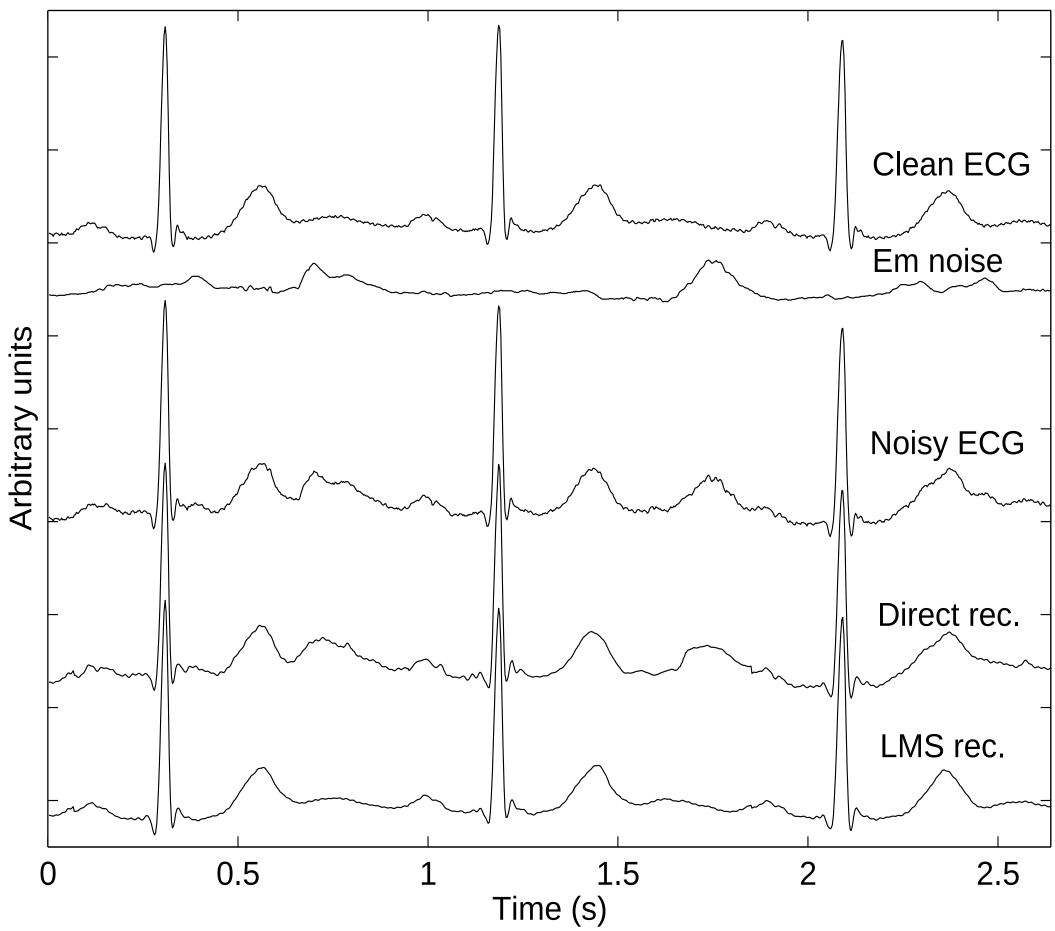


Figure 18: *Example of ECG signal reconstructions from record 103 with em noise SNR=10 dB and p=50 KL basis functions.*

data.

When the transform coefficients are calculated using the inner product the effect of noise in the reconstructed signal for physiological noise is higher than for white noise because signal and noise representations in the transformed domain are more overlapped in the first case, and the inner product estimation can not distinguish between signal and noise. Analytical results of distortion and SNR in the reconstructed signal for any value of number of functions are given.

We have also considered an adaptive estimation of the coefficients with the LMS algorithm in order to reduce the uncorrelated noise. Two different situations of the LMS algorithm are analyzed: stationary (without changes in the deterministic component of the signal) and non-stationary (actual ECG signals with beat-to-beat morphology changes) input signals.

In the ideal case of stationary signals we have analyzed the steady-state performance of the LMS algorithm making a comparison with inner product results. We calculate the value of the LMS step-size μ_{lim} that get the same performance than inner product. If the step-size is chosen as $\mu < \mu_{lim}$, the LMS algorithm gets better steady-state performance than inner product.

However, actual ECG records have beat-to-beat morphology changes, and there is not infinite time for the adaptation process. We define the distortion for each beat as the mean square error between the original signal and the reconstructed signal with the weight vector at the end of each occurrence. In the case of complete expansions, the analytical expression of the value of the step-size μ_{opt} that minimizes the distortion after a finite-time adaptation process is given. For non-complete expansions only experimental results are given.

With the shown methodology we can give some practical criteria for the selection of the more appropriate transform coefficient estimation method (inner product or adaptive estimation) and the choice of the step-size μ for the LMS algorithm. For example, in typical Holter recording, the value of SNR is around 10 dB, and therefore the optimum value of the LMS step-size should be around μ=0.3–0.5 if physiological noise is present with p around 40 KL basis functions. The improvement of the LMS algorithm over the inner product will be around 40% for these operating conditions.

References

[1] S. M. S. Jalaleddine, C. G. Hutchens, R. D. Strattan, and W. A. Coberly, "ECG data compresssion techniques: A unified approach," IEEE Trans. Biomed. Eng., vol. BME-37, pp. 329–341, Apr. 1990.

[2] J. P. Abenstein and W. J. Tompkins, "A new data reduction algorithm for real-Time ECG analysis," IEEE Trans. Biomed. Eng., vol. BME-29, no. 1, pp. 43–48, 1982.

[3] M. Ishijima, S. B. Shin, G. H. Hostetter, and J. Sklansky, "Scan Along Polygonal Approximation for data compression of electrocardiograms," IEEE Trans. Biomed. Eng., vol. BME-30, pp. 723–729, 1983.

[4] W. J. Tompkins, Biomedical Digital Signal Processing, ch. 10. Data reduction techniques, pp. 193–215. Prentice-Hall, 1993.

[5] P. S. Hamilton and W. J. Tompkins, "Compression of the ambulatory ECG by Average Beat Subtraction and Residual Differencing," IEEE Trans. Biomed. Eng., vol. BME-38, no. 3, pp. 253–259, 1991.

[6] G. Barlas and E. Skordalakis, "A novel family of compression algorithms for ECG and other semiperiodical, one-dimensional, biomedical signals," IEEE Trans. Biomed. Eng., vol. 43, pp. 820–828, Aug 1996.

[7] A. E. Çetin, H. Koymen, and M. C. Aydin, "Multichannel ECG data compression by multirate signal processing and transform domain coding techniques," IEEE Trans. Biomed. Eng., vol. 40, pp. 495–499, 1993.

[8] J. D. Bronzino, The Biomedical Engineering Handbook, ch. 56. Compression of Digital Biomedical Signals, pp. 853–865. CRC Press, 1995.

[9] B. Bradie, "Wavelet package-based compression of single lead ECG," IEEE Trans. Biomed. Eng., vol. 43, pp. 493–501, 1996.

[10] T. Blanchett, G. C. Kember, and G. A. Fenton, "KLT-based quality controlled compression of single-lead ECG," IEEE Trans. on Signal Proc., vol. 45, pp. 942–945, Jul 1998.

[11] L. Scharf and D. Tufts, "Rank reduction for modeling stationary signals," IEEE Trans. Acous. Sig. Proc., vol. BME-35, pp. 350–355, 1987.

[12] R. D. C. Lamberti, M. Zagnoni and G. Bortolan, "Evaluation of algorithms for real-time ECG data compression," in Proc. Computers in Cardiology, pp. 399–403, IEEE Computer Society Press, 1991.

[13] G. B. Moody and R. G. Mark, "The MIT-BIH arrhythmia database on CD-ROM and software for use with it," in Proc. Computers in Cardiology, pp. 185–188, IEEE Computer Society Press, 1990.

[14] G. B. Moody and R. G. Mark, "Development and evaluation of a 2-lead ECG analysis program," in Proc. Computers in Cardiology, pp. 39–44, IEEE Computer Society Press, 1982.

[15] S. Olmos, M. Millán, J. García, and P. Laguna, "ECG data compression with the Karhunen-Loève transform," in Proc. Computers in Cardiology 1996, pp. 253–256, IEEE Computer Society Press, 1996.

[16] N. S. Jayant and P. Noll, Digital Coding of Waveforms. Englewood Cliffs, N.J.: Prentice-Hall, 1984.

[17] V. A. Allen and J. Belina, "ECG data compression using the discrete cosine transform," in Proc. Computers in Cardiology, pp. 687–690, IEEE Computer Society Press, 1992.

[18] W. Philips and G. D. Jonghe, "Data compression of ECG's by High-Degree Polynomial Approximation," IEEE Trans. Biomed. Eng., vol. 39, no. 4, pp. 330–337, 1992.

[19] R. Jané, S. Olmos, P. Laguna, and P. Caminal, "Adaptive Hermite models for ECG data compression: Performance and evaluation with automatic wave detection," in Proc. Computers in Cardiology, pp. 389–392, IEEE Computer Society Press, 1993.

[20] P. Laguna, G. Moody, J. García, A. Goldberger, and R. Mark, "Analysis of the ST-T complex of the electrocardiogram using the Karhunen-Loève transform: adaptative monitoring and alternans detection," Med. Biol. Eng. Comput., vol. 37, pp. 175–189, 1999.

[21] C. W. Therrien, Discrete random signals and statistical signal processing. Prentice-Hall, 1992.

[22] A. Gersho and R. M. Gray, Vector Quantization and Signal Compression. Boston: Kluwer Academic Publishers, 1992.

[23] S. Olmos, J. Garcia, R. Jané, and P. Laguna, "Truncated orthogonal expansions of recurrent signals: Equivalence to a periodic time-variant filter," in Proc. IEEE Int. Conf. Acous. Speech Signal Proc., (Seattle (USA)), pp. 1709–, 1998.

[24] S. Haykin, Adaptive Filter Theory. New Jersey: Prentice-Hall, Englewood Cliffs, 3rd ed., 1996.

[25] B. Widrow and S. D. Stearns, Adaptive Signal Processing. New Jersey: Prentice-Hall, Englewood Cliffs, 1985.

[26] P. Laguna, R. Jané, S. Olmos, N. V. Thakor, H. Rix, and P. Caminal, "Adaptive estimation of QRS complex by the Hermite model for classification and ectopic beat detection," Medical and Biological Engineering and Computing, vol. 34, pp. 58–68, 1996.

[27] P. Laguna, R. Jané, O. Meste, P. W. Poon, P. Caminal, H. Rix, , and N. V. Thakor, "Adaptive filter for event-related bioelectric signals using an impulse correlated reference input: Comparison with signal averaging techniques," IEEE Trans. Biomed. Eng., vol. 39, no. 10, pp. 1032–1044, 1992.

[28] N. V. Thakor, X. Guo, C. Vaz, P. Laguna, R. Jané, P. Caminal, H. Rix, and D. Hanley, "Orthonormal (Fourier and Walsh) models of time-varying evoked potentials in neurological injury," IEEE Trans. Biomed. Eng, vol. 40, no. 3, pp. 213–221, 1993.

[29] P. Clarkson and P. White, "Simplified analysis of the LMS adaptive filter using a transfer function approximation," IEEE Trans. on Acous. Speech Sig. Proc., vol. 35, pp. 987–993, Jul 1987.

[30] C. Vaz, X. Kong, and N. V. Thakor, "An adaptive estimation of periodic signals using a Fourier linear combiner," IEEE Trans. on Signal Processing, vol. 42, pp. 1–10, Jan 1994.

[31] S. Olmos and P. Laguna, "Steady-state MSE convergence analysis in LMS adaptive filters with deterministic reference inputs for biomedical signals," IEEE Trans. on Signal Proc., 1999. Submitted.

[32] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," IEEE Trans. Acoust., Speech, Signal Processing, vol. 33, pp. 222–230, 1985.

[33] S. Olmos, Contributions to ECG data compression with orthogonal transforms. PhD thesis, Univ. Politec. of Catalonia, Barcelona (Spain), 1998. In Spanish.

[34] L. Horowitz and K. D. Senne, "Performance advantage of complex LMS for controlling narrow-band adaptive arrays," IEEE Trans. ASSP, vol. 29, no. 3, pp. 722–736, 1981.

[35] B. Fisher and N. J. Bershad, "The complex LMS adaptive algorithm — transient weight mean and covariance with application to the ale," IEEE Trans. ASSP, vol. 31, pp. 34–44, 1983.

[36] S. C. Douglas and W. Pan, "Exact expectation analysis of the LMS adaptive filter," IEEE Trans. on Signal Proc., vol. 43, pp. 2863–2872, Dec 1995.

[37] A. K. Barros and N. Ohnishi, "MSE behavior of biomedical event-related filters," IEEE Trans. on Biomed. Eng., vol. 44, pp. 848–855, Sep 1997.

[38] B. Widrow, J. McCool, M. G. Larimore, and R. Johnson, "Stationary and nonstationary learning characteristics of the LMS adaptive filter," IEEE Proc., vol. 64, pp. 1151–1162, Aug 1976.

[39] W. Gardner, "Learning characteristics of stochastic-gradient-descent algorithms: a general study, analysis and critique," Signal Proc., pp. 113–133, 1984.

[40] O. Macchi and E. Eweda, "Convergence analysis of self-adaptive equalizers," IEEE Trans. on Information Theory, vol. IT-30, Special Issue on Linear Adaptive Filtering, pp. 161–176, 1984.

[41] M. Hajivandi and W. Gardner, "Measures of tracking performance for the LMS algorithm," IEEE Trans. Acous. Sig. Proc., vol. 38, no. 11, pp. 1953 – 1958, 1990.

[42] O. Macchi, Adaptive Processing: The Least Mean Square Approach with applications in transmission. New York, US: John Wiley & Sons Ltd., 1995.

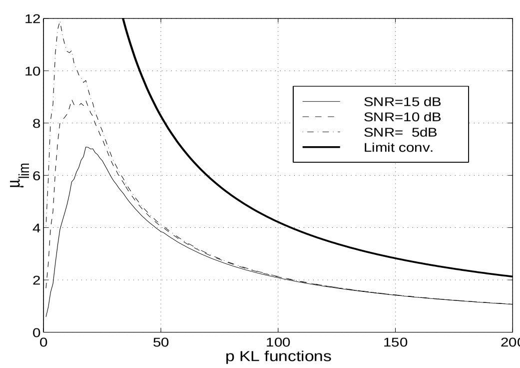
[43] O. Macchi, The theory of adaptive filtering in a random time-varying environment, pp. 171–206. London: A. Figueiras-Vidal, 1996.

[44] N. J. Bershad, "On the optimum gain parameter in LMS adaptation," IEEE Trans. Acoust., Speech, Signal Processing, vol. 35, pp. 1065–1068, 1987.

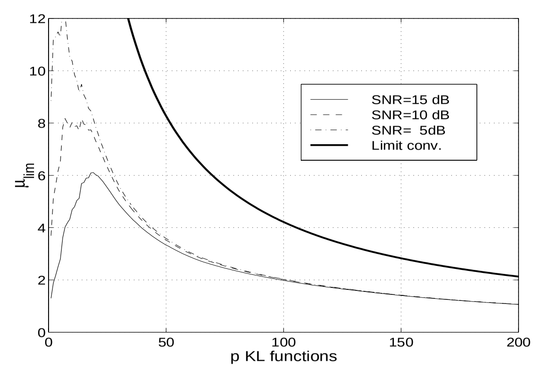
[45] P. Laguna, R. Jané, E. Masgrau, and P. Caminal, "The adaptive linear combiner with a periodic-impulse reference input as a linear comb filter," Signal Proc., vol. 48, pp. 193–203, 1996.

[46] S. J. Elliott and P. Darlington, "Adaptive cancellation of periodic, synchronously sampled interference," IEEE Trans. on Acoust. Sig. Proc., vol. 33, pp. 715–717, Jun 1985.

The *em* and *ma* noise records are non-stationary signals and the results will differ respect to the obtained for white noise. If we want to make an equivalent study for these kinds of noise we could consider the mean representation of the noise in the KL domain. Assuming this hypothesis, the μ_{lim} for the *em* and *ma* noises are presented in figures 19(a) and 19(b) respectively. It can be seen that μ_{lim} for physiological types of noise are higher than for white noise. The reason for this result is that ECG signal and physiological noise spectra are more overlapped between them than white noise (see figure 5). In this situation, the direct estimation will produce very noisy reconstructions while the LMS will be able to attenuate the uncorrelated noise. The value of μ for getting same performance is not restrictive in this case and it takes higher values. It can be noted that for very low values of signal to noise ratio the value of μ_{lim} can be higher than the limit of convergence (**??**). In this cases the LMS algorithm should not be used with higher values than the limit of convergence because its convergence is not guaranteed.



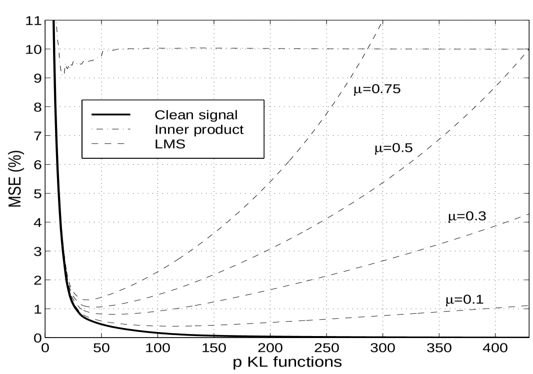
(a) Em noise



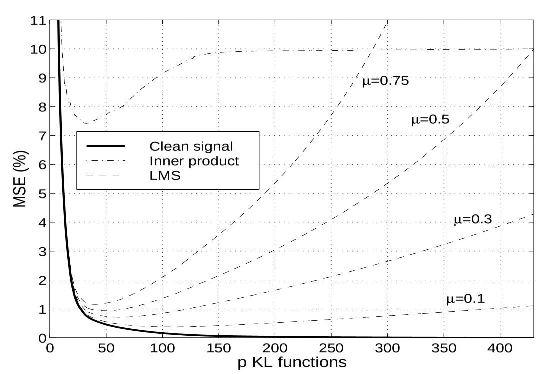
(b) Ma noise

Figure 19: *Values of μ_{lim} for several kinds of noise.*

We can calculate the value of MSE_p from expressions (16) and (3) using the eigenvalues of the covariance matrix of ECG signals from the training set and the mean of KL coefficients of *em* and *ma* noise. These values are represented in figure 20. We can see that MSE_p values for inner product are much higher than for white noise (see figure 10(a)) because both signal and noise spectra are more overlapped. In contrast, the LMS algorithm can attenuate uncorrelated noise with ECG signal and gets closer values to white noise.



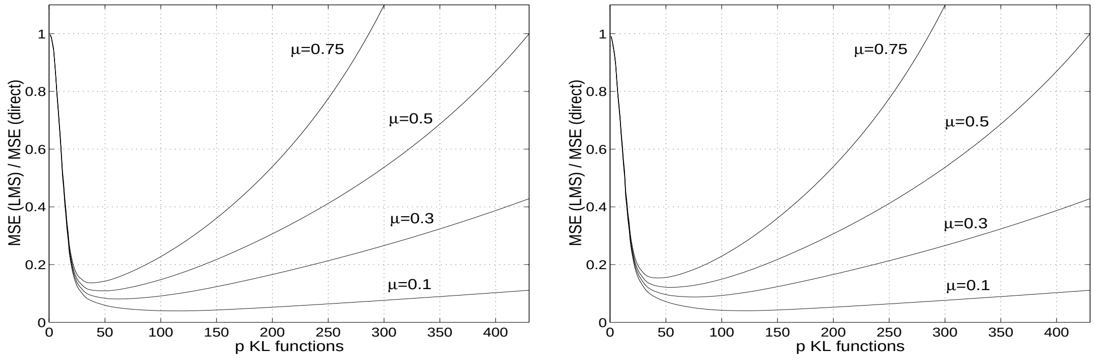
(a) Em noise with SNR=10 dB



(b) Ma noise SNR=10 dB

Figure 20: *MSE_p for inner product and LMS of signals in physiological noise.*

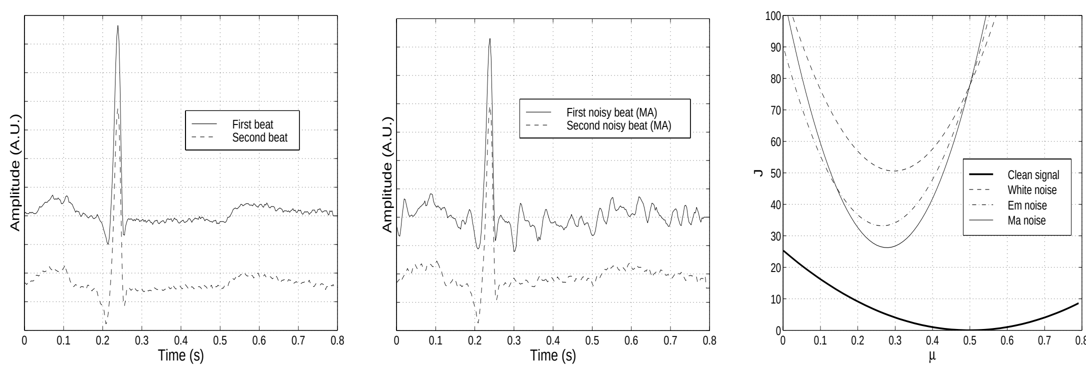
Finally we compare the MSE_p performance for LMS and inner product for the two physiological kinds of noise with a value of SNR=10 dB in figure 21. It can be seen that the maximum MSE improvement of LMS versus inner product is higher now than in white noise case and it occurs at lower values of functions p.



(a) Em noise with SNR=10 dB *(b) Ma noise SNR=10 dB*

Figure 21: *Values of* $MSE_p^{LMS/direct}$ *for em and ma noise with SNR=10 dB.*

We show in Fig. 22 one example with the values of the cost function J *versus* μ for a case of small change between two normal beats of record 100 from MIT-BIH Arrythmia database. Both clean signal and added noise cases are considered using the three kinds of noise (white noise, *em* and *ma*) with a value of SNR=10 dB. For the clean signal case, the optimum value of μ_{opt} is 0.5 (as it was announced before). When noise is added to the signal with SNR=10 dB, the level of noise is larger than the signal change from the first beat to the second one, and so the optimum values of μ_{opt} will be lower than 0.5 (similar case than subfigure 12(a)). The mean square error obtained by inner product with a complete expansion $J_N^{direct} = \mathbf{n}^T\mathbf{n}$ is the same than obtained by the LMS with μ=0.5, that is equal for the three kinds of noise because all of them have the same value of SNR.



(a) Normal beats. *(b) Normal beats with ma noise SNR=10 dB.* *(c) Values of J for both clean and noisy signal.*

Figure 22: *Example of J versus* μ *for a transition of two normal beats from record 100 with simulated noise.*

If we combine both situations, we can consider a more general initialization of the weight vector of LMS for non-stationary noisy ECG signals as

$$\mathbf{W}[0] = \alpha\mathbf{W}' + (1-\alpha)\mathbf{M}\,\mathbf{d} \tag{35}$$

where $\mathbf{W}'$ is the final weight vector of previous signal occurrence and α is a weighting factor that gives relative importance to noise or dynamic changes in the signal. The weighting factor α can be selected *a priori* if information of noise level or non-stationarity is available. For example, (decir ejemplos ... Nota
quizas en HRECG (poco ruido y estacionario por paciente quieto) o pruebas de esfuerzo (super ruidoso y noestacionario por cambios por movimientos) ambos factores tengan el mismo peso... ¿piensas algo a

priori?)

We can find the optimum values of μ and α that minimize the cost function J_N. Applying the value of the weight vector at the end occurrence time (28) ant the weight vector initialization in (35) we obtain

$$\mathbf{W}[N] = (1-2\,\mu)\,\alpha\,(\mathbf{W}' - \mathbf{M}\,\mathbf{d}) + \mathbf{M}\,\mathbf{d} = \beta\,(\mathbf{W}' - \mathbf{M}\,\mathbf{d}) + \mathbf{M}\,\mathbf{d}\,. \tag{36}$$

Calculating the cost functions and forcing zero at its first derivative $\frac{\partial J_N}{\partial \beta}$ we obtain

$$\beta_{opt} = (1-2\,\mu_{opt})\,\alpha_{opt} = \frac{(\mathbf{W}' - \mathbf{M}\,\mathbf{d})^T\mathbf{M}\,(\mathbf{s}-\mathbf{d})}{(\mathbf{W}' - \mathbf{M}\,\mathbf{d})^T(\mathbf{W}' - \mathbf{M}\,\mathbf{d})}\,. \tag{37}$$

This expression is similar than (30), where the weight vector is initialized with (35). This value correspond to a minimum point because the second derivative can be easily demonstrated that is positive

$$\frac{\partial^2 J_N}{\partial \beta^2} = 8\,(\mathbf{W}' - \mathbf{M}\,\mathbf{d})^T(\mathbf{W}' - \mathbf{M}\,\mathbf{d})\,. \tag{38}$$

8.1 Comparison of inner product *versus* the LMS algorithm

One of the objectives of this manuscript is to compare the performance of two ways for estimating the coefficients of orthogonal expansions of noisy ECG signals. Stationary signal case was studied in section 5. Now we will make an approximated analysis for non-stationary ECG signals.

We would like to have a criteria for selecting the step-size μ that gets same performance than inner product (μ_{lim}). The performace can be measured with the cost functions J defined in equation (29). The inner product coefficients are estimated as the projection of the observed noisy signal $\mathbf{d}$ onto the subspace generated by the p basis functions $\mathbf{W}^{direct} = \mathbf{M}^T\mathbf{d}$. The mean square error for inner product J_p^{direct} will be

$$\begin{aligned} J_p^{direct} = & \quad (\mathbf{s}-\mathbf{y}')^T(\mathbf{s}-\mathbf{y}') = (\mathbf{s}-\mathbf{M}^T\mathbf{M}\mathbf{d})^T(\mathbf{s}-\mathbf{M}^T\mathbf{M}\mathbf{d}) \\ = & \quad \left[(\mathbf{I}-\mathbf{M}^T\mathbf{M})\mathbf{s} - \mathbf{M}^T\mathbf{M}\mathbf{n}\right]^T\left[(\mathbf{I}-\mathbf{M}^T\mathbf{M})\mathbf{s} - \mathbf{M}^T\mathbf{M}\mathbf{n}\right] \end{aligned} \tag{39}$$

When all basis functions are used ($p\!=\!N$), we will have $\mathbf{M}^T\mathbf{M} = \mathbf{I}$ and $J_N^{direct} = \mathbf{n}^T\mathbf{n}$.

A similar expression for mean square error J_p can be obtained for the LMS algorithm

$$J_p^{LMS} = (\mathbf{s}-\mathbf{y}')^T(\mathbf{s}-\mathbf{y}') = (\mathbf{s}-\mathbf{M}^T\mathbf{W}[N])^T(\mathbf{s}-\mathbf{M}^T\mathbf{W}[N]) \tag{40}$$

where $\mathbf{W}[N]$ is given in (26). J_p^{LMS} will depend on the step-size μ, the noise present in the observed signal $\mathbf{n}\!=\!\mathbf{d}\!-\!\mathbf{s}$ and the initial weight vector $\mathbf{W}[0]$. The last dependence on $\mathbf{W}[0]$ is the difference from the stationary case, because now time adaptation is finite in order to track dynamic changes of the signal.

We could find a value of $\mu = \mu_{lim}$ that achieves $J_p^{LMS} = J_p^{direct}$ for non-stationary signals, that is

$$\mu_{lim} = \frac{1}{2}\frac{(\mathbf{W}[0]-\mathbf{M}\,\mathbf{d})^T(\mathbf{W}[0]-\mathbf{M}\,\mathbf{s}) \pm \sqrt{\text{expresion que falta}}}{(\mathbf{W}[0]-\mathbf{M}\,\mathbf{d})^T(\mathbf{W}[0]-\mathbf{M}\,\mathbf{d})} \tag{41}$$

But in this case the expression of μ_{lim} is not very useful because now we can not say that J_p^{LMS} is lower than J_p^{direct} for $\mu < \mu_{lim}$ because of optimum weight vector is time variant due to the beat-to-beat morphology changes. We shown in Fig. **??** the value of μ_{lim} for some normal beats of record 100 with SNR=10 dB of white noise. Falta

The second approach for using non-complete expansions (a bit different from the first one) can be interpreted as an approximation for low values of μ of equation (27), reducing only to linear terms $\mathbf{A}_1$ and $\mathbf{B}_1$. This is equivalent to apply the impulse correlated filter (AICF) [27, 45] in order to remove non-correlated noise with the recurrent signal, and afterwards making the truncation of p coefficients, equivalent to apply the inner product to the output of the adaptive filter. This approach can be analyzed in more detail than the first one, because we can use similar expressions than for complete expansion case (28)-(30), but now the matrix of orthogonal basis functions $\mathbf{M}$ will have dimensions $(p \times N)$, and all the vector norms and inner products will be evaluated in the signal subspace spanned by the reduced number of functions p. In this case the misadjustment will be larger than in the first approach, because all the weights are used in the adaptation with the LMS algorithm

---

When we compress noisy signals high values of MSE can be obtained due to the presence of noise that do not represent real error between original clean signal and reconstruction. Two approaches have been considered for estimating the coefficients: classical inner product and adaptive estimation with the LMS algorithm. Inner product has a higher convergence rate, but the LMS algorithm achieves greater noise reduction. Three different types of noise have been considered: white noise, electrode motion and motion artifact.

Steady state performance of both approaches have been compared. Analytical expressions for MSE_p and SNR_p are derived. The limit value of the LMS step-size μ_{lim} is obtained to get same performance than inner product in stationary signals. If a value of $\mu < \mu_{lim}$ is selected, LMS gets lower values of MSE. The maximum achievable improvement in the MSE of LMS *versus* inner product is achieved when the number of functions is low. The improvement reconstruction MSE of LMS versus inner product is higher for physiological types of noise because in this case signal and noise "KL-spectra" are more overlapped. So, in actual ECG records where physiological noise is important the improvement of LMS versus inner product is significant and lead us to consider it as the appropriate tool to compress noisy ECG signals.

In general, an event-related signal can be considered as an stochastic process which can be decomposed into an invariant deterministic signal time-locked to a stimulus, and an additive uncorrelated noise with the signal.