CASCADED ALL-PASS SECTIONS FOR LMS ADAPTIVE FILTERING

H.J.W. Belt and H.J. Butterweck
Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands
H.J.W.Belt@ele.tue.nl and H.J.Butterweck@ele.tue.nl

ABSTRACT

The behaviour of the LMS adaptive algorithm is analyzed for a class of adaptive filters that is based on a cascade of identical N-th order all-pass sections. The well-known tapped-delay-line is a special case of this class. We look at the rate of convergence and the steady-state weight fluctuations. It is shown that in the steady state the weight-error correlation matrix satisfies a Lyapounov equation for sufficiently small values of the step-size. Sometimes *a priori* knowledge of the unknown reference system is available that can be used to select the N parameters of the all-pass section. In these cases the LMS adaptive filter based on a cascade of identical all-pass sections can outperform the LMS adaptive tapped-delay-line.

1 INTRODUCTION

The widespread use of the LMS adaptive algorithm applied to a tapped-delay-line (TDL) is a direct consequence of its simplicity and good performance. For the analysis of the behaviour of the LMS algorithm most authors rely on an "independence assumption" stating statistical independence of successive input vectors [1]. It is argued that for the analysis to be tractable this assumption must be made. However, due to the (deterministic) coherence between successive input vectors, the independence assumption cannot be justified. This logical inconsistency has recently led to alternative approaches. Particularly the weight fluctuations in the fully adapted state are studied in [2] *without* making use of the independence assumption. An iterative procedure is used to arrive at a power series of the weight-error correlation matrix (WECM) in terms of the step-size.

In this paper we study the behaviour of the LMS algorithm applied to a more general class of adaptive filters. The delay line is replaced by a cascade of identical all-pass sections of order N and each tap signal is filtered to yield N outputs. In Section 2 this structure is introduced and a few relevant properties are given. In Section 3 some notational aspects concerning the LMS algorithm applied to the all-pass filter bank are treated. Section 4 deals with the WECM, which is shown to be the solution of a Lyapounov equation. Two explicitly solvable cases of the Lyapounov equation are considered. Section 5 contains a performance analysis of the LMS algorithm applied to the all-pass filter bank in terms of adaptation speed and misadjustment. Finally, some concluding remarks are made in the Discussion.

2 THE ALL-PASS FILTER BANK

Consider a filter bank of P sections, cf. Fig. 1. Compared to the TDL each delay element has been replaced by an N-th order all-pass section with the transfer function

$$A(z) = \frac{\prod_{i=1}^{N}(1 - z\,p_i)}{\prod_{i=1}^{N}(z - p_i)}.$$

The (not necessarily distinct) poles p_i are real or occur in complex conjugate pairs and lie inside the unit circle. Each tap signal in the cascade is filtered by a section $B(z)$ with one input and N outputs. The latter is chosen such that the NP impulse responses of the filter bank form an orthonormal set. These functions are usually referred to as "Kautz functions" [3],[4].

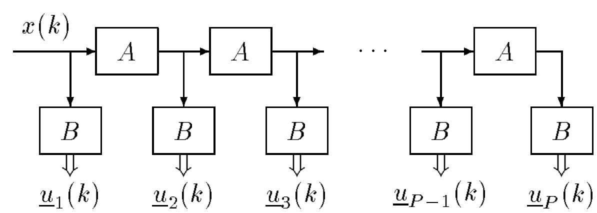

Figure 1: Orthonormal filter bank based on a cascade of identical N-th order all-pass sections

If we take $A(z) = (1 - \xi z)/(z - \xi)$ and $B(z) = \sqrt{1-\xi^2}\, z/(z - \xi)$ with $|\xi| < 1$ we have $N = 1$ and we obtain a "Laguerre filter bank" [5]. Its orthonormal impulse responses are called the "discrete Laguerre functions" [4]. Note that for $\xi = 0$ the Laguerre filter bank degenerates into a TDL.

Although most of the following analysis applies to more general structures, that of Fig. 1 deserves special attention: all-passes imply energy conservation along the filter line and good numerical properties, while the repetitive structure can be easily implemented.

The Kautz filter bank is assumed to be excited by a zero-mean, stationary stochastic signal $x(k)$ with autocorrelation function $X_l = E\{ x(k)x(k-l) \}$, power $\sigma_x^2 = X_0$, and power spectral density function $\Phi_{xx}(\Omega)$. Furthermore, it operates in the steady state such that also the $M = NP$ outputs of the filter bank can be viewed as stationary. These outputs are written in vector notation as $\underline{u}(k) = \{ \underline{u}_1^t(k), \ldots, \underline{u}_P^t(k) \}^t$, where $\underline{u}_p(k)$ denotes the $N \times 1$ output vector of the p-th section ($p = 1, 2, \ldots, P$). The corresponding $M \times M$ covariance matrix $\boldsymbol{R} = E\{ \underline{u}(k)\, \underline{u}^t(k) \}$ is symmetric and (semi-) positive definite. It has a block-Toeplitz structure where the size of the blocks is $N \times N$. Its eigenvectors are orthogonal and its eigenvalues λ_1 to λ_M are non-negative. For white $x(k)$ the covariance matrix becomes $\boldsymbol{R} = \sigma_x^2 \boldsymbol{I}$ because the filter bank is orthonormal.

We now show that the smallest eigenvalue $\lambda_{\min}$ of $\boldsymbol{R}$ satisfies the inequality [6]

$$\lambda_{\min} \;=\; \min_{\underline{a} \neq \underline{0}} \frac{\underline{a}^t \boldsymbol{R} \underline{a}}{\underline{a}^t \underline{a}} \;\geq\; \min\{ \Phi_{xx}(\Omega) \}. \qquad (1)$$

Let $g_m(k)$ with $m = 1, 2, \ldots, M$ denote the impulse response of the m-th filter bank output, and $G_m(e^{j\Omega})$ its Fourier transform. We then find

$$\begin{aligned} \underline{a}^t \boldsymbol{R} \underline{a} \;&=\; \frac{1}{2\pi} \int_{-\pi}^{\pi} |\sum_{m=1}^{M} a_m G_m(e^{j\Omega})|^2\, \Phi_{xx}(\Omega)\, d\Omega \\ &\geq\; \frac{1}{2\pi} \int_{-\pi}^{\pi} |\sum_{m=1}^{M} a_m G_m(e^{j\Omega})|^2 d\Omega\;\; \min\{\Phi_{xx}(\Omega)\}, \end{aligned}$$

where the a_m are the M components of the vector $\underline{a}$. Using the orthogonality of the impulse responses $g_m(k)$ implying that of the system functions $G_m(e^{j\Omega})$, we easily obtain

$$\underline{a}^t \boldsymbol{R} \underline{a} \;\geq\; \underline{a}^t \underline{a}\;\; \min\{\Phi_{xx}(\Omega)\},$$

which implies (1). The counterpart of (1), not used in this paper, reads as [6]

$$\lambda_{\max} \;\leq\; \max\{ \Phi_{xx}(\Omega) \}.$$

Hereafter we assume that $x(k)$ is "persistently exciting", which means that $\Phi_{xx}(\Omega) > 0$ for all frequencies, or equivalently, $\lambda_{\min} > 0$.

3 THE LMS ADAPTIVE ALL-PASS FILTER

Let $x(k)$ simultaneously excite an adaptive filter of the above-mentioned type and a linear and time-invariant "reference system" whose output, corrupted by additive noise $n(k)$, yields the "desired signal" $d(k)$. The noise $n(k)$ is a zero-mean, stationary stochastic process with autocorrelation function $N_l = E\{n(k)n(k-l)\}$, power $\sigma_n^2 = N_0$, and power spectral density function $\Phi_{nn}(\Omega)$. The weight vector of the adaptive filter is denoted by $\underline{w}(k) = \{ w_1(k), \ldots, w_M(k) \}^t$. The output $y(k)$ of the adaptive filter can be written as $\underline{w}^t(k)\, \underline{u}(k)$. By subtracting $y(k)$ from the desired signal we construct the "error signal" $e(k) = d(k) - y(k)$, which can be used to update $\underline{w}(k)$. In the (optimal) Wiener solution $\underline{w} = \underline{w}_o$ the weights are such that the Mean-Squared Error (MSE) $E\{ e^2(k) \}$ is minimal.

The weight error vector $\underline{v}(k)$ is defined as the instantaneous difference between the weight vector of the adaptive filter and the Wiener solution: $\underline{v}(k) = \underline{w}(k) - \underline{w}_o$. Assuming that the reference system can be modelled by the Kautz filter set used for the adaptive filter, so that its output equals $\underline{w}_0^t\, \underline{u}(k)$, the error signal can be written as

$$\begin{aligned} e(k) \;&=\; n(k) \;+\; \underline{w}_o^t\, \underline{u}(k) \;-\; \underline{w}^t\, \underline{u}(k) \\ &=\; n(k) \;-\; \underline{v}^t(k)\, \underline{u}(k). \end{aligned}$$

The LMS adaptive algorithm is given by

$$\underline{w}(k+1) \;=\; \underline{w}(k) \;+\; \mu\, e(k)\, \underline{u}(k).$$

Here μ is the adaptation constant or step-size. We may write for the weight-error vector:

$$\begin{aligned} \underline{v}(k+1) \;&=\; \underline{v}(k) \;-\; \mu\, \underline{u}(k)\, \underline{u}^t(k)\, \underline{v}(k) \;+\; \mu\, n(k)\, \underline{u}(k) \\ &=\; (\boldsymbol{I} \;-\; \mu\, \boldsymbol{R}(k))\, \underline{v}(k) \;+\; \mu\, \underline{f}(k), \qquad (2) \end{aligned}$$

where we have used the short-hand notation $\boldsymbol{R}(k) = \underline{u}(k)\underline{u}^t(k)$ and $\underline{f}(k) = n(k)\underline{u}(k)$. As in [2], equation (2) is the starting point for the determination of the positive (semi-)definite WECM, given by $\boldsymbol{V} = E\{ \underline{v}(k)\, \underline{v}^t(k) \}$.

4 THE WECM

For sufficiently small values of μ the time-varying term $\boldsymbol{R}(k)$ in (2) can be replaced by its constant average $\boldsymbol{R}$. Then in the limiting case $\mu \to 0$, the WECM satisfies a Lyapounov equation (henceforth all summations extend from $-\infty$ to ∞ unless stated otherwise):

$$\boldsymbol{R}\boldsymbol{V} \;+\; \boldsymbol{V}\boldsymbol{R} \;=\; \mu\, \boldsymbol{F}, \qquad (3)$$

where $\boldsymbol{F} = \sum_l \boldsymbol{F}_l$, $\boldsymbol{F}_l = N_l \boldsymbol{R}_l$, $\boldsymbol{R}_l = E\{\underline{u}(k)\, \underline{u}^t(k-l)\}$, and $\boldsymbol{R} = \boldsymbol{R}_0$. For the proof the reader is referred to the Appendix. This Lyapounov equation was already found in [2] and [7] for the special case of the TDL. We consider two explicitly solvable cases:

- If $n(k)$ is white, the right-hand side of equation (3) equals $\mu\sigma_n^2 \boldsymbol{R}$, and the WECM becomes

$$\boldsymbol{V} \;=\; \tfrac{1}{2}\, \mu\, \sigma_n^2\, \boldsymbol{I}.$$

- If $x(k)$ is white we get

$$\boldsymbol{V} \;=\; \tfrac{1}{2}\, \mu\, \tilde{\boldsymbol{R}}, \qquad (4)$$

where $\tilde{\boldsymbol{R}}$ is the covariance matrix of the outputs $u_1(k)$ to $u_M(k)$ of the Kautz filter in the imaginary situation that it is excited by $n(k)$ instead of $x(k)$. Here we have used that the impulse responses of the filter bank are orthonormal.

Proof: For the $\{p,q\}$-th element of the matrix $\boldsymbol{F}$ in (3) we may write

$$\begin{aligned}\{\boldsymbol{F}\}_{pq} &= \sum_l N_l E\{\, u_p(k)u_q(k-l)\,\} \\ &= \sum_l N_l \sum_{s=0}^{\infty}\sum_{t=0}^{\infty} g_p(s)g_q(t)X_{s-l-t} \\ &= \sigma_x^2 \sum_{s=0}^{\infty}\sum_{t=0}^{\infty} g_p(s)g_q(t)N_{s-t}, \qquad (5)\end{aligned}$$

which represents σ_x^2 times the covariance between the p-th and the q-th output of the filter bank as if it had been excited by $n(k)$ instead of $x(k)$. Also, from (3) with $\boldsymbol{R} = \sigma_x^2 \boldsymbol{I}$ we know that

$$\mu\{\boldsymbol{F}\}_{pq} \;=\; 2\sigma_x^2\{\boldsymbol{V}\}_{pq}. \qquad (6)$$

Combination of (5) and (6) concludes the proof of (4).

5 PERFORMANCE ANALYSIS

In this section our interest will be focussed on the terminal behaviour of the LMS algorithm. The performance of an adaptation algorithm is usually characterized by two measures: the rate of convergence and the misadjustment. As is well known, the rate of convergence of the LMS algorithm is determined by the smallest eigenvalue of $\boldsymbol{R}$, which, in our case, is lower-bounded by the minimum of $\Phi_{xx}(\Omega)$, see (1). Therefore, the rate of convergence of the LMS algorithm remains bounded even when the number of adaptive weights M is continuously increased.

For a discussion of the misadjustment we decompose the steady-state output of the adaptive filter as

$$\begin{aligned} y(k) &= \underline{w}_o^t\, \underline{u}(k) \;+\; \underline{v}^t(k)\, \underline{u}(k) \\ &= y_w(k) \;+\; y_f(k), \end{aligned}$$

where $y_w(k)$ is the contribution due to the Wiener coefficients $\underline{w}_o$ and $y_f(k)$ is that due to the weight fluctuations. As proposed in [8], for a small step-size μ the misadjustment should be defined as

$$\text{misadjustment} \;=\; \frac{E\{\, y_f^2(k)\,\}}{\sigma_n^2}.$$

We then obtain

$$E\{\, y_f^2(k)\,\} \;=\; E\{\, \underline{u}^t(k)\, \underline{v}(k)\, \underline{v}^t(k)\, \underline{u}(k)\,\}.$$

For sufficiently small values of μ the vector $\underline{v}(k)$ varies so much slower than $\underline{u}(k)$ that we may write [7], using (3)

$$\begin{aligned} E\{\, y_f^2(k)\,\} &\approx E\{\, \underline{u}^t(k)\, \boldsymbol{V}\, \underline{u}(k)\,\} \\ &= \frac{1}{2}\,\text{trace}\{\, \boldsymbol{R}\boldsymbol{V} \;+\; \boldsymbol{V}\boldsymbol{R}\,\} \\ &= \frac{\mu}{2}\sum_l N_l\, E\{\, \underline{u}^t(k)\, \underline{u}(k-l)\,\} \\ &= \frac{\mu}{2}\sum_{m=1}^{M}\frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_{nn}(\Omega)\,\Phi_{xx}(\Omega)\,|G_m(e^{j\Omega})|^2 d\Omega, \qquad (7)\end{aligned}$$

Note the symmetry with respect to the excitation signal and the additive noise signal. Because all contributions to the summation in (7) are strictly positive, we can draw the important conclusion that the misadjustment grows without bound with increasing M.

In the special case "$x(k)$ white" the result in (7) becomes $\frac{1}{2}\mu M\sigma_x^2 \text{trace}\{\tilde{\boldsymbol{R}}\}$. When $n(k)$ is white we obtain $\frac{1}{2}\mu M\sigma_n^2 \text{trace}\{\boldsymbol{R}\}$. When both $x(k)$ and $n(k)$ are white the result in (7) reduces to $\frac{1}{2}\mu M\sigma_n^2\sigma_x^2$.

In the Laguerre case where $N = 1$ we have that $|G_m(e^{j\Omega})|^2 = |B(e^{j\Omega})|^2$ for all frequencies, which is a consequence of the all-pass structure. As a result we obtain

$$E\{y_f^2(k)\} = \frac{\mu M}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_{nn}(\Omega)\,\Phi_{xx}(\Omega)\,|B(e^{j\Omega})|^2\, d\Omega.$$

For $\xi = 0$ the Laguerre filter reduces to the TDL, and we find (see also [7])

$$E\{\, y_f^2(k)\,\} \;=\; \frac{\mu M}{2}\,\frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_{nn}(\Omega)\,\Phi_{xx}(\Omega)\, d\Omega,$$

which becomes $\frac{1}{2}\mu M\sigma_n^2\sigma_x^2$ when either $x(k)$ or $n(k)$ is white.

6 DISCUSSION

We have analyzed the LMS adaptive algorithm for a class of adaptive filters that is based on a cascade of identical N-th order all-pass sections. We have shown that the rate of convergence remains bounded even when the number of weights M is continuously increased. We have also shown that the misadjustment grows without bound with increasing M, and thus for a small misadjustment M should be kept small.

Observe, however, that besides the rate of convergence and the misadjustment a third criterion has to be reckoned with, viz. a good approximation of the reference system. In the Wiener solution, the adaptive filter imitates the reference system such that a properly defined approximation error is minimized. For a given reference system each of the competitive adaptive filter structures yields an approximation error that decreases monotonically with the number of filter sections M. Conversely, for a given approximation error the minimally required number of filter sections M_{min} depends on the chosen structure. If this structure contains one or more extra free parameters (in our case the poles p_1 to p_N), M_{min} becomes a function of these parameters and can be minimized when appropriate *a priori* knowledge of the reference system is available, see e.g. [9]. When this knowledge is not available, it is sometimes possible to adaptively optimize the free parameters. This is done for the pole of an adaptive Laguerre filter in [10]. Compared with the classical TDL the more general all-pass filter banks often turn out to yield substantially lower values of M_{min}. As we have shown, a smaller M_{min} leads to a better LMS adaptive filter.

References

[1] S. Haykin, Adaptive filter theory, Third Edition, Prentice-Hall Int. London 1996.

[2] H.J. Butterweck, "A steady-state analysis of the LMS adaptive algorithm without use of the independence assumption", *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Detroit, Michigan, USA, May 9-12, 1995, Vol. 2, pp. 1404-1407.

[3] W. Kautz, "Transient synthesis in the time domain", *IRE Trans. Circuit Theory*, Vol. 1, pp. 29-39, 1954.

[4] P.W. Broome, "Discrete orthonormal sequences", *J. ACM*, Vol. 12, No. 2, pp. 151-168, 1965.

[5] R.E. King and P.N. Paraskevopoulos, "Digital Laguerre filters", *Int. J. Circuit Theory and Appl.*, Vol. 5, pp. 81-91, 1977.

[6] T. Oliveira e Silva, "Kautz filters", English translation of a work written in Portuguese for the "Prémio Científico IBM 94". *ftp://inesca.inesca.pt/pub/tos/English/revdetua.ps.gz.*

[7] H.J. Butterweck, "An approach to LMS adaptive filtering without use of the independence assumption", *Proc. VIII European Signal Processing Conference (EUSIPCO)*, Trieste, Italy, Sept. 10-13, 1996.

[8] H.J. Butterweck, "A new interpretation of the misadjustment in adaptive filtering", *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Atlanta, Georgia, USA, May 7-10, 1996, Vol. 3, pp. 1641-1643.

[9] N. Tanguy, P. Vilbé, and L.C. Calvez, "Optimum Choice of Free Parameter in Orthonormal Approximations", *IEEE Trans. Automatic Control*, Vol. 40, pp. 1811-1813, 1995.

[10] H.J.W. Belt and A.C. den Brinker "Adaptive Laguerre filters with pole optimization", to appear in: *Proc. IEEE Int. Symp. on Circuits and Systems (ISCAS)*, Atlanta, Georgia, USA, May 12-15, 1996.

APPENDIX

We prove that the WECM satisfies a Lyapounov equation in the limiting case $\mu \to 0$. It is assumed that the filter operates in the steady state since $-\infty$, such that the influence of the initial conditions can be neglected. With (2) we may write for the weight error vector

$$\underline{v}(k+1) \;=\; \mu\sum_{l=0}^{\infty}(I-\mu\,\boldsymbol{R})^l\,\underline{f}(k-l) \;=\; \mu\sum_{l=0}^{\infty}\boldsymbol{D}^l\,\underline{f}(k-l),$$

where $\boldsymbol{D}$ is the symmetric "damping matrix". We will use the short-hand notation $\boldsymbol{D}_l = \boldsymbol{D}^l$ for $l \geq 0$ and $\boldsymbol{D}_l = 0$ for $l < 0$. For later use the following properties of $\boldsymbol{D}_l$ are noted here:

$$\boldsymbol{D}_l^t = \boldsymbol{D}_l \text{ and } \boldsymbol{D}\boldsymbol{D}^l = \boldsymbol{D}^l\boldsymbol{D} \text{ and } \boldsymbol{D}_{l+1} = \boldsymbol{D}\boldsymbol{D}_l + \delta_{l+1}\boldsymbol{I},$$

where δ_l is the Kronecker delta function. We find for the WECM

$$\begin{aligned}\boldsymbol{V} &= E\{\,\underline{v}(k)\,\underline{v}^t(k)\,\}\\ &= \mu^2\sum_l\sum_m \boldsymbol{D}_l\,E\{\,\underline{f}(k-l)\underline{f}^t(k-m\,\}\boldsymbol{D}_m^t\\ &= \mu^2\sum_l\sum_m \boldsymbol{D}_l\,\boldsymbol{F}_m\,\boldsymbol{D}_{m+l} \;=\; \mu^2\sum_m \boldsymbol{T}_m,\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{T}_m &= \sum_l \boldsymbol{D}_l\,\boldsymbol{F}_m\,\boldsymbol{D}_{m+l}\\ &= \sum_l \boldsymbol{D}_{l+1}\,\boldsymbol{F}_m\,\boldsymbol{D}_{m+l+1}\\ &= \boldsymbol{D}\,\boldsymbol{T}_m\,\boldsymbol{D} + \boldsymbol{F}_m\,\boldsymbol{D}\,\boldsymbol{D}_{m-1} +\\ &\quad\; \boldsymbol{D}_{-m-1}\,\boldsymbol{F}_m \;+ \boldsymbol{F}_0\,\delta_m.\end{aligned} \tag{8}$$

For the first term in (8) we may use the following approximation when μ is sufficiently small:

$$\begin{aligned}\boldsymbol{D}\boldsymbol{T}_m\boldsymbol{D} &= (\boldsymbol{I}-\mu\boldsymbol{R})\boldsymbol{T}_m(\boldsymbol{I}-\mu\boldsymbol{R})\\ &\approx \boldsymbol{T}_m - \mu\boldsymbol{R}\boldsymbol{T}_m - \mu\boldsymbol{T}_m\boldsymbol{R}\end{aligned} \tag{9}$$

The second and the third term in (8) can be approximated as follows:

$$\boldsymbol{F}_m\boldsymbol{D}\boldsymbol{D}_{m-1} \;\approx\; \boldsymbol{F}_m\epsilon(m-1), \tag{10}$$

$$\boldsymbol{D}\boldsymbol{D}_{-m-1}\boldsymbol{F}_m \;\approx\; \boldsymbol{F}_m\epsilon(-m-1), \tag{11}$$

where $\epsilon(m)$ is the unit step function. This last approximation is valid when it is assumed that the matrix sequence $\boldsymbol{F}_m$ damps faster than $\boldsymbol{D}_m$ (this is true for small values of μ). We now obtain from (8) with (9), (10) and (11) that

$$\begin{aligned}&\mu\boldsymbol{R}\boldsymbol{T}_m + \mu\boldsymbol{T}_m\boldsymbol{R}\\ &\quad= \boldsymbol{F}_m\epsilon(m-1) + \boldsymbol{F}_m\epsilon(-m-1) + \boldsymbol{F}_0\delta_m\\ &\quad= \boldsymbol{F}_m.\end{aligned}$$

Multiplying both sides by μ and summating from $-\infty$ to ∞ we find

$$\mu^2\sum_m \boldsymbol{R}\boldsymbol{T}_m + \mu^2\sum_m \boldsymbol{T}_m\boldsymbol{R} \;=\; \mu\sum_m \boldsymbol{F}_m$$

from which the Lyapounov equation in (3) follows directly.