Proceedings – 23rd Annual Conference – IEEE/EMBS Oct.25-28, 2001, Istanbul, TURKEY

BLOCK LMS ADAPTIVE FILTER WITH DETERMINISTIC REFERENCE INPUTS FOR EVENT-RELATED SIGNALS

S. Olmos[1,2], L. Sörnmo[1], P. Laguna[2]

[1]Dept. of Electroscience, Lund University, Sweden

[2]Dept. of Electronics Eng. and Communications, Zaragoza Univ., Spain

***Abstract*—Adaptive estimation of the linear coefficient vector in truncated expansions is considered for the purpose of modeling noisy, recurrent signals. The block LMS (BLMS) algorithm, being the solution of the steepest descent strategy for minimizing the mean square error in a complete signal occurrence, 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 performance of the BLMS algorithm is studied on an ECG signal and the results show that its performance is superior to that of the LMS algorithm.**

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

I. INTRODUCTION

The problem of noise reduction in recurrent signals is well-studied and has traditionally been solved by ensemble averaging, or by one of the many variations on this technique. The time reference of each occurrence is often synchronized to a known, external stimulus; in certain signals, however, 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 the recurrent signal, while still achieving a reduction of the noise level, it is desirable to develop methods which incorporate a priori information on possible morphologies. More recently, modeling of each occurrence as a signal which 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 signal and the signal model represents a suitable cost function since it is related to signal energy. The optimal coefficients are determined on an individual occurrence basis, thus constituting a memory-less 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]. The reference inputs to the LMS algorithm are deterministic functions and 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 for every new sample based on an instantaneous gradient estimation. In a recent study, however, a steady-state 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 [2]. 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 for the case with random reference inputs and has, when the input is stationary, the same steady-state misadjustment and convergence speed as the LMS algorithm [3, 4]. 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.

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, sine and cosine functions, complex exponentials and Walsh functions, whereas the QRST complexes of the electrocardiogram (ECG) have been modeled by Hermite functions or basis functions that resulted from the Karhunen-Loève (KL) expansion. The ultimate purpose of the basis function description is not necessarily noise reduction, as mentioned above, but may as well be considered for data compression, feature extraction and monitoring.

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 sub-index k denotes the occurrence number. When a truncated orthogonal expansion is used, the estimated signal $\mathbf{y}_k$ is a linear combination of

This work was supported by P40/98 from CONSI+D-DGA and 2FD97-1197-CO2-01 from CICYT-FEDER and a post-doctoral grant to S. Olmos PF 00 21479412 from MECD (Spain).

basis functions

$$\mathbf{y}_k = \mathbf{T}\,\mathbf{w}_k \qquad (1)$$

where $\mathbf{T}$ 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 find the linear optimal coefficient vector $\mathbf{w}_k^{\mathrm{o}}$ is to minimize the cost function defined by the mean square error between $\mathbf{d}_k$ and $\mathbf{y}_k$

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

Applying differentiation we obtain that the Wiener solution for the linear coefficient vector is

$$\mathbf{w}_k^{\mathrm{o}} = \mathbf{T}^T E\left\{\mathbf{d}_k\right\} . \qquad (3)$$

This solution can be easily understood because the optimal signal description in the transform domain is the projection of the expected value of the observed signal.

The observed signal $\mathbf{d}_k$ is commonly contaminated by noise. Assuming an additive-noise model, each signal occurrence $\mathbf{d}_k$ can be decomposed as $\mathbf{d}_k = \mathbf{s}_k + \mathbf{n}_k$where $\mathbf{s}_k$ is a deterministic signal and $\mathbf{n}_k$ is zero-mean random noise. Since the clean signal $\mathbf{s}_k$ is unavailable, $\mathbf{w}_k^{\mathrm{o}}$ needs to be estimated according to (3) from the observed signal $\mathbf{d}_k$. A very simple way is to approximate $E\left\{\mathbf{d}_k\right\} \simeq \mathbf{d}_k$ in (3), implying that the linear coefficient vector is estimated by

$$\mathbf{w}_k^{\mathrm{IP}} = \mathbf{T}^T \mathbf{d}_k , \qquad (4)$$

where IP denotes the inner product between each basis function and the observed signal. This kind of estimation is memory-less since only information from the k-th occurrence is used to estimate $E\left\{\mathbf{d}_k\right\}$, 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. The 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 trade-off between convergence speed and signal to noise ratio 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, not necessarily obtained in a contiguous fashion. 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. The adaptive system estimates at each signal occurrence the amount of each reference input (columns of the $N \times p$ matrix $\mathbf{T}$) present in the input signal $\mathbf{d}_k$.

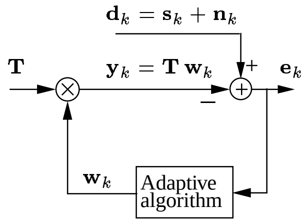


Fig. 1. Adaptive block-wise estimation of truncated expansions.

In order to minimize the cost function in (2), the optimal weight vector can be estimated by using an iterative algorithm based on the steepest descent strategy: the weight vector is updated once every occurrence according to

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu\,\frac{\partial J_k}{\partial \mathbf{w}_k} , \qquad (5)$$

where μ is the step-size that controls stability and convergence speed of the algorithm. The weight vector update equation can be obtained using the classical gradient approximation as

$$\mathbf{w}_{k+1} = \left(1 - 2\mu\right)\mathbf{w}_k + 2\mu\mathbf{T}^T\mathbf{d}_k . \qquad (6)$$

This algorithm is named BLMS because it is equivalent to the LMS algorithm but with a block-wise gradient estimation. In other words, the BLMS is equivalent to exponential averaging in the subspace spanned by $\mathbf{T}$. Note that the BLMS is equivalent to IP when $\mu = \frac{1}{2}$.

We will now consider the bias and variance of the BLMS algorithm, since these quantities are useful for comparison with other estimation methods. The weight error vector at the k-th occurrence $\mathbf{v}_k = \mathbf{w}_k - \mathbf{w}^{\mathrm{o}}$ can be written as

$$\mathbf{v}_k = \left(1-2\mu\right)^k \mathbf{v}_0 + 2\mu \sum_{j=0}^{k-1} \left(1-2\mu\right)^{k-j-1} \mathbf{T}^T \mathbf{n}_j . \quad (7)$$

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 (7) will be important. Taking $\lim_{k\to\infty}$ and the expected value we obtain

$$\lim_{k\to\infty} E\left\{\mathbf{v}_k\right\} = E\left\{\mathbf{v}_\infty\right\} = \mathbf{0} , \qquad (8)$$

since the noise $\mathbf{n}_j$ is assumed to be zero-mean. Accordingly the steady-state weight vector $\mathbf{w}_\infty$ is an unbiased estimator of the Wiener solution (3).

The steady-state weight error vector variance will be

$$E\left\{\mathbf{v}_\infty^T \mathbf{v}_\infty\right\} = \lim_{k\to\infty} 4\mu^2 tr \left\{ \sum_{j=0}^{k-1}\sum_{l=0}^{k-1} \left(1-2\mu\right)^{k-l-1} \mathbf{T}^T E\left\{\mathbf{n}_l \mathbf{n}_j^T\right\} \mathbf{T} \; \left(1-2\mu\right)^{k-j-1} \right\} . \quad (9)$$

If the noise signal is assumed to be stationary with correlation function shorter than the gap between consecutive occurrences[1], then

$$E\left\{\mathbf{n}_l\mathbf{n}_j\right\} = \delta_{lj}\,\mathbf{N} \tag{10}$$

being $\mathbf{N}$ the $N\times N$ noise covariance matrix. Accordingly, the steady-state weight error vector energy is

$$\begin{aligned} E\left\{\mathbf{v}_\infty^T\mathbf{v}_\infty\right\} &= \lim_{k\to\infty} 4\mu^2 \sum_{j=0}^{k-1} (1-2\mu)^{2(k-j-1)}\, tr\left\{\mathbf{T}^T\mathbf{N}\mathbf{T}\right\} \\ &= \frac{\mu}{1-\mu}\, tr\left\{\mathbf{T}^T\mathbf{N}\mathbf{T}\right\}. \end{aligned} \tag{11}$$

The MSE is usually decomposed as $J_k^{\mathrm{BLMS}} = J_k^{\mathrm{o}} + J_k^{\mathrm{ex}}$ where J_k^{o} is the MSE at the optimum and the excess MSE J_k^{ex} can be written as

$$J_k^{\mathrm{ex}} = E\left\{\mathbf{v}_k^T\mathbf{v}_k\right\} - 2\,E\left\{\mathbf{v}_k^T\mathbf{T}^T\mathbf{e}_k^{\mathrm{o}}\right\} \tag{12}$$

At the optimal solution, the orthogonality principle applies

$$\mathbf{T}^T\mathbf{e}_k^{\mathrm{o}} = \mathbf{T}^T\left(\mathbf{I} - \mathbf{T}\mathbf{T}^T\right)\mathbf{s} + \mathbf{T}^T\mathbf{n}_k = \mathbf{T}^T\mathbf{n}_k\,. \tag{13}$$

Hence the cost functions can be written as

$$J_k^{\mathrm{BLMS}} = J_k^{\mathrm{o}} + E\left\{\mathbf{v}_k^T\mathbf{v}_k\right\} - 2\,E\left\{\mathbf{v}_k^T\mathbf{T}^T\mathbf{n}_k\right\} \tag{14}$$

Using the same noise assumption as in (10) the cross term $E\left\{\mathbf{v}_k^T\mathbf{T}^T\mathbf{n}_k\right\}$ at steady-state will be null because

$$\lim_{k\to\infty}\sum_{j=0}^{k-1}(1-2\mu)^{k-j-1}\, tr\left\{\mathbf{T}^T E\left\{\mathbf{n}_k\mathbf{n}_j^T\right\}\mathbf{T}\right\} = 0\,. \tag{15}$$

Summing up, the steady-state MSE will be

$$J_\infty^{\mathrm{BLMS}} = J_\infty^{\mathrm{o}} + \frac{\mu}{1-\mu}\, tr\left\{\mathbf{T}^T\mathbf{N}\mathbf{T}\right\}. \tag{16}$$

In the case of complete expansions we have that $tr\left\{\mathbf{T}^T\mathbf{N}\mathbf{T}\right\} = tr\left\{\mathbf{N}\right\}$ which is equal to the noise energy. In the case of white noise and incomplete expansions

$$J_\infty^{\mathrm{BLMS}} = J_\infty^{\mathrm{o}} + \frac{\mu}{1-\mu}\, p\,\sigma^2\,. \tag{17}$$

It may be worthwhile to point out certain relationships to the LMS algorithm. In the case of the LMS algorithm, J_k^{ex} is composed of three terms [2, Eq. (12-13)] while for the BLMS algorithm only two terms are present in (14) because the truncation signal error is orthogonal to the input basis functions $\mathbf{T}$. Moreover, the LMS algorithm converges to a biased estimate for truncated expansions [2], while the BLMS estimation is steady-state unbiased. When complete expansions $p = N$ are used it can be noted that $J_\infty^{\mathrm{BLMS}} = J_\infty^{\mathrm{LMS}}$ [2]. This result agrees with the fact that the algorithms become identical when complete expansions are used (compare (6) with [2, Eq. (8)]).

[1] The very-low frequency components of biomedical signals, e.g. baseline wander in the ECG, are usually removed in a pre-processing stage because they do not convey any valuable clinical information.

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 [5] 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 equations of BLMS (6) the term $\mathbf{T}^T\mathbf{d}_k$ represents the IP estimation of $\mathbf{w}^{\mathrm{o}}$ using only information from the k-th occurrence. The first term accounts for the estimation done at the previous occurrence. Consequently, 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 [5]. 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 (6) by $\mathbf{T}$

$$\mathbf{y}_{k+1} = (1-2\mu)\,\mathbf{y}_k + 2\mu\mathbf{T}\mathbf{T}^T\mathbf{d}_k\,. \tag{18}$$

In the case of complete expansions $\mathbf{T}\mathbf{T}^T = \mathbf{I}_N$, and the coefficients in (18) are scalar and time-invariant. When truncated orthogonal expansions $p < N$ are considered, a coupled system of finite difference equations is obtained from (18) because $\mathbf{T}\mathbf{T}^T \neq \mathbf{I}_N$.

V. Results

The performance of the three estimation methods (IP, LMS and BLMS) is illustrated by a simulation example in which the characteristics of an ECG signal is studied. In particular, the ECG is analyzed with respect to the ST-T complex (Fig. 2) since this part of the cardiac cycle frequently reflects myocardial ischemia (this condition is caused by a lack of blood supply in a certain region of the heart wall). Changes that occur in the ST-T complex due to ischemia are traditionally quantified by the amplitude measurement "ST60" obtained 60 milliseconds after the depolarisation phase has ended.

Basis functions derived by using the KL expansion have been found useful for monitoring of ischemia [6]. 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. 2. 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 [2], to which Gaussian white noise was added with an $\mathrm{SNR} = 20$ dB. The three estimation methods (IP,

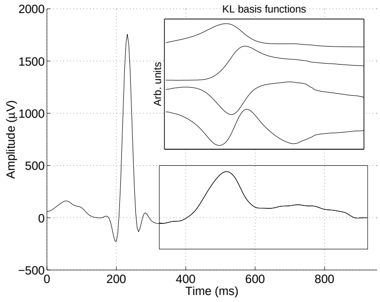
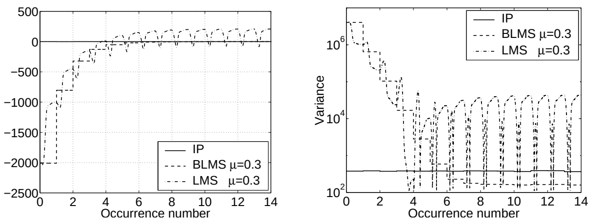


Fig. 2. ST-T complex selected for the simulation from a normal heartbeat.



(a) Bias of the first coef. $\mathbf{w}_k[1]$ (b) Variance $E\left\{\mathbf{v}_k^T\mathbf{v}_k\right\}$

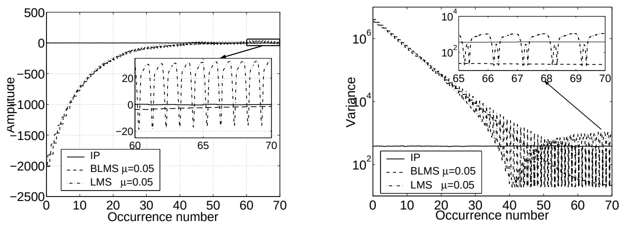
Fig. 3. Bias and variance for small memory (μ=0.3) and large truncation error ($p = 1$) which corresponds to 29.5% of the signal energy.

LMS and BLMS) were then applied to the simulated signals. Average results from a set of 5000 trials are shown in Figs. 3–5, with several values of the number of basis functions p and the step-size μ. The results below presents 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 study is outside the scope of the present paper.

The first component weight error vector trajectory is illustrated in Fig. 3(a) when only one basis function is used in the expansion model with a large step-size ($\mu = 0.3$). The steady-state bias of the LMS algorithm is large due to the truncation signal error and the large value of the step-size. In contrast, BLMS obtains an steady-state unbiased estimate; IP is unbiased at any occurrence. The variance of the LMS algorithm is shown at every time instant. The large variance is due to the combination of large truncation error and large step-size.

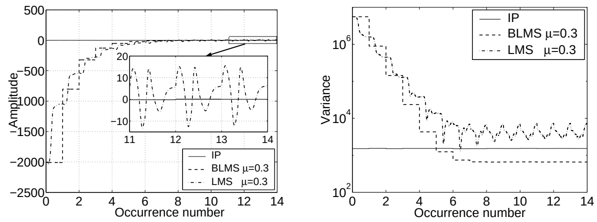
If more memory is used by the adaptive algorithms (lower value of μ), the steady-state variance will be lower, but the convergence speed will decrease, as it is illustrated in Fig. 4. It can be checked that the LMS and BLMS performance are very similar when very small value of the step-size are used, but 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. 5), even for large values of μ (note that when complete expansions are used LMS and BLMS are equivalent for any step-size). It is also illustrated in Figs. 3 and 5 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, or BLMS, where only the variance is affected in a linear way by the number of basis functions p.



(a) Bias of the first coef. $\mathbf{w}_k[1]$ (b) Variance $E\left\{\mathbf{v}_k^T\mathbf{v}_k\right\}$

Fig. 4. Bias and variance for large memory (μ=0.05) and large truncation error ($p = 1$) which corresponds to 29.5% of the signal energy.



(a) Bias of the first coef. $\mathbf{w}_k[1]$ (b) Variance $E\left\{\mathbf{v}_k^T\mathbf{v}_k\right\}$

Fig. 5. Bias and variance for small memory (μ=0.3) and small truncation error ($p = 4$) which corresponds to 3.2% of the signal energy.

VI. Conclusions

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

References

[1] 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] S. Olmos and P. Laguna, "Steady-state MSE convergence analysis in LMS adaptive filters with deterministic reference inputs for biomedical signals," *IEEE Trans. Signal Proc.*, vol. 48, pp. 2229–2241, 2000.

[3] G. Clark, S. Mitra, and S. Parker, "Block implementation of adaptive digital filters," *IEEE Trans. Circuits and Systems*, vol. CAS-28, no. 6, pp. 584–592, 1981.

[4] S. Haykin, *Adaptive Filter Theory*. New Jersey: Prentice-Hall, 3rd ed., 1996.

[5] 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, 1999.

[6] 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: adaptive monitoring and alternans detection," *Med. Biol. Eng. Comput.*, vol. 37, pp. 175–189, 1999.