# Block Adaptive Filters With Deterministic Reference Inputs for Event-Related Signals: BLMS and BRLS

Salvador Olmos, Member, IEEE, Leif Sörnmo, Member, IEEE, and Pablo Laguna, Member, IEEE

Abstract—Adaptive estimation of the linear coefficient vector in truncated expansions is considered for the purpose of modeling noisy, recurrent signals. Two different criteria are studied for block-wise processing of the signal: the mean square error (MSE) and the least squares (LS) error. The block LMS (BLMS) algorithm, being the solution of the steepest descent strategy for minimizing the MSE, is shown to be steady-state unbiased and with a lower variance than the LMS algorithm. It is demonstrated that BLMS is equivalent to an exponential averager in the subspace spanned by the truncated set of basis functions. The block recursive least squares (BRLS) solution is shown to be equivalent to the BLMS algorithm with a decreasing step size. The BRLS is unbiased at any occurrence number of the signal and has the same steady-state variance as the BLMS but with a lower variance at the transient stage. The estimation methods can be interpreted in terms of linear, time-variant filtering. The performance of the methods is studied on an ECG signal, and the results show that the performance of the block algorithms is superior to that of the LMS algorithm. In addition, measurements with clinical interest are found to be more robustly estimated in noisy signals.

*Index Terms*—Adaptive filters, deterministic input, event-related signal, orthogonal expansions.

## I. INTRODUCTION

T HE problem of noise reduction in recurrent signals is well studied and has traditionally been solved by ensemble averaging or by one of its many variations. The time reference of each occurrence is often synchronized to a known, external stimulus; however, in certain signals, the time reference is difficult to observe, and therefore, a fiducial point needs to be established for each occurrence by some kind of estimation procedure. A major disadvantage with ensemble averaging is that efficient noise reduction is typically achieved at the expense of using a large number of occurrences for averaging. In order to better track short-term changes in morphology of a recurrent signal while still achieving a reduction of the noise level, it is desirable to develop methods that incorporate *a priori* information on possible morphologies. More recently, modeling of each

Manuscript received May 22, 2001; revised January 29, 2002. This work was supported by P075/2001 from CONSI+D-DGA, TIC2001-2167-C02-02 from CICYT, and a Post-Doctoral grant to S. Olmos PF 0021479421 from MECD (Spain). The associate editor coordinating the review of this paper and approving it for publication was Dr. Dennis R. Morgan.

S. Olmos was with the Department of Electroscience, Lund University, Lund, Sweden. He is now with the Electronics Engineering and Communication Department, Zaragoza University, Zaragoza, Spain.

L. Sörnmo is with the Department of Electroscience, Lund University, Lund, Sweden.

P. Laguna is with the Electronics Engineering and Communication Department, Zaragoza University, Zaragoza, Spain.

Publisher Item Identifier S 1053-587X(02)03286-5.

occurrence as a signal that is well described by a truncated linear expansion of orthonormal basis functions has been studied.

The coefficients of the linear expansion can be estimated using different approaches. In many situations, the mean square error (MSE) between the observed and the modeled signals represents a suitable cost function since it is related to signal energy. The optimal coefficients are commonly referred to as the Wiener solution and are determined on an individual occurrence basis, thus constituting memoryless estimation in the sense that none of the previous occurrences are included in the current coefficient estimation. By introducing memory in the estimator, the variance of the coefficient estimates can be considerably reduced while the capability of tracking signal changes in a noisy environment is still available.

Several papers have been presented in the area of biomedical signal processing where an adaptive solution based on the LMS algorithm is suggested; see, e.g., [1]-[4]. The reference inputs to the LMS algorithm are deterministic functions and are defined by a periodically extended, truncated set of orthonormal basis functions. In these papers, the LMS algorithm operates on an "instantaneous" basis such that the weight vector is updated every new sample within the occurrence, based on an instantaneous gradient estimate. In a recent study, however, a steadystate convergence analysis for the LMS algorithm with deterministic reference inputs showed that the steady-state weight vector is biased, and thus, the adaptive estimate does not approach the Wiener solution [5]. To handle this drawback, we consider another strategy for estimating the coefficients of the linear expansion, namely, the block LMS (BLMS) algorithm, in which the coefficient vector is updated only once every occurrence based on a block gradient estimation. The BLMS algorithm has already been proposed in the case of random reference inputs and has, when the input is stationary, the same steadystate misadjustment and convergence speed as the LMS algorithm [6]-[11]. A major advantage of the block, or the transform domain, LMS algorithm is that the input signals are approximately uncorrelated (or orthogonal in a more general sense). To the best of our knowledge, block adaptation has not been considered previously within the context of deterministic reference input signals.

An alternative solution to the MSE criterion is to use the least squares (LS) criterion, where no assumptions on the statistics of the input signals are invoked. This approach gives the best linear unbiased estimate, assuming that the measurement error process is white and zero mean [11], [12]. The LS solution can be calculated recursively by means of the RLS algorithm, which has an initially faster convergence speed than the LMS algorithm. When the inputs are random, the improvement in performance is achieved at the expense of a large increase in computational complexity. In contrast, when deterministic and orthonormal inputs are used, the complexity is about the same as for the LMS algorithm.

The selection of orthonormal basis functions is, of course, dependent on the application of interest. In the area of biomedical signal processing, the analysis of evoked potentials in the electroencephalogram has been based on impulse functions [13], [14], sine and cosine functions [1], [15], complex exponentials [16], [17] and Walsh functions [18], whereas the QRST complexes of the electrocardiogram (ECG) have been modeled by Hermite functions [4], [19] or basis functions that result from the Karhunen–Loève (KL) expansion [20]. The purpose of the basis function description is not only noise reduction, as mentioned above, but may also be applied to data compression [21], [22], feature extraction [19], and monitoring [20].

The paper is organized as follows. The Wiener solution for the linear expansion coefficient is briefly reviewed in Section II as is its memoryless estimation [the inner product (IP)]. The BLMS algorithm with deterministic reference inputs is then presented in Section III. Section IV presents the equivalent transfer function of the BLMS algorithm and its relation to the IP and LMS estimators. Section V describes the estimation of linear coefficients using a different cost function-the weighted least squares (LS) error criterion. The resulting optimal solution defines the block recursive least squares (BRLS) algorithm; its relationship to the optimal MSE solution is established in Section V. A comparative performance analysis of the four estimators (IP, LMS, BLMS, and BRLS) is presented in Section VI in terms of bias, variance, and definitions of signal-to-noise ratio (SNR). Finally, the performance of the estimation methods is illustrated in Section VII using KL basis functions to characterize the ST-T segment of the ECG.

# II. MSE ESTIMATION OF EXPANSION COEFFICIENTS

An observed event-related signal  $\mathbf{d}_k$  can be represented as a  $N \times 1$  vector, where the subindex k denotes the occurrence number. When a truncated orthogonal expansion is used, the estimated signal  $\mathbf{y}_k$  is a linear combination of p basis functions

$$\mathbf{y}_k = \mathbf{T}\mathbf{w}_k \tag{1}$$

where **T** is a  $N \times p$  matrix whose columns are the basis functions, and  $\mathbf{w}_k$  is the  $p \times 1$  coefficient vector with  $p \leq N$ . One approach to finding the optimal linear coefficient vector  $\mathbf{w}_k^o$  is to minimize the cost function defined by the mean square error between  $\mathbf{d}_k$  and  $\mathbf{y}_k$ 

$$J_k = E\{(\mathbf{d}_k - \mathbf{T}\mathbf{w}_k)^T (\mathbf{d}_k - \mathbf{T}\mathbf{w}_k)\}.$$
 (2)

Applying differentiation, we obtain

$$\frac{\partial J_k}{\partial \mathbf{w}_k} = 2\mathbf{T}^T \mathbf{T} \mathbf{w}_k - 2\mathbf{T}^T E\{\mathbf{d}_k\} = \mathbf{0}.$$
 (3)

Since we have that  $\mathbf{T}^T \mathbf{T} = \mathbf{I}_p$  for any subset of orthogonal basis functions,<sup>1</sup> the Wiener solution for the linear coefficient vector is

$$\mathbf{w}_k^o = \mathbf{T}^T E\{\mathbf{d}_k\}.$$
 (4)

<sup>1</sup>It should be noted that for truncated expansions  $\mathbf{TT}^T \neq \mathbf{I}_N$ ; equality only holds for complete expansions.

This solution can be easily understood because the optimal signal description in the transformed domain is the projection of the expected value of the observed signal. This solution corresponds to a minimum because the Hessian matrix is positive definite.

The observed signal  $d_k$  is commonly contaminated by noise. Assuming an additive-noise model, each signal occurrence  $d_k$  can be decomposed as

$$\mathbf{d}_k = \mathbf{s}_k + \mathbf{n}_k \tag{5}$$

where  $\mathbf{s}_k$  is a deterministic signal, and  $\mathbf{n}_k$  is zero-mean random noise. The Wiener solution (4) for this signal model is

$$\mathbf{w}_k^o = \mathbf{T}^T \mathbf{s}_k \tag{6}$$

i.e., the projection of the deterministic signal in the transformed domain. Hence, the modeled signal is given by

3

$$\boldsymbol{v}_{k}^{o} = \mathbf{T}\mathbf{T}^{T}\mathbf{s}_{k} \tag{7}$$

which can be interpreted as the output of a linear, time-variant filter defined by the matrix  $\mathbf{TT}^{T}$  with the clean signal  $\mathbf{s}_{k}$  as input [23]. The cost function at the optimum will be

$$J_k^o = E\left\{\mathbf{n}_k^T \mathbf{n}_k\right\} + \mathbf{s}_k^T (\mathbf{I} - \mathbf{T}\mathbf{T}^T)\mathbf{s}_k \tag{8}$$

which is the sum of the noise energy and the truncated signal error.

Since the clean signal  $\mathbf{s}_k$  is unavailable,  $\mathbf{w}_k^o$  needs to be estimated according to (4) from the observed signal  $\mathbf{d}_k$ . A very simple way is to approximate  $E\{\mathbf{d}_k\} \simeq \mathbf{d}_k$  in (4), implying that the linear coefficient vector is estimated by

$$\mathbf{w}_k^{\mathrm{IP}} = \mathbf{T}^T \mathbf{d}_k \tag{9}$$

where IP denotes the inner product between each basis function and the observed signal. This kind of estimation is memoryless since only information from the *k*th occurrence is used to estimate  $E\{\mathbf{d}_k\}$ , and as a result, sudden changes in signal shape can be tracked. On the other hand,  $\mathbf{w}_k^{\mathrm{IP}}$  will be sensitive to the presence of noise.

## **III. BLMS ESTIMATION**

One way to reduce the influence of noise is to include adaptive algorithms in the coefficient estimation since this type of estimators has memory of previous occurrences. When the deterministic signal is repetitive with slow occurrence-to-occurrence shape changes, the amount of noise can be reduced at the expense of a slower convergence. The tradeoff between convergence speed and SNR improvement is controlled by the memory used in the estimation.

The structure of the vector-based adaptive filter is shown in Fig. 1. The primary input  $\mathbf{d}_k = \mathbf{s}_k + \mathbf{n}_k$  consists of successive concatenated signal occurrences. For the steady-state analysis of the algorithm, we assume that the deterministic signal  $\mathbf{s}_k$  remains unchanged during all occurrences, i.e.,  $\mathbf{s}_k = \mathbf{s}$ . In practice,  $\mathbf{s}_k$  will be occurrence variant, and the algorithm will track signal changes in a finite adaptation time. At each signal occurrence, the adaptive system estimates the amount of each reference input (columns of the  $N \times p$  matrix **T**) present in the input signal  $\mathbf{d}_k$ .

In order to minimize the cost function (2), the optimal weight vector can be estimated using an iterative algorithm based on



Fig. 1. Adaptive block-wise estimation of truncated expansions.

the steepest descent strategy. The weight vector is updated once every signal occurrence according to

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu \frac{\partial J_k}{\partial \mathbf{w}_k} \tag{10}$$

where  $\mu$  is the step size, which controls stability and convergence speed of the algorithm. Using the definition of  $\mathbf{e}_k = \mathbf{d}_k - \mathbf{y}_k = \mathbf{d}_k - \mathbf{T}\mathbf{w}_k$  and decomposing the weight vector as the sum of the optimum value<sup>2</sup> and the weight error vector

$$\mathbf{w}_k = \mathbf{w}^{\mathrm{o}} + \mathbf{v}_k \tag{11}$$

we can write

$$J_{k} = E\left\{ (\mathbf{d}_{k} - \mathbf{T}(\mathbf{w}^{o} + \mathbf{v}_{k}))^{T} (\mathbf{d}_{k} - \mathbf{T}(\mathbf{w}^{o} + \mathbf{v}_{k})) \right\}$$
$$= E\left\{ (\mathbf{e}_{k}^{o} - \mathbf{T}\mathbf{v}_{k})^{T} (\mathbf{e}_{k}^{o} - \mathbf{T}\mathbf{v}_{k}) \right\}$$
(12)

where  $\mathbf{e}_{k}^{o} = \mathbf{d}_{k} - \mathbf{T}\mathbf{w}^{o}$  denotes the error signal obtained with the optimum weight solution. Using the Wiener solution (6), the minimum error signal can be written as

$$\mathbf{e}_{k}^{\mathrm{o}} = \mathbf{s} + \mathbf{n}_{k} - \mathbf{T}\mathbf{w}^{\mathrm{o}} = (\mathbf{I} - \mathbf{T}\mathbf{T}^{T})\mathbf{s} + \mathbf{n}_{k}$$
(13)

and thus, two independent sources can be considered in the error signal: the error due to truncation and the observed noise.

The cost function (12) can equally be written as

$$J_{k} = E\left\{\left(\mathbf{e}_{k}^{\mathrm{o}}\right)^{T}\mathbf{e}_{k}^{\mathrm{o}}\right\} + E\left\{\mathbf{v}_{k}^{T}\mathbf{v}_{k}\right\} - 2E\left\{\mathbf{v}_{k}^{T}\mathbf{T}^{T}\mathbf{e}^{\mathrm{o}}\right\} \quad (14)$$

where the quadratic dependence on the weight error vector is evident. At the optimal solution, the signal truncation error is orthogonal to the reference inputs, and

$$\mathbf{T}^{T}\mathbf{e}_{k}^{\mathrm{o}} = \mathbf{T}^{T}(\mathbf{I} - \mathbf{T}\mathbf{T}^{T})\mathbf{s} + \mathbf{T}^{T}\mathbf{n}_{k} = \mathbf{T}^{T}\mathbf{n}_{k}.$$
 (15)

Hence, the cost function  $J_k$  can be written as

$$I_{k} = E\left\{\left(\mathbf{e}_{k}^{\mathrm{o}}\right)^{T}\mathbf{e}_{k}^{\mathrm{o}}\right\} + E\left\{\mathbf{v}_{k}^{T}\mathbf{v}_{k}\right\} - 2E\left\{\mathbf{v}_{k}^{T}\mathbf{T}^{T}\mathbf{n}_{k}\right\} \quad (16)$$

where the first term is called the minimum error  $J_k^o$ , and the sum of the two following terms is referred to as the excess of MSE  $J_k^{\text{ex}}$ .

The gradient can be calculated as

$$\frac{\partial J_k}{\partial \mathbf{v}_k} = 2E\left\{\mathbf{v}_k\right\} - 2\mathbf{T}^T E\left\{\mathbf{n}_k\right\}.$$
(17)

When only the kth signal occurrence is available, a simple gradient estimation considers

$$\frac{\partial J_k}{\partial \mathbf{v}_k} \simeq 2\mathbf{v}_k - 2\mathbf{T}^T \mathbf{n}_k. \tag{18}$$

<sup>2</sup>The dependence of  $\mathbf{w}^{\circ}$  on the occurrence number k is omitted because for the steady-state analysis, it is assumed that the deterministic signal s is invariant for all signal occurrences.

Accordingly, the steepest descent algorithm (10) can be written as

$$\mathbf{v}_{k+1} = (1 - 2\mu)\mathbf{v}_k + 2\mu\mathbf{T}^T\mathbf{n}_k.$$
 (19)

This algorithm is named block LMS (BLMS) because it is equivalent to the LMS algorithm but with blockwise gradient estimation. In a similar way, the update equation for the weight vector can be easily obtained as

$$\mathbf{w}_{k+1} = (1 - 2\mu)\mathbf{w}_k + 2\mu\mathbf{T}^T\mathbf{d}_k.$$
 (20)

In other words, the BLMS is equivalent to exponential averaging in the subspace spanned by **T**. Note that the BLMS is equivalent to IP when  $\mu = 0.5$ .

We will now consider the bias and variance of the BLMS algorithm since these quantities are needed for comparison with other estimation methods. The weight error vector at the kth occurrence can be written as

$$\mathbf{v}_{k} = (1 - 2\mu)^{k} \mathbf{v}_{0} + 2\mu \sum_{j=0}^{k-1} (1 - 2\mu)^{k-j-1} \mathbf{T}^{T} \mathbf{n}_{j}.$$
 (21)

The first term is clearly a transient, which, for  $0 < \mu < 1$ , will vanish after a sufficiently large number of signal occurrences. Therefore, at steady state, only the second term in (21) will be important. Taking  $\lim_{k\to\infty}$  of the expected value, we obtain

$$\lim_{k \to \infty} E\{\mathbf{v}_k\} = E\{\mathbf{v}_\infty\} = \lim_{k \to \infty} 2\mu$$
$$\times \sum_{j=0}^{k-1} (1 - 2\mu)^{k-j-1} \mathbf{T}^T E\{\mathbf{n}_j\} = \mathbf{0} \quad (22)$$

since the noise  $n_j$  is assumed to be zero mean. Accordingly, the steady-state weight vector  $w_{\infty}$  is an unbiased estimator of the Wiener solution (6).

In order to analyze the steady-state variance, we need to quantify the excess MSE  $J_k^{\text{ex}}$  in the cost function (16). The energy of the weight error vector  $E\{\mathbf{v}_k^T\mathbf{v}_k\}$  can be calculated as

$$E\left\{\mathbf{v}_{k}^{T}\mathbf{v}_{k}\right\} = (1-2\mu)^{2k}E\left\{\mathbf{v}_{0}^{T}\mathbf{v}_{0}\right\} + 2(1-2\mu)^{k}2\mu$$

$$\times E\left\{\mathbf{v}_{0}^{T}\left(\sum_{j=0}^{k-1}(1-2\mu)^{k-j-1}\mathbf{T}^{T}\mathbf{n}_{j}\right)\right\}$$

$$+ 4\mu^{2}E\left\{\left(\sum_{j=0}^{k-1}(1-2\mu)^{k-j-1}\mathbf{n}_{j}^{T}\mathbf{T}\right)$$

$$\times \left(\sum_{l=0}^{k-1}\mathbf{T}^{T}\mathbf{n}_{l}(1-2\mu)^{k-l-1}\right)\right\}.$$
(23)

At steady-state, the first two terms will be null if appropriate values of the step-size  $\mu$  are selected, and then

$$E\left\{\mathbf{v}_{\infty}^{T}\mathbf{v}_{\infty}\right\} = E\left\{\operatorname{tr}\left\{\mathbf{v}_{\infty}\mathbf{v}_{\infty}^{T}\right\}\right\}$$
$$= \lim_{k \to \infty} 4\mu^{2}\operatorname{tr}\left\{\sum_{j=0}^{k-1}\sum_{l=0}^{k-1}(1-2\mu)^{k-l-1}\right\}$$
$$\times \mathbf{T}^{T}E\{\mathbf{n}_{l}\mathbf{n}_{j}^{T}\}\mathbf{T}(1-2\mu)^{k-j-1}\right\}. (24)$$

If the noise signal is assumed to be stationary with correlation function shorter than the gap between consecutive occurrences,<sup>3</sup> then

$$E\{\mathbf{n}_{l}\mathbf{n}_{j}^{T}\} = \delta_{lj}\mathbf{N}$$
<sup>(25)</sup>

where N is the  $N \times N$  noise covariance matrix. Accordingly, the steady-state weight error vector energy is

$$E\left\{\mathbf{v}_{\infty}^{T}\mathbf{v}_{\infty}\right\} = \lim_{k \to \infty} 4\mu^{2} \sum_{j=0}^{k-1} (1-2\mu)^{2(k-j-1)} \operatorname{tr}\left\{\mathbf{T}^{T}\mathbf{N}\mathbf{T}\right\}$$
$$= \frac{\mu}{1-\mu} \operatorname{tr}\left\{\mathbf{T}^{T}\mathbf{N}\mathbf{T}\right\}.$$
(26)

For the particular case of white noise with variance  $\sigma^2$ , (26) is simplified to

$$E\left\{\mathbf{v}_{\infty}^{T}\mathbf{v}_{\infty}\right\} = \frac{\mu p\sigma^{2}}{1-\mu}.$$
(27)

In order to complete the evaluation of the cost function (16), the cross term  $E\{\mathbf{v}_k^T\mathbf{T}^T\mathbf{n}_k\}$  needs to be quantified as

$$E\left\{\mathbf{v}_{k}^{T}\mathbf{T}^{T}\mathbf{n}_{k}\right\} = (1-2\mu)^{k}E\left\{\mathbf{v}_{0}^{T}\mathbf{T}^{T}\mathbf{n}_{k}\right\}$$
$$+ \sum_{j=0}^{k-1} (1-2\mu)^{k-j-1}E\left\{\mathbf{n}_{j}^{T}\mathbf{T}\mathbf{T}^{T}\mathbf{n}_{k}\right\}. \quad (28)$$

The first term is again a transient, and it will be null at steady state. The second term is zero because

$$E\left\{\mathbf{n}_{j}^{T}\mathbf{T}\mathbf{T}^{T}\mathbf{n}_{k}\right\} = \operatorname{tr}\left\{\mathbf{T}^{T}E\left\{\mathbf{n}_{k}\mathbf{n}_{j}^{T}\right\}\mathbf{T}\right\} = 0, \quad k \neq j$$
(29)

which follows from (25). Summing up, the steady-state cost function is

$$J_{\infty}^{\text{BLMS}} = J_{\infty}^{\text{o}} + \frac{\mu}{1-\mu} \text{tr}\{\mathbf{T}^{T}\mathbf{NT}\}.$$
 (30)

In the case of complete expansions, we have  $tr{T^TNT} = tr{N}$ , which is equal to the noise energy. In the case of white noise and incomplete expansions

$$J_{\infty}^{\text{BLMS}} = J_{\infty}^{\text{o}} + \frac{\mu}{1-\mu} p \sigma^2.$$
(31)

It may be worthwhile to point out certain relationships to the LMS algorithm. In the case of the LMS algorithm,  $J_k^{\text{ex}}$  is composed of three terms [5, Eq. (12) and (13)], whereas for the BLMS algorithm, only two terms are present in (16) because the signal truncation error is orthogonal to the input basis functions **T**. Moreover, the LMS algorithm converges to a biased estimate in the case of truncated expansions [5], whereas the BLMS estimation is steady-state unbiased. When complete expansions p = N are used, it can be noted that  $J_{\infty}^{\text{BLMS}} = J_{\infty}^{\text{LMS}}$  [5]. This result agrees with the fact that the algorithms become identical when complete expansions are used (compare (20) with [5, Eq. (8)]).

## **IV. EQUIVALENT TRANSFER FUNCTION**

Truncated orthogonal expansions can be understood as linear time-variant filters. The equivalent instantaneous impulse and frequency responses were calculated in [23], where the linear coefficients were estimated using the IP and the LMS algorithm. This section will extend the analysis for the BLMS algorithm.

In the update equation of the BLMS algorithm (20), the term  $\mathbf{T}^T \mathbf{d}_k$  represents the IP estimation of  $\mathbf{w}^o$  using only information from the *k*th occurrence. The first term accounts for the estimation done at the previous occurrence. As a consequence, the BLMS algorithm can be understood as a transform-domain exponential averager. It is well known that exponential averaging is equivalent to a linear time-invariant filter whose transfer function is a comb filter. On the other hand, truncated orthogonal expansions estimated with inner product are equivalent to a linear time-variant filter [23]. Therefore, the combination of both systems is a linear time-variant filter.

An alternative demonstration can be done by looking at the reconstructed signal. A first-order finite difference equation is obtained by premultiplying both sides of (20) by  $\mathbf{T}$ 

$$\mathbf{y}_{k+1} = (1 - 2\mu)\mathbf{y}_k + 2\mu\mathbf{T}\mathbf{T}^T\mathbf{d}_k.$$
 (32)

In the case of complete expansions  $\mathbf{TT}^T = \mathbf{I}_N$ , the coefficients in (32) are scalar and time invariant. Accordingly, the system can be described with a single transfer function, which is a comb filter. The same transfer function was obtained for the LMS algorithm fed with a complete set of impulse functions in [14] and, more generally, with any complete set of orthogonal functions in [23].

When truncated orthogonal expansions p < N are considered, a coupled system of finite difference equations is obtained from (32) because  $\mathbf{TT}^T \neq \mathbf{I}_N$ . The equation system can be written in a scalar way as

$$y[kN+l] = (1-2\mu)y[(k-1)N+l] + 2\mu \sum_{m=0}^{N-1} r_{lm}d[(k-1)N+m] \quad l = 0, 1, \dots, N-1 \quad (33)$$

where the coefficients  $r_{lm}$  are the elements of the matrix  $\mathbf{TT}^T$ . Thus, a set of linear Nth-order finite difference equations with time-variant coefficients is obtained. If null initial conditions are used, a linear time-variant system can be defined from (33). The analysis of this equation system is simpler than the one obtained in [23, Eq. (14)] for the LMS algorithm.

One way to solve the equivalent time-variant impulse response is to observe the output response f[a, n] to impulse functions  $\delta[n - a]$  located at different time instants a with  $0 \le a \le N - 1$ . Then, the equivalent time-variant impulse response h[m, n] will accomplish

$$f[a,n] = \sum_{m=-\infty}^{\infty} h[m,n]\delta[n-a-m] = h[n-a,n].$$
 (34)

Let  $\delta_a = \begin{bmatrix} 0 \cdots 0 & 1 \\ 1 & 0 \cdots 0 \end{bmatrix}^T$  be the occurrence representation of the impulse input  $\delta[n - a]$ . The BLMS output at the *k*th occurrence  $\mathbf{f}_k$  follows the recursion

$$\begin{cases} \mathbf{f}_k = \mathbf{0}, & k = 0\\ \mathbf{f}_k = (1 - 2\mu)^{k-1} 2\mu \mathbf{T} \mathbf{T}^T \boldsymbol{\delta}_a, & k \ge 1 \end{cases}$$
(35)

<sup>&</sup>lt;sup>3</sup>The very-low frequency components of biomedical signals, e.g., baseline wander in the ECG, are usually removed in a preprocessing stage because they do not convey any valuable clinical information.

Step-size/Forgetting factor relationship

which is equal to the linear convolution of the FIR impulse response of the IP at time instant a,  $\mathbf{TT}^T \boldsymbol{\delta}_a$  and the comb filter with impulse response

$$\sum_{k=1}^{\infty} 2\mu (1-2\mu)^{k-1} \delta[n-kN].$$
 (36)

The same impulse response was obtained in [23] for the LMS algorithm when small values of step-size are used (quadratic terms of  $\mu$  were neglected).

# V. LEAST SQUARES COEFFICIENT VECTOR

An alternative solution to the MSE criterion is to use the least squares (LS) error criterion, where the cost function to be minimized depends on the observed signal in a deterministic way. The problem can be defined as finding the  $p \times 1$  coefficient vector w that minimizes the cost function

$$\mathcal{E}_k = \mathbf{e}_k^T \mathbf{e}_k = (\mathbf{d}_k - \mathbf{T}\mathbf{w})^T (\mathbf{d}_k - \mathbf{T}\mathbf{w}).$$
(37)

The least squares solution is given by [24]

$$\mathbf{w}_k^{\mathrm{LS}} = \mathbf{T}^T \mathbf{d}_k \tag{38}$$

which is identical to the solution obtained by using the inner product estimation in (9).

When several signal occurrences are jointly analyzed, several extensions of the cost function can be defined to take the information from previous signal occurrences into account. The classical approach is to include all the available past signal occurrences, possibly weighted by a forgetting factor to add tracking capability to the algorithm. Then, the cost function at the *k*th occurrence is the squared error weighted sum from the first occurrence to the current time. Therefore, the coefficient vector at the *k*th occurrence is chosen to minimize

$$\mathcal{E}_k = \sum_{i=0}^k \lambda^{k-i} (\mathbf{d}_i - \mathbf{T} \mathbf{w})^T (\mathbf{d}_i - \mathbf{T} \mathbf{w})$$
(39)

where the constant  $\lambda, 0 < \lambda \leq 1$  is the forgetting factor. Note that the coefficient vector is held constant during the observation interval. The coefficient vector obtained by minimizing (39) is denoted by  $\mathbf{w}_k^{\text{LS}}$  and provides the LS coefficient vector at the *k*th occurrence. A necessary condition for the optimum is

$$\frac{\partial \mathcal{E}_k}{\partial \mathbf{w}} = \sum_{i=0}^k \lambda^{k-i} (2\mathbf{w} - 2\mathbf{T}^T \mathbf{d}_i) = \mathbf{0}$$
(40)

and then

$$\mathbf{w}_{k+1}^{\mathrm{LS}} = \frac{1-\lambda}{1-\lambda^{k+1}} \mathbf{T}^T \sum_{i=0}^k \lambda^{k-i} \mathbf{d}_i.$$
 (41)

Every time a new occurrence is available, the cost function (39) needs to be minimized. Fortunately, the LS coefficient vector



Fig. 2. Relationship between the step size  $\mu_k$  in BLMS and the forgetting factor  $\lambda$  in BRLS.

can be estimated in a recursive way using a block RLS (BRLS) approach by rewriting (41) as

$$\mathbf{w}_{k+1}^{\text{BRLS}} = \lambda \frac{1 - \lambda^k}{1 - \lambda^{k+1}} \mathbf{w}_k^{\text{BRLS}} + \frac{1 - \lambda}{1 - \lambda^{k+1}} \mathbf{T}^T \mathbf{d}_k.$$
 (42)

The LS coefficient vector consists of updating the solution of the previous occurrence with the new data  $d_k$ . The difference from the BLMS is that the update coefficients are time variant, and therefore, the recursive LS solution (BRLS) can be understood as the BLMS algorithm in (20) with an occurrence-varying step size

$$\mu_k = \frac{1 - \lambda}{2(1 - \lambda^{k+1})}.$$
(43)

The equivalent occurrence-variant step size  $\mu_k$  of the BLMS algorithm is illustrated in Fig. 2 for several values of the forgetting factor  $\lambda$ . The convergence is fast at the first signal occurrences because the equivalent step-size is large (in particular,  $\mu_1 = (1/2)$  and is therefore equivalent to IP). For later occurrences, the step-size decreases, and finally, the steady-state value is  $\mu_{\infty} = (1 - \lambda)/2$ . The strategy of a decreasing step-size implies that larger step sizes at the beginning provide a faster approximation to the optimum, and later, smaller values of  $\mu$  are used to reduce the variance. A decreasing step size has heuristically been included in some variants of the LMS algorithm [25].

Although the least squares cost function (39) is deterministic, the presence of noise in the observed signal yields some deviation with respect to the ideal noise-free solution. Then, bias and variance could be used to quantify these deviations.

The weight error vector can be written as

$$\mathbf{v}_{k}^{\text{BRLS}} = \mathbf{w}_{k}^{\text{BRLS}} - \mathbf{T}^{T}\mathbf{s}$$
$$= \frac{1-\lambda}{1-\lambda^{k}}\mathbf{T}^{T}\sum_{i=0}^{k-1}\lambda^{k-i}\mathbf{n}_{i}.$$
(44)

Using the zero-mean noise assumption, the BRLS yields an unbiased estimate

$$E\left\{\mathbf{v}_{k}^{\mathrm{BRLS}}\right\} = \mathbf{0} \tag{45}$$

TABLE IWEIGHT ERROR VECTOR  $\mathbf{v}_k$  FOR DIFFERENT ESTIMATION METHODS

IP	$\mathbf{v}_k = \mathbf{T}^T \mathbf{n}_k$		
LMS	$\mathbf{v}[(k-1)N+j] = \mathbf{F}_{N+i,j}^{k}\mathbf{v}_{j} + 2\mu \sum_{m=0}^{kN+l-1} \mathbf{F}_{kN+j,m+1}\boldsymbol{\tau}_{m} (n[m] + c[m])$	$1\leq j\leq N$	
BLMS	$\mathbf{v}_{k} = (1 - 2\mu)^{k} \mathbf{v}_{0} + 2\mu \mathbf{T}^{T} \sum_{i=0}^{k-1} (1 - 2\mu)^{k-i-1} \mathbf{n}_{i}$		
BRLS	$\mathbf{v}_{k} = \frac{1-\lambda}{1-\lambda^{k}} \mathbf{T}^{T} \sum_{i=0}^{k-1} \lambda^{k-i} \mathbf{n}_{i}$		

for any occurrence index k. The steady-state variance can be calculated as

$$E\left\{\left(\mathbf{v}_{\infty}^{\text{BRLS}}\right)^{T}\mathbf{v}_{\infty}^{\text{BRLS}}\right\}$$
  
=  $\lim_{k \to \infty} E\left\{\operatorname{tr}\left\{\mathbf{v}_{k}^{\text{BRLS}}\left(\mathbf{v}_{k}^{\text{BRLS}}\right)^{T}\right\}\right\}$   
=  $(1 - \lambda)^{2}\lim_{k \to \infty}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\lambda^{2k-i-j}\operatorname{tr}\left\{\mathbf{T}^{T}E\left\{\mathbf{n}_{i}\mathbf{n}_{j}^{T}\right\}\mathbf{T}\right\}$   
=  $\frac{1 - \lambda}{1 + \lambda}\operatorname{tr}\{\mathbf{T}^{T}\mathbf{NT}\}$  (46)

where the noise assumption (25) is also used. In the white noise case,  $\mathbf{N} = \sigma^2 \mathbf{I}$ , and the variance can be written as

$$E\left\{\left(\mathbf{v}_{\infty}^{\mathrm{BRLS}}\right)^{T}\mathbf{v}_{\infty}^{\mathrm{BRLS}}\right\} = \frac{1-\lambda}{1+\lambda}p\sigma^{2}.$$
 (47)

If the forgetting factor is selected according to  $1 - \lambda = 2\mu$ , then BRLS and BLMS have the same steady-state variance

$$E\left\{\left(\mathbf{v}_{\infty}^{\mathrm{BRLS}}\right)^{T}\mathbf{v}_{\infty}^{\mathrm{BRLS}}\right\} = \frac{\mu}{1-\mu}\mathrm{tr}\{\mathbf{TNT}^{T}\}.$$
 (48)

Finally, we note that the BRLS can be described as a linear time-variant filter, similar to the BLMS algorithm. The first-order finite difference equation that characterizes the BRLS algorithm is obtained by premultiplying (42) with  $\mathbf{T}$ 

$$\mathbf{y}_{k+1}^{\text{BRLS}} = \lambda \frac{1 - \lambda^k}{1 - \lambda^{k+1}} \mathbf{y}_k^{\text{BRLS}} + \frac{1 - \lambda}{1 - \lambda^{k+1}} \mathbf{T} \mathbf{T}^T \mathbf{d}_k.$$
 (49)

This finite difference equation has occurrence-variant coefficients and, therefore, defines a linear time-variant filter.

#### VI. ESTIMATOR PERFORMANCE COMPARISON

A performance comparison of estimation algorithms is usually done in terms of bias and variance, but additional factors could also be considered, such as convergence speed, computational complexity, and delay. In this section, we will compare bias and variance when the deterministic signal  $s_k$  is assumed to be occurrence invariant, and we will present definitions of the SNR.

The output signal  $y_k$  for any linear coefficient estimation method can be written as

$$\mathbf{y}_k = \mathbf{T}(\mathbf{w}^{\mathrm{o}} + \mathbf{v}_k) \tag{50}$$

where  $\mathbf{v}_k$  depends on the selected estimation method. For the LMS algorithm, there is no compact expression of  $\mathbf{v}_k$  because the weight vector  $\mathbf{w}[n]$  is updated at every sample, and the output signal at the *k*th occurrence in (50) needs to be written in a scalar way

$$y[kN+j] = \boldsymbol{\tau}_j^T(\mathbf{w}^{\mathrm{o}} + \mathbf{v}[kN+j]) \quad 1 \le j \le N$$
 (51)

where  $\tau_j^T$  denotes the *j*th row of **T**. The expression of the weight error vector  $\mathbf{v}_k$  for each of the estimation methods is given in Table I. The expression for the LMS algorithm is taken from [5],  $\mathbf{F}_{i,j}$  denotes the transition matrix between the *i*th and the *j*th time instants [5]

$$\mathbf{F}_{i,j} = \prod_{n=j}^{i} \left( \mathbf{I} - 2\mu \boldsymbol{\tau}_n \boldsymbol{\tau}_n^T \right), \quad i \ge j$$
 (52)

and c[n] is the instantaneous signal truncation error within the signal subspace

$$c[n] = s[n] - \boldsymbol{\tau}_{n-kN}^T \mathbf{w}^{\text{o}}.$$
(53)

The estimation based on the inner product is sensitive to noise because the noise is directly projected on the subspace spanned by **T**. In contrast, the weight error vector for LMS and BLMS is composed of two terms: a transient (which depends on initial conditions and which will be null at steady-state) and a filtered version of the previous noise occurrences  $\mathbf{n}_i$ ,  $i = 0, 1, \dots, k-1$ . In the case of the BLMS algorithm,  $\mathbf{v}_k$  is an exponential average of previous noise occurrences. For the LMS algorithm, the expression is more cumbersome but conceptually similar. One main difference is that for the LMS, two terms are averaged: the instantaneous noise signal n[m] and the signal truncation error c[m]. Another difference is that the update averaging in BLMS is uncoupled because the coefficients  $1 - 2\mu$ are scalars, whereas in LMS, the update is coupled because the transition matrices  $\mathbf{F}_{i,j}$  are not diagonal. When small values of the step-size are used,  $\mathbf{F}_{N+j,j}$  can be approximated by  $(1-2\mu)\mathbf{I}$ [23, (25) and (26)], and then, both algorithms are approximately equivalent. However, the weight error vector for the BRLS algorithm is composed of only an exponential average of previous noise occurrences without the transient term.

The bias  $E{\mathbf{v}_k}$  and the variance  $E{\mathbf{v}_k^T \mathbf{v}_k}$  can be analyzed at any occurrence k using the expressions given in Table I. If the deterministic signal  $\mathbf{s}_k$  is assumed to be constant for all signal occurrences, the comparison can be made at steady state. For example, the expressions of steady-state bias and variance for the cases of zero-mean white noise with variance  $\sigma^2$  and zero-mean colored noise are given in Tables II and III, respectively. In the case of the LMS algorithm, both bias and variance are time-variant, i.e., different values are obtained at every time instant  $j, 1 \le j \le N$  of the kth signal occurrence.

When the deterministic signal s is constant only within a shorter interval, the bias and variance analysis should be done after a finite number of signal occurrences. Similar expressions of bias and variance could be obtained for this case using Table I, including the transient terms for LMS and BLMS.

The main objective of using adaptive algorithms in the linear expansion coefficient estimation is to reduce the noise in the

	Bias: $E \{ \mathbf{w}_{\infty} \} - \mathbf{w}^{\mathrm{o}} = E \{ \mathbf{v}_{\infty} \}$	Variance: $E\left\{ \left( \mathbf{w}_{\infty} - \mathbf{w}^{\circ} \right)^{T} \left( \mathbf{w}_{\infty} - \mathbf{w}^{\circ} \right) \right\} = E\left\{ \mathbf{v}_{\infty}^{T} \mathbf{v}_{\infty} \right\}$
IP	0	$p \sigma^2$
LMS	$2\mu \mathbf{B}_j \sum_{m=j}^{N+j-1} c[m] \mathbf{F}_{N+j-1,m+1} \boldsymbol{\tau}_m,$	$\frac{4\mu^2tr\big\{\mathbf{D}_j+\sigma^2\sum_{s=0}^\infty(\mathbf{F}_{N+j-1,j})^s\mathbf{Q}_j\big(\mathbf{F}_{N+j-1,j}^T\big)^s\big\},$
	where $\mathbf{B}_{j} = (\mathbf{I} - \mathbf{F}_{N+j-1,j})^{-1}$	where $\mathbf{Q}_{j} = \sum_{m=j}^{N+j-1} \mathbf{F}_{N+j-1,m+1} \boldsymbol{\tau}_{m} \boldsymbol{\tau}_{m}^{T} \mathbf{F}_{N+j-1,m+1}^{T}$ and
		$\mathbf{D}_{j} = \mathbf{B}_{j} \left( \sum_{m=j}^{N+j-1} \sum_{l=j}^{N+j-1} c[m] c[l] \mathbf{F}_{N+j-1,m+1} \boldsymbol{\tau}_{m} \boldsymbol{\tau}_{l}^{T} \mathbf{F}_{N+j-1,l+1}^{T} \right) \mathbf{B}_{j}^{T}$
BLMS	0	$\frac{\mu}{1-\mu} p \sigma^2$
BRLS	0	$\frac{1-\lambda}{1+\lambda} p \sigma^2$

 TABLE II

 Steady-State Bias and Variance for White Noise

TABLE III STEADY-STATE BIAS AND VARIANCE FOR COLORED NOISE

	Bias: $E\left\{\mathbf{v}_{\infty} ight\}$	Variance: $E\left\{\mathbf{v}_{\infty}^{T}\mathbf{v}_{\infty}\right\}$
IP	0	$E\left\{\mathbf{n}_{k}^{T}\mathbf{T}\mathbf{T}^{T}\mathbf{n}_{k}\right\} = \operatorname{tr}\left\{\mathbf{T}^{T}\mathbf{N}\mathbf{T}\right\}$
LMS	$2\mu \mathbf{B}_j \sum_{m=i}^{N+j-1} c[m] \mathbf{F}_{N+j-1,m+1} \boldsymbol{\tau}_m$	complex formula [5]
BLMS	0	$\frac{\mu}{1-\mu}\operatorname{tr}\left\{\mathbf{T}^{T}\mathbf{N}\mathbf{T}\right\}$
BRLS	0	$\frac{1-\lambda}{1+\lambda}\operatorname{tr}\left\{\mathbf{T}^{T}\mathbf{N}\mathbf{T}\right\}$

observed signal  $\mathbf{d}_k = \mathbf{s} + \mathbf{n}_k$ . A natural performance index will be the improvement of the SNR between input and output signals obtained by any of the estimation methods (IP, LMS, BLMS, and BRLS) at the *k*th occurrence. Let SNR<sup>i</sup><sub>k</sub> be the SNR of the input signal at the *k*th occurrence

$$\mathrm{SNR}_{k}^{\mathrm{i}} = \frac{E\{\mathbf{s}^{T}\mathbf{s}\}}{E\{\mathbf{n}_{k}^{T}\mathbf{n}_{k}\}}.$$
(54)

The SNR of the output signal  $y_k$  can be written using (50) as

$$\operatorname{SNR}_{k}^{\mathrm{o}} = \frac{E\{\mathbf{s}^{T}\mathbf{T}\mathbf{T}^{T}\mathbf{s}\}}{E\{\mathbf{v}_{k}^{T}\mathbf{v}_{k}\}}$$
(55)

and thus, the improvement in SNR will be

$$\Delta \text{SNR}_{k} = \frac{\text{SNR}_{k}^{\text{o}}}{\text{SNR}_{k}^{\text{i}}} = \frac{K}{E\left\{\mathbf{v}_{k}^{T}\mathbf{v}_{k}\right\}}$$
(56)

where K is a constant for a given signal subspace  $\mathbf{T}$  and a noisy observed signal  $\mathbf{d}_k$ . Accordingly, the improvement of the SNR is inversely proportional to the weight error variance. The steady-state  $\Delta$ SNR for each estimation method can be easily obtained using (56) and Tables II or III. For example, the steadystate improvement of SNR of BLMS versus IP is, for both cases of white and colored noise

$$\Delta \text{SNR}_{\infty}^{\text{BLMS/IP}} = \frac{1-\mu}{\mu}.$$
(57)

When the BLMS algorithm is used with  $\mu < 0.5$ , the steadystate output signal  $\mathbf{y}_{\infty}^{\text{BLMS}}$  is cleaner than  $\mathbf{y}_{\infty}^{\text{IP}}$ . On the other hand, the convergence speed of the BLMS algorithm will be low for small values of  $\mu$ .

In the case of the LMS algorithm, the weight vector is updated on a sample-by-sample basis, and the output signal at time instant n = kN + j is  $y[n] = \tau_j^T \mathbf{w}[n]$ . Accordingly, the instantaneous SNR of the output signal is evaluated as

$$SNR^{LMS}[n] = \frac{b^2[n]}{\boldsymbol{\tau}_j^T E\{\mathbf{v}[n]\mathbf{v}^T[n]\}\boldsymbol{\tau}_j \mathbf{x}[n]}$$
(58)



Fig. 3. ST-T complex selected for the simulation from a normal heartbeat.

where b[n] is the projection of the deterministic signal s[n]onto the subspace spanned by  $\mathbf{T}, b[n] = \boldsymbol{\tau}_{n-kN}^T \mathbf{w}^o$ . Then,  $\mathrm{SNR}^{\mathrm{LMS}}[n]$  is time variant, even at steady-state, because both numerator and denominator have different values at different time instants j of the signal occurrence. However, a comparison among different estimation methods should be done using the same temporal basis either for occurrences or samples. When the scenario is evaluated on an occurrence-by-occurrence basis, then (58) could be averaged over all instants of the k-th occurrence, and then

$$\operatorname{SNR}_{k}^{\operatorname{LMS}} = \frac{\sum_{j=1}^{N} b^{2}[j]}{\sum_{j=1}^{N} \tau_{j}^{T} E\{\mathbf{v}[kN+j]\mathbf{v}^{T}[kN+j]\}\boldsymbol{\tau}_{j}}.$$
 (59)

# VII. RESULTS

The performance of the four estimation methods (IP, LMS, BLMS, and BRLS) is illustrated by a simulation example in which the characteristics of an ECG signal are studied. In particular, the ECG is analyzed with respect to the ST-T complex (Fig. 3) since this part of the cardiac cycle frequently reflects myocardial ischemia. Ischemic heart disease constitutes one of



Fig. 4. Bias and variance for short memory ( $\mu = 0.3$ ;  $\lambda = 0.4$ ) and large truncation error (p = 1) that corresponds to 29.5% of signal energy.



Fig. 5. Bias and variance for long memory ( $\mu = 0.05$ ;  $\lambda = 0.9$ ) and large truncation error (p = 1) that corresponds to 29.5% of signal energy.

the most common fatal diseases in the western hemisphere. Myocardial ischemia is caused by a lack of sufficient blood flow to the contractile cells and may lead to myocardial infarction with its severe sequellae of heart failure, arrythmias, and death. Changes that occur in the ST-T complex due to ischemia are traditionally quantified by the amplitude measurement "ST60" obtained 60 ms after the depolarization phase has ended [26].

Basis functions derived by using the KL expansion [24] have been found useful for monitoring of ischemia [20]. The KL basis functions used in the present study were estimated from a training set of signals including several databases in order to adapt the basis functions to a large variety of ECG morphologies. The four most significant basis functions are also plotted in Fig. 3. It should be emphasized that although the KL basis functions have been selected here, other orthogonal expansion can be used as well.

The signal analyzed below was synthesized as a sequence of identical ST-T complexes, in the same way as was done in [5], to which white Gaussian noise was added with an  $SNR^i = 20$  dB. The four estimation methods (IP, BLMS, LMS, and BRLS) were then applied to the simulated signals. Average results from a set of 5000 trials are shown in Figs. 4–6, with several values

of the number of basis functions p and the step-size  $\mu$ . The results below present the performance during "steady-state" heart conditions; however, it is naturally of interest to also study the performance during changes in the ST-T segment; such a study is outside the scope of the present paper.

The first component weight error vector trajectory is illustrated in Fig. 4(a) when only one basis function is used in the expansion model with a large step-size ( $\mu = 0.3$ ;  $\lambda = 0.4$ ). The large steady-state bias of the LMS algorithm is due to the signal truncation error and the large value of the step-size, whereas the BLMS yields a steady-state unbiased estimate. On the other hand, IP and BRLS are unbiased at any occurrence. The variance evolution shown in Fig. 4(b) pinpoints the steady-state equivalence between BLMS and BRLS and their advantage versus IP and LMS. The variance of the LMS algorithm is shown at every time instant with a very large steady-state value due to the combination of large truncation error and large step-size.

If a larger amount of memory is used by the adaptive algorithms (lower value of  $\mu$  or higher value of  $\lambda$ ), the steady-state variance will be lower, but the convergence speed will decrease; see Fig. 5. It can be checked that the LMS and BLMS performance are very similar when very small value of the step-size



Fig. 6. Bias and variance for short memory ( $\mu = 0.3$ ;  $\lambda = 0.4$ ) and small truncation error (p = 4) that corresponds to 3.2% of signal energy.



Fig. 7. ST60 trends for several values of the number of basis functions and step-size and SNR = 20 dB. The ST60 amplitude of the clean signal was  $-47 \mu V$ .

are used, but there are still some differences due to the truncation error: The LMS is biased and with a slightly higher variance at steady state.

When a larger number of basis functions is used in the expansion, most of the signal energy is contained in the signal subspace spanned by  $\mathbf{T}$ , and the effect of the truncation error on the LMS is much less important (see Fig. 6), even for large values of  $\mu$  (note that when complete expansions are used, LMS and BLMS are equivalent for any step size). It is also illustrated in Figs. 4 and 6 that the number of basis functions used in the

expansion has a critical impact on the bias and variance performance of the LMS algorithm but not in IP, BLMS, or BRLS, where only the variance is affected in a linear way by the number of basis functions *p*.

Fig. 7 shows ST60 trends measured from the signals estimated by IP, LMS, BLMS, and BRLS for different conditions of the number of basis functions and the step size. It can be seen that the LMS yields a biased estimate, which is especially pronounced for large signal truncation error (low p) and large step size. It is also illustrated that the performance of LMS and



Fig. 8. ST60 trends with SNR = 10 dB and low-memory estimation. The ST60 amplitude of the clean signal was  $-47 \mu V$ .

BLMS is similar for low values of the step size ( $\mu = 0.05$ ). The variances of the four estimation methods are proportional to the number of basis functions.

In many situations, the SNR is much lower than 20 dB, whereas the signal properties may be changing. Fig. 8 examplifies this case by presenting the performance for two different step sizes. The number of basis functions is set to 4 in order to provide a sufficiently good signal characterization.

#### VIII. CONCLUSION

In this paper, the problem of adaptive estimation of linear transform coefficients for event-related signals was analyzed for a block structure with deterministic inputs. The BLMS algorithm was derived using the steepest descent strategy with block gradient estimation to minimize the mean square error. Its performance was found to be better than the LMS algorithm, providing a steady-state unbiased estimation of the Wiener solution and a lower steady-state variance that is unaffected by the signal truncation error.

Using instead a block-wise least-squares approach, the resulting BRLS algorithm yields an unbiased estimate for any occurrence and with lower variance than BLMS at the transient stage but with identical steady-state variance. The BRLS was shown to be equivalent to the BLMS with a decreasing step-size (larger values at the transient state to get a fast approximation to the optimum and lower values at steady-state to reduce the variance). It was shown that BRLS and BLMS have the same steady-state variance when  $2\mu = 1 - \lambda$ .

#### ACKNOWLEDGMENT

The authors would like to thank to the anonymous reviewers for their valuable comments.

#### REFERENCES

- C. Vaz and N. V. Thakor, "Adaptive Fourier estimation of time-varying evoked potentials," *IEEE Trans. Biomed. Eng.*, vol. 36, pp. 448–455, Apr. 1989.
- [2] N. V. Thakor, C. Vaz, R. W. McPherson, and D. F. Hanley, "Adaptive Fourier series modeling of time-varying evoked potentials: Study of human somatosensory evoked response to etomidate anesthetic," *Electroencephal. Clin. Neurophysiol.*, vol. 80, no. 2, pp. 108–118, Mar. 1991.
- [3] A. K. Barros, M. Yoshizawa, and Y. Yasuda, "Filtering noncorrelated noise in impedance cardiography," *IEEE Trans. Biomed. Eng.*, vol. 42, pp. 324–327, Mar. 1995.

- [4] 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," *Med. Biol. Eng. Comput.*, vol. 34, no. 1, pp. 58–68, Jan. 1996.
- [5] S. Olmos and P. Laguna, "Steady-state MSE convergence analysis in LMS adaptive filters with deterministic reference inputs for biomedical signals," *IEEE Trans. Signal Processing*, vol. 48, pp. 2229–2241, Aug. 2000.
- [6] G. A. Clark, S. K. Mitra, and S. R. Parker, "Block implementation of adaptive digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-28, pp. 584–592, June 1981.
- [7] S. S. Narayan, A. M. Peterson, and M. J. Narasimha, "Transform domain LMS filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 609–615, June 1983.
- [8] J. C. Lee and C. K. Un, "Performance of transform-domain LMS adaptive digital filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 499–510, June 1986.
- [9] D. F. Marshall, W. K. Jenkins, and J. J. Murphy, "The use of orthogonal transforms for improving performance of adaptive filters," *IEEE Trans. Circuits Syst.*, vol. 36, pp. 474–484, Apr. 1989.
- [10] J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Processing Mag.*, vol. 9, pp. 14–37, Jan. 1992.
- [11] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
- [12] T. D. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 2000.
- [13] 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, pp. 1032–1044, Oct. 1992.
- [14] 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 Process.*, vol. 48, no. 3, pp. 193–203, Feb. 1996.
- [15] P. Poon and N. V. Thakor, "Adaptive Fourier modeling of evoked potentials: Application to cerebral focal ischemia," *Proc. 12th Annual Conf. IEEE EMBS*, pp. 885–886, 1990.
- [16] C. Vaz, X. Kong, and N. V. Thakor, "An adaptive estimation of periodic signals using a Fourier linear combiner," *IEEE Trans. Signal Processing*, vol. 42, pp. 1–10, Jan. 1994.
- [17] A. K. Barros and N. Ohnishi, "MSE behavior of biomedical event-related filters," *IEEE Trans. Biomed. Eng.*, vol. 44, pp. 848–855, Sept. 1997.
- [18] 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, pp. 213–221, Mar. 1993.
- [19] L. Sörnmo, P. O. Börjesson, M. E. Nygårds, and O. Pahlm, "A method for evaluation of QRS shape features using a mathematical model for the ECG," *IEEE Trans. Biomed. Eng.*, vol. 28, pp. 713–717, Oct. 1981.
- [20] P. Laguna, G. B. Moody, J. García, A. L. Goldberger, and R. G. Mark, "Analysis of the ST-T complex of the electrocardiogram using the Karhunen-Loève transform: Adaptive monitoring and alternans detection," *Med. Biol. Eng. Comput.*, vol. 37, no. 2, pp. 175–189, Mar. 1999.

- [21] V. A. Allen and J. Belina, "ECG data compression using the discrete cosine transform," in *Comput. Cardiol.*, 1992, pp. 687–690.
- [22] W. Philips and G. De Jonghe, "Data compression of ECG's by highdegree polynomial approximation," *IEEE Trans. Biomed. Eng.*, vol. 39, pp. 330–337, Apr. 1992.
- [23] S. Olmos, J. Garcia, R. Jané, and P. Laguna, "Truncated orthogonal expansions of recurrent signals: Equivalence to a linear time-variant periodic filter," *IEEE Trans. Signal Processing*, vol. 47, pp. 3164–3172, Nov. 1999.
- [24] C. W. Therrien, Discrete Random Signals and Statistical Signal Processing. Englewood Cliffs, NJ: Prentice-Hall, 1992.
- [25] G. Glentins, K. Berberidis, and S. Theodoridis, "Efficient least squares adaptive algorithms for FIR transversal filtering," *IEEE Signal Processing Mag.*, vol. 16, pp. 13–41, July 1999.
- [26] G. M. Vincent, J. A. Abildskov, and M. J. Burgess, "Mechanism of ischemic ST-segment displacement: Evaluation by direct current recordings," *Circulation*, vol. 56, pp. 559–566, Oct. 1977.

**Salvador Olmos** (M'01) was born in Valencia, Spain, in 1969. He received the Industrial Engineering and Ph.D.degrees from the Polytechnic University of Catalonia, Barcelona, Spain, in 1993 and 1998, respectively.

He is currently an Associate Professor of signal processing and communications with the Department of Electronics and Engineering and Communications, Zaragoza University, Zaragoza, Spain. From August 2000 to August 2001, he was with the Department of Electroscience, Lund University, Lund, Sweden, supported by a post-doctoral research grant from Spanish Government. His professional research interests involve signal processing of biomedical signals. Leif Sörnmo (S'80–M'85) received the M.Sc. and D.Sc. degrees in electrical engineering from Lund University, Lund, Sweden, in 1978 and 1984, respectively.

He held a research position with the Department of Clinical Physiology, Lund University, from 1983 to 1995, where he worked on computer-based ECG analysis. Since 1990, he has been with the Signal Processing Group, Department of Electroscience, Lund University, where he now holds a position as a Professor in biomedical signal processing. He was on the editorial board of *Computers in Biomedical Research* from 1997 to 2000. His main research interests include statistical signal processing and modeling of biomedical signals. Current research projects include high-resolution ECG analysis, methods in ischemia monitoring, time–frequency analysis of atrial fibrillation, power efficient signal processing in pacemakers, and detection of otoacoustic emissions.

Dr. Sörnmo has been an Associate Editor of the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING since 2001.

**Pablo Laguna** (M'91) was born in Jaca (Huesca), Spain, in 1962. He received the M.S. degree in physics and the Ph.D. degree in physical science from the Science Faculty at the University of Zaragoza, Zaragosa, Spain, in 1985 and 1990, respectively. His Ph.D. thesis was developed at the Biomedical Engineering Division of the Institute of Cybernetics, Politecnic University of Catalonia (UPC-CSIC), Barcleona, Spain, under the direction of P. Caminal.

He is currently an Associate Professor of signal processing and communications with the Department of Electronics Engineering and Comunications at the Centro Politécnico Superior (CPS), University of Zaragoza. From 1987 to 1992, he worked as an Asistant Professor of automatic control with the Department of Control Engineering, and as a Researcher at the Biomedical Engineering Division of the Institute of Cybernetics, UPC-CSIC From 1999 to 2002, he was responsible for international relations at the CPS. His professional research interests are in signal processing, in particular applied to biomedical applications.