AN APPROACH TO LMS ADAPTIVE FILTERING WITHOUT USE OF THE INDEPENDENCE ASSUMPTION

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

ABSTRACT

Without use of the well-known "independence assumption" an exact analysis of the LMS-type tapped-delay line adaptive filter is provided, valid for small adaptation constants. For arbitrarily coloured excitations, the steady-state weight-error correlation matrix satisfies a Lyapounov equation, which under special conditions admits a closed-form solution.

1 INTRODUCTION

In the past twenty years, the basic theory of LMS-type adaptive filtering using a tapped-delay line (TDL) structure has been throughout based on an "independence assumption" [1] stating statistical independence of successive input vectors. But this assumption is questionable: within an updating cycle all input vector components are merely shifted to the next place with the last component removed and the first component renewed. Such a strong deterministic coherence between successive input vectors in a TDL structure obviously conflicts with the independence assumption. Nevertheless, justified by a lack of competitive methods, the assumption was and is still widely accepted, despite its lack of consistency and despite the general, growing awareness of this deficiency. Concerning the assumption Gardner [2] states that "in order for such a relatively comprehensive analysis to be tractable, there is one simplifying assumption that cannot be removed" thus expressing a general feeling that it is indispensable for any analytic approach of the LMS algorithm. It leads to conclusions that agree fairly with experimental observations, particularly for small "adaptation constants", and can be supported by a number of sophisticated plausibility arguments [3].

Recently, two ways have been proposed to avoid the independence assumption, thus liberating adaptive filter theory from an unsatisfactory tool and enabling a logically consistent teaching in this field. The first way owing to Douglas et al. [4,5] provides an exact computer-aided mean and mean-square performance analysis, which, however, becomes rather laborious for multi-tap filters. The second method [6] yields analytic results, but is confined to the limit of small adaptation constants. A generalization set up as a power series in terms of the adaptation constant soon becomes intractable [7]; in fact, it predicts not more than some (experimentally verified) weak higher-order effects under specific operating conditions. Thus only the zero-order theory pertinent to the theoretical limit of a vanishing adaptation constant seems to deserve sufficient consideration. This confinement is supported by the broad range of validity of the zero-order solution. In fact it provides reliable results for all adaptation constants sufficiently distant from the stability boundary. Remarkably, various statements of the zero-order theory are not confirmed by the independence theory which, therefore, cannot claim general validity even for a vanishing adaptation constant. It is only under rather special assumptions concerning the spectral distribution of the exciting signals, that the two approaches arrive at the same results.

In the present paper we address the zero-order theory for an LMS adaptive filter of the TDL type under excitation by stationary (input and reference) signals of any colouring. We concentrate upon the steady state, in which the weight coefficients remain fluctuating after completion of the adaptation phase. Thus we do not address adaptation transients and tracking problems. Extending the general outline presented in [6] and adopted in [1] we derive a Lyapounov equation for the "weight-error correlation matrix" in the frequency and time domain, with and without using the eigenvectors of the input correlation matrix. Under special conditions the equation is shown to have a closed-form solution. Additional attention is paid to the external effects of the weight fluctuations commonly summarized under the name "misadjustment".

2 BASIC DYNAMICS AND SMALL-SIGNAL APPROXIMATION

Consider a configuration, in which an adaptive filter tries to imitate a reference filter. The filters, both of the TDL type, are assumed to have equal length M with a constant $M \times 1$ weight vector $\underline{h}$ of the reference filter and a time-varying $M \times 1$ weight vector $\underline{w}_k = \underline{h} + \underline{v}_k$ of the adaptive filter (for unequal lengths we have to write $\underline{w}_k = \underline{h}^{(w)} + \underline{v}_k$, where $\underline{h}^{(w)}$ denotes the Wiener solution pertinent to $\underline{h}$). Both filters are excited by the common "input signal" x_k. The output of the reference filter is superimposed by an external "reference signal" n_k that, after subtraction of the output signal y_k of the adaptive filter, yields the "error signal" e_k. The input and the reference signal x_k, n_k are assumed to be sample functions of statistically independent, stationary zero-mean random processes with unspecified colouring. If at $k = 0$ these random signals are applied to the system and if $\underline{w}_0 \neq \underline{h}$, an adaptation process is initiated which, in global terms, directs the weight $\underline{w}_k$ towards $\underline{h}$. However, $\underline{w}_k$ does not reach $\underline{h}$ asymptotically as a limiting value, but oscillates around it with random fluctuations $\underline{v}_k$. Eventually also this $M \times 1$ "weight error" vector $\underline{v}_k$ becomes a stationary, zero-mean random signal, whose statistics are the main subject of the present paper. Specifically we study the $M \times M$ weight-error correlation matrix

$$V = E\{\underline{v}_k \underline{v}_k^t\}, \tag{1}$$

whose diagonal elements V_{mm} denote the "powers" of the pertinent weight fluctuations, while the off-diagonal elements V_{mn} stand for the mutual correlations. Like any other correlation matrix, V is symmetric ($V = V^t$) and positive (semi-)definite ($V \geq 0$).

For further use we define the $M \times 1$ input vector

$$\underline{x}_k = (x_k, x_{k-1}, \ldots, x_{k-M+1})^t \tag{2}$$

made up of the scalar input signal and its $(M-1)$ past values. The output signal is defined as the inner product

$$y_k = \underline{w}_k^t \underline{x}_k = \underline{h}^t \underline{x}_k + \underline{v}_k^t \underline{x}_k, \tag{3}$$

while the error signal is given by

$$e_k = n_k + \underline{h}^t \underline{x}_k - y_k = n_k - \underline{v}_k^t \underline{x}_k. \tag{4}$$

Now we discuss the weight updating rule, which for the LMS algorithm reads as

$$\underline{v}_{k+1} = \underline{v}_k + 2\mu e_k \underline{x}_k = \underline{v}_k + 2\mu(\, n_k \underline{x}_k - \underline{x}_k \underline{x}_k^t \underline{v}_k\,), \tag{5}$$

where μ is the adaptation constant. Further analysis is eased by making use of the normalized signals $\sqrt{2\mu}n_k$ and $\sqrt{2\mu}\underline{x}_k$, which, for the sake of simplicity, are again denoted by n_k and $\underline{x}_k$, respectively. The statement "μ is small", henceforth often tacitly presupposed, is then phrased as "the power of $\underline{x}_k$ is small". (The concomitant statement concerning n_k is of minor importance because of the linear dependence of $\underline{v}_k$ on n_k). After normalization (5) passes into

$$\underline{v}_{k+1} = \underline{v}_k - \underline{x}_k \underline{x}_k^t \underline{v}_k + n_k \underline{x}_k. \tag{6}$$

This relation defines a deterministic operator $(n_k, \underline{x}_k) \rightarrow (\underline{v}_k)$ such that $\underline{v}_k$ is uniquely determined by the past values of n_k and $\underline{x}_k$. Remember, however, that only the scalar input signal x_k can be freely chosen, which results in an inherent coherence in the vector signal $\underline{x}_k$.

With $R_k = \underline{x}_k \underline{x}_k^t$ and $\underline{f}_k = n_k \underline{x}_k$ the system under consideration belongs to a more general class governed by the difference equation

$$\underline{v}_{k+1} = \underline{v}_k - R_k \underline{v}_k + \underline{f}_k. \tag{7}$$

Again, we are interested in the steady state, where besides R_k and $\underline{f}_k$ also $\underline{v}_k$ is a stationary random signal. In conformity with our special situation, the $M \times 1$ excitation vector $\underline{f}_k$ is assumed to have zero mean, while R_k is a symmetric, positive (semi-)definite time-varying $M \times M$ matrix with the mean value $R = E\{R_k\}$ which, like R_k, is positive (semi-)definite. In our special case R means the "input correlation matrix".

It is important to recognize that, due to $R_k \geq 0$, the term $(-R_k v_k)$ in (7) represents a time-dependent system damping. Remembering our aim to study the filter behaviour for small adaptation constants, i.e. for small input signals, we have to examine the limiting case where this damping R_k and herewith R is small (compared to the unit matrix). Then the system behaves as an extreme low-pass filter [7] implying that the variations of $\underline{v}_k$ are much slower than those contained in $\underline{f}_k$ and R_k. Consequently, the time-dependent damping R_k in (7) can be replaced with its average R:

$$\underline{v}_{k+1} = \underline{v}_k - R\underline{v}_k + \underline{f}_k. \tag{8}$$

Thus for $R \rightarrow 0$ we can solve the simple difference equation (8) with constant coefficients instead of (7). For moderate values of R (i.e. for moderate values of the adaptation constant μ) this zero-order solution approximately solves (7).

From (8) it can be concluded that in the limiting case of a vanishing R the required $M \times M$ weight-error correlation matrix V as given by (1) satisfies the Lyapounov equation

$$RV + VR = F. \tag{9}$$

Here F denotes an "excitation matrix" defined as

$$F = \sum_{l=-\infty}^{\infty} E\{\underline{f}_k \underline{f}_{k-l}^t\} = \sum_{l=-\infty}^{\infty} E\{n_k n_{k-l}\} E\{\underline{x}_k \underline{x}_{k-l}^t\}. \tag{10}$$

Like R, it is symmetric, positive definite, and of Toeplitz structure. In the next two sections proofs are provided of (9), one in the frequency domain using the eigenvectors of R, another in the time domain without using the eigenvector representation.

3 FREQUENCY-DOMAIN TREATMENT

The signal transformation $\underline{f}_k \rightarrow \underline{v}_k$ pertinent to (8) can be elegantly formulated in the frequency domain. Denoting frequency functions by capitals with a tilde, we find for the associated matrix system function

$$\tilde{H}(z) = ((z-1)I + R)^{-1}. \tag{11}$$

Let $\tilde{F}(\Omega)$, $\tilde{V}(\Omega)$ denote the matrix power spectral densities of $\underline{f}_k$ and $\underline{v}_k$ (i.e. the Fourier transforms of the autocorrelations $F^{(l)} = E\{\underline{f}_k \underline{f}_{k-l}^t\}$ and $V^{(l)} = E\{\underline{v}_k \underline{v}_{k-l}^t\}$), then we have the "local" spectral relationship

$$\tilde{V}(\Omega) = \tilde{H}(e^{j\Omega})\tilde{F}(\Omega)\tilde{H}^H(e^{j\Omega}). \tag{12}$$

Notice that, in contrast with the familiar scalar case, the factors in (12) do not commute, implying that the central factor cannot be extracted from the product. The required weight-error correlation matrix V is found by averaging the pertinent spectrum over all frequencies:

$$V = \frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{V}(\Omega)d\Omega. \tag{13}$$

Now let $\underline{q}_i$ and λ_i denote the orthonormal eigenvectors and positive eigenvalues of the input correlation matrix R, then we can formulate $\tilde{H}$, $\tilde{F}$, $\tilde{V}$ in normal coordinates, i.e. with respect to the basis of eigenvectors:

$$\tilde{H} = \sum_i \sum_j \underline{q}_i \tilde{H}_{ij} \underline{q}_j^t, \qquad \tilde{H}_{ij} = \underline{q}_i^t \tilde{H} \underline{q}_j, \tag{14}$$

with similar expressions for $\tilde{F}$ and $\tilde{V}$. A special result is found for the system matrix $\tilde{H}$, where the representation (14) becomes diagonal: $\tilde{H}_{ij} = 0$ for $i \neq j$, in accordance with (11). The diagonal elements for $z = e^{j\Omega}$ read as

$$\tilde{H}_{ii} = (e^{j\Omega} - 1 + \lambda_i)^{-1}. \tag{15}$$

Next we rewrite the basic relation (12) in normal coordinates, observing the low-pass character of $\tilde{H}_{ii}$:

$$\begin{aligned} \tilde{V}_{ij}(\Omega) &= \tilde{H}_{ii}(e^{j\Omega})\tilde{F}_{ij}(\Omega)\tilde{H}_{jj}^*(e^{j\Omega}) \\ &\approx \tilde{H}_{ii}(e^{j\Omega})\tilde{H}_{jj}^*(e^{j\Omega})\tilde{F}_{ij}(0). \end{aligned} \tag{16}$$

Performing the Ω-integration (13) we obtain

$$V_{ij} = \frac{\tilde{F}_{ij}(0)}{\lambda_i + \lambda_j - \lambda_i\lambda_j} \approx \frac{\tilde{F}_{ij}(0)}{\lambda_i + \lambda_j},$$

valid for small λ's, i.e. for small input signals. Rewriting this "local" proportionality between V_{ij} and $\tilde{F}_{ij}(0)$ as

$$(\lambda_i + \lambda_j)V_{ij} = \tilde{F}_{ij}(0) \tag{17}$$

we recognize (9) in normal-coordinate formulation. Indeed, (17) follows from (9) through simultaneous pre- and post-multiplication by $\underline{q}_i^t$ and $\underline{q}_j$ recognizing that F in (10) represents the low-frequency value $\tilde{F}(0)$ of the power spectrum of $\underline{f}_k$. The higher frequencies are lost due to the low-pass character of the adaptive system.

4 TIME-DOMAIN TREATMENT

The solution of (8) has the convolutional form

$$\underline{v}_k = H_k \otimes \underline{f}_k = \sum_{m=-\infty}^{\infty} H_{k-m}\underline{f}_m, \tag{18}$$

with H_k the $M \times M$ matrix impulse response of the system. This is determined from (8) as

$$H_k = u_{k-1}(I-R)^{k-1} = H_k^t, \tag{19}$$

where u_k denotes the unit step (u_k=1 for $k \geq 0$ and zero elsewhere).

In our context we are interested in the transmission of stationary stochastic signals, particularly in terms of their autocorrelations $F^{(l)} = E\{\underline{f}_k \underline{f}_{k-l}^t\}$, $V^{(l)} = E\{\underline{v}_k \underline{v}_{k-l}^t\}$. We find a linear relationship of the form

$$V^{(l)} = \sum_m \sum_n H_m F^{(n)} H_{m+n-l}^t, \tag{20}$$

from which we obtain the desired "weight-error correlation matrix" V:

$$V = V^{(0)} = \sum_m \sum_n H_m F^{(n)} H_{m+n}^t. \tag{21}$$

Next we decompose (21) in the form

$$V = \sum_n T^{(n)}; \quad T^{(n)} = \sum_m H_m F^{(n)} H_{m+n}^t. \tag{22}$$

In an attempt to sum up the second series (22) we encounter the difficulty that the matrices H_m and $F^{(n)}$ in general do not commute, thus prohibiting the extraction of $F^{(n)}$ from the sum. In fact (22) does not admit an explicit summation. Instead we show that in the small-signal approximation $T^{(n)}$ satisfies a Lyapounov equation. To this end we write (19) in the recursive form (where the Dirac function δ_m equals unity for $m = 0$ and zero elsewhere)

$$H_{m+1} = (I-R)H_m + \delta_m I = H_m(I-R) + \delta_m I. \tag{23}$$

This enables us to rewrite (22) as

$$\begin{aligned} T^{(n)} &= \sum_m H_m F^{(n)} H_{m+n}^t \;=\; \sum_m H_{m+1} F^{(n)} H_{m+n+1}^t \\ &= \sum_m [(I-R)H_m + \delta_m I]F^{(n)}[(I-R)H_{m+n}^t + \delta_{m+n} I] \\ &= (I-R)T^{(n)}(I-R) + F^{(n)}(I-R)H_n^t + \\ &\quad (I-R)H_{-n}F^{(n)} + \delta_n F^{(n)}. \end{aligned} \tag{24}$$

For small input signals, i.e. for a small R, we can neglect the terms $RT^{(n)}R$, $-F^{(n)}RH_n^t$, $-RH_{-n}F^{(n)}$, and approximate $F^{(n)}H_n^t$ and $H_{-n}F^{(n)}$ by $u_{n-1}F^{(n)}$ and $u_{1-n}F^{(n)}$, respectively. Thus we arrive at the Lyapounov equation

$$RT^{(n)} + T^{(n)}R = F^{(n)}. \tag{25}$$

Finally, summing up (25) over all n and using (22) and (10), we find (9). Notice that the above derivation does not explicitly use the eigenvector representation of R.

Due to the statistical independence of n_k and $\underline{x}_k$ the autocorrelation of $\underline{f}_k$ equals the product of those of n_k

and $\underline{x}_k$. With Parseval's theorem the sum over all $F^{(l)}$, as required in (10), equals the average of the product of the pertinent power spectra. This implies that all frequencies contained in n_k and $\underline{x}_k$ equally contribute to the right-hand term of (10) representing the zero-frequency spectral component of $\underline{f}_k$.

5 EXPLICITLY SOLVABLE CASES

There are only few cases, which admit a closed-form solution of (9). One occurs for a white process n_k, where the sum (10) consists of the single term $E\{n_k^2\}R$ which leads to

$$V = \tfrac{1}{2}E\{n_k^2\}I. \tag{26}$$

Thus the weight fluctuations are uncorrelated and have equal power $\frac{1}{2}E\{n_k^2\}$. This result is also predicted by the independence assumption [1]. Notice that the power has a finite limit for $R \to 0$, i.e. for a vanishing input signal $\underline{f}_k$ in (8). This can be explained by an amplifying property of the low-pass system, which, due to (11), has a high "resonance peak" at $z = 1$: $\tilde{H}(1) = R^{-1}$.

Another explicitly solvable case is found for a white x_k. Then $R = E\{x_k^2\}I$ and F and V become Toeplitz matrices with $F_{mn} = E\{x_k^2\}E\{n_k n_{k-(m-n)}\}$ and $V_{mn} = \frac{1}{2}E\{n_k n_{k-(m-n)}\}$. Like the former situation all weight fluctuations have the same power $\frac{1}{2}E\{n_k^2\}$, while the correlation between two weight fluctuations equals the autocorrelation of n_k with the time shift replaced with the spatial distance between the points under consideration. In this sense the autocorrelation of n_k is mapped into a spatial distribution on the delay line.

Returning to the more general case we observe that the correlation matrices R, F, and V are symmetric and positive definite. With regard to V, it is a basic property of the Lyapounov equation that it transfers these properties from the given matrices R, F to the unknown V. However, this is not true for the Toeplitz property. It is only for very large filter lengths M that V becomes almost Toeplitz (i.e. except for the vicinity of the matrix border). A general consequence of the Toeplitz property of R, F is a double symmetry in V with respect to the main and the side diagonal [7].

6 MISADJUSTMENT

We close with a brief discussion of the external effects of the fluctuating filter weights usually designated as "misadjustment". According to [8] this should not be based upon the excess mean-squared error, but be directly related to the partial output signal owing to the weight fluctuations:

$$\text{misadjustment} \approx \frac{E\{[\underline{v}_k^t \underline{x}_k]^2\}}{E\{n_k^2\}}, \tag{27}$$

where the numerator can be rewritten as

$$E\{\underline{v}_k^t \underline{x}_k \underline{x}_k^t \underline{v}_k\} \approx E\{\underline{v}_k^t E\{\underline{x}_k \underline{x}_k^t\}\underline{v}_k\} = E\{\underline{v}_k^t R \underline{v}_k\}. \tag{28}$$

Here we have exploited the fact that $\underline{x}_k$ and $\underline{v}_k$ fluctuate on extremely different time scales. During some time interval of length N, $\underline{v}_k$ can be considered as constant, whereas N is sufficient to evaluate the time average with respect to $\underline{x}_k$ (which equals the ensemble average, due to ergodicity). This leads to a local $E\{[\underline{v}_k^t \underline{x}_k]^2\}$ pertinent to the "frozen" $\underline{v}_k$. In a second averaging operation with respect to $\underline{v}_k$ the final $E\{[\underline{v}_k^t \underline{x}_k]^2\}$ is determined. This runs as follows. With the identities

$$\underline{v}_k^t R \underline{v}_k = \mathrm{tr}(R\underline{v}_k \underline{v}_k^t) = \mathrm{tr}(\underline{v}_k \underline{v}_k^t R) = \tfrac{1}{2}\mathrm{tr}(R\underline{v}_k \underline{v}_k^t + \underline{v}_k \underline{v}_k^t R)$$

and using (9) we find

$$\begin{aligned} &E\{[\underline{v}_k^t \underline{x}_k]^2\} \\ &= \tfrac{1}{2}\mathrm{tr}(RV + VR) = \tfrac{1}{2}\sum_l \mathrm{tr} F^{(l)} = \tfrac{1}{2}\sum_l E\{\underline{f}_k^t \underline{f}_{k-l}\} \\ &= \tfrac{1}{2}\sum_l E\{n_k n_{k-l}\}E\{\underline{x}_k^t \underline{x}_{k-l}\} \\ &= \tfrac{1}{2}M\sum_l E\{n_k n_{k-l}\}E\{x_k x_{k-l}\}, \end{aligned}$$

which (formulated in normalized terms) leads to

$$\text{misadjustment} = \frac{\frac{1}{2}M\sum_l E\{n_k n_{k-l}\}E\{x_k x_{k-l}\}}{E\{n_k^2\}}. \tag{29}$$

Observe that the sum in the numerator can be rewritten as the average over the spectra of the input and the reference signal [8].

References

[1] S. Haykin, Adaptive filter theory, Third Edition, Prentice-Hall Int. London 1996, Chapter 9, Appendix I.

[2] W.A. Gardner, "Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique", Signal Processing, vol. 6, pp. 113-133, 1984.

[3] J.E. Mazo, "On the independence theory of equalizer convergence", Bell System Technical J., vol. 58, pp. 963-993, 1979.

[4] S.C. Douglas and T.H.Y. Meng, "Exact expectation analysis of the LMS adaptive filter without the independence assumption", IEEE Proc. Int. Conf. Acoustics, Speech, and Signal Processing, vol. 4, pp. 61-64, 1992.

[5] S.C. Douglas and Weimin Pan, "Exact expectation analysis of the LMS adaptive filter", IEEE Trans. Signal Processing, vol. SP-43, no. 12, pp. 2863-2871, Dec. 1995.

[6] H.J. Butterweck, "A steady-state analysis of the LMS adaptive algorithm without use of the independence assumption", IEEE Proc. Int. Conf. on Acoustics, Speech, and Signal Processing, Detroit 1995, pp. 1404-1407.

[7] H.J. Butterweck, "Iterative analysis of the steady-state weight fluctuations in LMS-type adaptive filters", Eindhoven University of Technology, Report 96-E-299, ISBN 90-6144-299-0, June 1996.

[8] H.J. Butterweck, "A new interpretation of the misadjustment in adaptive filtering", IEEE Proc. Int. Conf. on Acoustics, Speech, and Signal Processing, Atlanta 1996, pp. 1641-1643.