We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Transient Analysis of a Combination of Two Adaptive Filters

Tõnu Trump

Department of Radio and Telecommunication Engineering,
Tallinn University of Technology
Estonia

1. Introduction

The Least Mean Square (LMS) algorithm is probably the most popular adaptive algorithm. The algorithm has since its introduction in Widrow & Hoff (1960) been widely used in many applications like system identification, communication channel equalization, signal prediction, sensor array processing, medical applications, echo and noise cancellation etc. The popularity of the algorithm is due to its low complexity but also due to its good properties like e.g. robustness Haykin (2002); Sayed (2008). Let us explain the algorithm in the example of system identification.
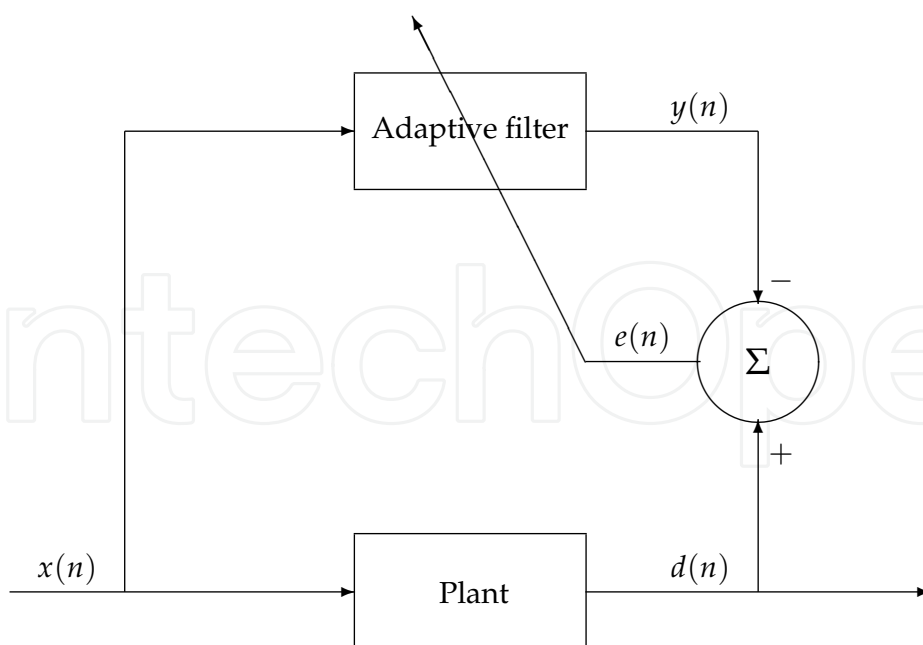

Fig. 1. Usage of an adaptive filter for system identification.

As seen from Figure 1, the input signal to the unknown plant, is $x(n)$ and the output signal from the plant is $d(n)$. We call the signal $d(n)$ the desired signal. The signal $d(n)$ also contains noise and possible nonlinear effects of the plant. We would like to estimate the impulse

response of the plant observing its input and output signals. To do so we connect an adaptive filter in parallel with the plant. The adaptive filter is a linear filter with output signal $y(n)$. We then compare the output signals of the plant and the adaptive filter and form the error signal $e(n)$. Obviously one would like have the error signal to be as small as possible in some sense. The LMS algorithm achieves this by minimizing the mean squared error but doing this in instantaneous fashion. If we collect the impulse response coefficients of our adaptive filter computed at iteration n into a vector $\mathbf{w}(n)$ and the input signal samples into a vector $\mathbf{x}(n)$, the LMS algorithm updates the weight vector estimate at each iteration as

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu e^*(n)\mathbf{x}(n), \tag{1}$$

where μ is the step size of the algorithm. One can see that the weight update is in fact a low pass filter with transfer function

$$H(z) = \frac{\mu}{1 - z^{-1}} \tag{2}$$

operating on the signal $e^*(n)\mathbf{x}(n)$. The step size determines in fact the extent of initial averaging performed by the algorithm. If μ is small, only a little of new information is passed into the algorithm at each iteration, the averaging is thus over a large number of samples and the resulting estimate is more reliable but building the estimate takes more time. On the other hand if μ is large, a lot of new information is passed into the weight update each iteration, the extent of averaging is small and we get a less reliable estimate but we get it relatively fast.

When designing an adaptive algorithm, one thus faces a trade–off between the initial convergence speed and the mean–square error in steady state. In case of algorithms belonging to the Least Mean Square family this trade–off is controlled by the step-size parameter. Large step size leads to a fast initial convergence but the algorithm also exhibits a large mean–square error in the steady state and in contrary, small step size slows down the convergence but results in a small steady state error.

Variable step size adaptive schemes offer a possible solution allowing to achieve both fast initial convergence and low steady state misadjustment Arenas-Garcia et al. (1997); Harris et al. (1986); Kwong & Johnston (1992); Matthews & Xie (1993); Shin & Sayed (2004). How successful these schemes are depends on how well the algorithm is able to estimate the distance of the adaptive filter weights from the optimal solution. The variable step size algorithms use different criteria for calculating the proper step size at any given time instance. For example squared instantaneous errors have been used in Kwong & Johnston (1992) and the squared autocorrelation of errors at adjacent time instances have been used in Arenas-Garcia et al. (1997). The reference Matthews & Xie (1993) ivestigates an algorithm that changes the time–varying convergence parameters in such a way that the change is proportional to the negative of gradient of the squared estimation error with respect to the convergence parameter. In reference Shin & Sayed (2004) the norm of projected weight error vector is used as a criterion to determine how close the adaptive filter is to its optimum performance.

Recently there has been an interest in a combination scheme that is able to optimize the trade–off between convergence speed and steady state error Martinez-Ramon et al. (2002). The scheme consists of two adaptive filters that are simultaneously applied to the same inputs as depicted in Figure 2. One of the filters has a large step size allowing fast convergence and the other one has a small step size for a small steady state error. The outputs of the filters are combined through a mixing parameter λ. The performance of this scheme has been studied for some parameter update schemes Arenas-Garcia et al. (2006); Bershad et al. (2008);

Candido et al. (2010); Silva et al. (2010). The reference Arenas-Garcia et al. (2006) uses convex combination i.e. λ is constrained to lie between 0 and 1. The references Silva et al. (2010) and Candido et al. (2010) present transient analysis of a slightly modified versions of this scheme. The parameter λ is in those papers found using an LMS type adaptive scheme and possibly computing the sigmoidal function of the result. The reference Bershad et al. (2008) takes another approach computing the mixing parameter using an affine combination. This paper uses the ratio of time averages of the instantaneous errors of the filters. The error function of the ratio is then computed to obtain λ.

In Mandic et al. (2007) a convex combination of two adaptive filters with different adaptation schemes has been investigated with the aim to improve the steady state characteristics. One of the adaptive filters in that paper uses LMS algorithm and the other one Generalized Normalized Gradient Decent algorithm. The combination parameter λ is computed using stochastic gradient adaptation. In Zhang & Chambers (2006) the convex combination of two adaptive filters is applied in a variable filter length scheme to gain improvements in low SNR conditions. In Kim et al. (2008) the combination has been used to join two affine projection filters with different regularization parameters. The work Fathiyan & Eshghi (2009) uses the combination on parallel binary structured LMS algorithms. These three works use the LMS like scheme of Azpicueta-Ruiz et al. (2008b) to compute λ.

It should be noted that schemes involving two filters have been proposed earlier Armbruster (1992); Ochiai (1977). However, in those early schemes only one of the filters have been adaptive while the other one has used fixed filter weights. Updating of the fixed filter has been accomplished by copying of all the coefficients from the adaptive filter, when the adaptive filter has been performing better than the fixed one.

In this chapter we compute the mixing parameter λ from output signals of the individual filters. The scheme was independently proposed in Trump (2009a) and Azpicueta-Ruiz et al. (2008a), the steady state performance of it was investigated in Trump (2009b) and the tracking performance in Trump (2009c). The way of calculating the mixing parameter is optimal in the sense that it results from minimization of the mean-squared error of the combined filter. In the main body of this chapter we present a transient analysis of the algorithm. We will assume throughout the chapter that the signals are complex–valued and that the combination scheme uses two LMS adaptive filters. The italic, bold face lower case and bold face upper case letters will be used for scalars, column vectors and matrices respectively. The superscript $*$ denotes complex conjugation and H Hermitian transposition of a matrix. The operator $E[\cdot]$ denotes mathematical expectation, $tr[\cdot]$ stands for trace of a matrix and $\mathcal{Re}\{\cdot\}$ denotes the real part of a complex variable.

2. Algorithm

Let us consider two adaptive filters, as shown in Figure 2, each of them updated using the LMS adaptation rule

$$\mathbf{w}_i(n) = \mathbf{w}_i(n-1) + \mu_i e_i^*(n)\mathbf{x}(n), \tag{3}$$

$$e_i(n) = d(n) - \mathbf{w}_i^H(n-1)\mathbf{x}(n), \tag{4}$$

$$d(n) = \mathbf{w}_o^H\mathbf{x}(n) + v(n). \tag{5}$$

In the above equations the vector $\mathbf{w}_i(n)$ is the length N vector of coefficients of the i-th adaptive filter, with $i = 1, 2$. The vector $\mathbf{w}_o$ is the true weight vector we aim to identify with our adaptive scheme and $\mathbf{x}(n)$ is the N input vector, common for both of the adaptive filters.
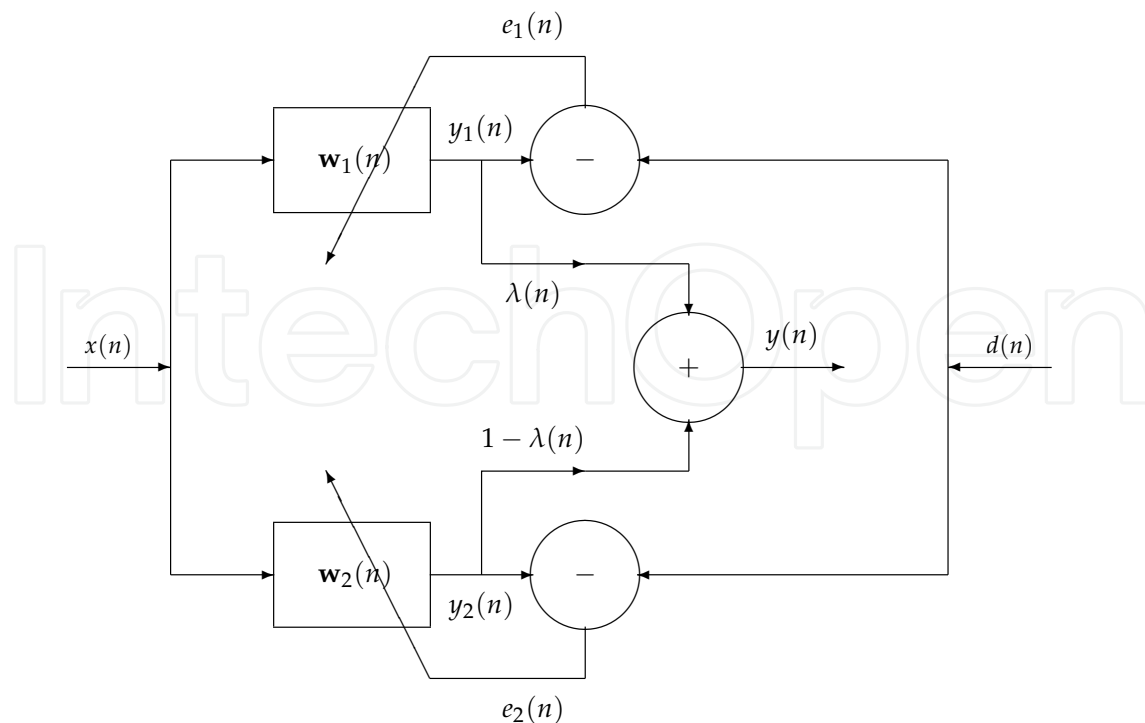
Fig. 2. The combined adaptive filter.

The input process is assumed to be a zero mean wide sense stationary Gaussian process. The desired signal $d(n)$ is a sum of the output of the filter to be identified and the Gaussian, zero mean i.i.d. measurement noise $v(n)$. We assume that the measurement noise is statistically independent of all the other signals. μ_i is the step size of i–th adaptive filter. We assume without loss of generality that $\mu_1 > \mu_2$. The case $\mu_1 = \mu_2$ is not interesting as in this case the two filters remain equal and the combination renders to a single filter.

The outputs of the two adaptive filters are combined according to

$$y(n) = \lambda(n)y_1(n) + [1 - \lambda(n)]y_2(n), \tag{6}$$

where $y_i(n) = \mathbf{w}_i^H(n-1)\mathbf{x}(n)$ and the mixing parameter λ can be any real number.

We define the *a priori* system error signal as difference between the output signal of the true system at time n, given by $y_o(n) = \mathbf{w}_o^H\mathbf{x}(n) = d(n) - v(n)$, and the output signal of our adaptive scheme $y(n)$

$$e_a(n) = y_o(n) - \lambda(n)y_1(n) - (1 - \lambda(n))y_2(n). \tag{7}$$

Let us now find $\lambda(n)$ by minimizing the mean square of the *a priori* system error. The derivative of $E[|e_a(n)|^2]$ with respect to $\lambda(n)$ reads

$$\begin{aligned}
\frac{\partial E[|e_a(n)|^2]}{\partial \lambda(n)} &= 2E[(y_o(n) - \lambda(n)y_1(n) - (1 - \lambda(n))y_2(n)) \\
&\quad \cdot (-y_1(n) + y_2(n))^*] \\
&= 2E[\mathcal{R}e\{(y_o(n) - y_2(n))(y_2(n) - y_1(n))^*\} \\
&\quad + \lambda(n)|(y_2(n) - y_1(n))|^2].
\end{aligned}$$

Setting the derivative to zero results in

$$\lambda(n) = \frac{E[\mathcal{R}e\{(d(n) - y_2(n))(y_1(n) - y_2(n))^*\}]}{E[|(y_1(n) - y_2(n))|^2]}, \tag{8}$$

where we have replaced the true system output signal $y_o(n)$ by its observable noisy version $d(n)$. Note however, that because we have made the standard assumption that the input signal $\mathbf{x}(n)$ and measurement noise $v(n)$ are independent random processes, this can be done without introducing any error into our calculations.

The denominator of equation (8) comprises expectation of the squared difference of the two filter output signals. This quantity can be very small or even zero, particularly in the beginning of adaptation if the two step sizes are close to each other. Correspondingly λ computed directly from (8) may be large. To avoid this from happening we add a small regularization constant ϵ to the denominator of (8).

3. Transient analysis

In this section we are interested in finding expressions that characterize transient performance of the combined algorithm i.e. we intend to derive formulae that characterize entire course of adaptation of the algorithm. Before we can proceed we need, however, to introduce some notations. First let us denote the weight error vector of i–th filter as

$$\tilde{\mathbf{w}}_i(n) = \mathbf{w}_o - \mathbf{w}_i(n). \tag{9}$$

Then the equivalent weight error vector of the combined adaptive filter will be

$$\tilde{\mathbf{w}}(n) = \lambda \tilde{\mathbf{w}}_1(n) + (1 - \lambda)\tilde{\mathbf{w}}_2(n). \tag{10}$$

The mean square deviation of the combined filter is given by

$$\begin{aligned} MSD = E[\tilde{\mathbf{w}}^H(n)\tilde{\mathbf{w}}(n)] = {} & \lambda^2 E[\tilde{\mathbf{w}}_1^H(n)\tilde{\mathbf{w}}_1(n)] \\ & + 2\lambda(1 - \lambda)\mathcal{R}e\{E[\tilde{\mathbf{w}}_2^H(n)\tilde{\mathbf{w}}_1(n)]\} \\ & + (1 - \lambda)^2 E[\tilde{\mathbf{w}}_2^H(n)\tilde{\mathbf{w}}_2(n)]. \end{aligned} \tag{11}$$

The *a priori* estimation error of an individual filter is defined as

$$e_{i,a}(n) = \tilde{\mathbf{w}}_i^H(n-1)\mathbf{x}(n). \tag{12}$$

It follows from (7) that we can express the *a priori* error of the combination as

$$e_a(n) = \lambda(n)e_{1,a}(n) + (1 - \lambda(n))\,e_{2,a}(n) \tag{13}$$

and because $\lambda(n)$ is according to (8) a ratio of mathematical expectations and, hence, deterministic, we have for the excess mean square error of the combination

$$E[|e_a(n)|^2] = \lambda^2 E[|e_{1,a}(n)|^2] + 2\lambda(1 - \lambda)E[\mathcal{R}e\{e_{1,a}(n)e_{2,a}^*(n)\}] + (1 - \lambda)^2 E[|e_{2,a}(n)|^2]. \tag{14}$$

As $e_{i,a}(n) = \tilde{\mathbf{w}}_i^H(n-1)\mathbf{x}(n)$, the expression of the excess mean square error becomes

$$\begin{aligned} E[|e_a(n)|^2] = {} & \lambda^2 E[\tilde{\mathbf{w}}_1^H(n-1)\mathbf{x}\mathbf{x}^H\tilde{\mathbf{w}}_1(n-1)] \\ & + 2\lambda(1 - \lambda)E[\mathcal{R}e\{\tilde{\mathbf{w}}_1^H(n-1)\mathbf{x}\mathbf{x}^H\tilde{\mathbf{w}}_2(n-1)\}] \\ & + (1 - \lambda)^2 E[\tilde{\mathbf{w}}_2^H(n-1)\mathbf{x}\mathbf{x}^H\tilde{\mathbf{w}}_2(n-1)]. \end{aligned} \tag{15}$$

In what follows we often drop the explicit time index n as we have done in (15), if it is not necessary to avoid a confusion.

Noting that $y_i(n) = \mathbf{w}_i^H(n-1)\mathbf{x}(n)$, we can rewrite the expression for $\lambda(n)$ in (8) as

$$\lambda(n) = \frac{E[\tilde{\mathbf{w}}_2^H \mathbf{x}\mathbf{x}^H \tilde{\mathbf{w}}_2] - E[\mathcal{R}e\{\tilde{\mathbf{w}}_2^H \mathbf{x}\mathbf{x}^H \tilde{\mathbf{w}}_1\}]}{E[\tilde{\mathbf{w}}_1^H \mathbf{x}\mathbf{x}^H \tilde{\mathbf{w}}_1] - 2E[\mathcal{R}e\{\tilde{\mathbf{w}}_1^H \mathbf{x}\mathbf{x}^H \tilde{\mathbf{w}}_2\}] + E[\tilde{\mathbf{w}}_2^H \mathbf{x}\mathbf{x}^H \tilde{\mathbf{w}}_2]}. \tag{16}$$

We thus need to investigate the evolution of the individual terms of the type $EMSE_{k,l} = E[\tilde{\mathbf{w}}_k^H(n-1)\mathbf{x}(n)\mathbf{x}^H(n)\tilde{\mathbf{w}}_l(n-1)]$ in order to reveal the time evolution of $EMSE(n)$ and $\lambda(n)$. To do so we, however, concentrate first on the mean square deviation defined in (11).

For a single LMS filter we have after subtraction of (3) from $\mathbf{w}_o$ and expressing $e_i(n)$ through the error of the corresponding Wiener filter $e_o(n)$

$$\tilde{\mathbf{w}}_i(n) = \left(\mathbf{I} - \mu_i \mathbf{x}\mathbf{x}^H\right)\tilde{\mathbf{w}}_i(n-1) - \mu_i \mathbf{x} e_o^*(n). \tag{17}$$

We next approximate the outer product of input signal vectors by its correlation matrix $\mathbf{x}\mathbf{x}^H \approx \mathbf{R_x}$. The approximation is justified by the fact that with small step size the weight error update of the LMS algorithm (17) behaves like a low pass filter with a low cutoff frequency. With this approximations we have

$$\tilde{\mathbf{w}}_i(n) \approx \left(\mathbf{I} - \mu_i \mathbf{R_x}\right)\tilde{\mathbf{w}}_i(n-1) - \mu_i \mathbf{x} e_o^*(n). \tag{18}$$

This means in fact that we apply the small step size theory Haykin (2002) even if the assumption of small step size is not really true for the fast adapting filter. In our simulation study we will see, however, that the assumption works in practice rather well.

Let us now define the eigendecomposition of the correlation matrix as

$$\mathbf{Q}^H \mathbf{R}_x \mathbf{Q} = \mathbf{\Omega}, \tag{19}$$

where $\mathbf{Q}$ is a unitary matrix whose columns are the orthogonal eigenvectors of $\mathbf{R}_x$ and $\mathbf{\Omega}$ is a diagonal matrix having eigenvalues associated with the corresponding eigenvectors on its main diagonal. We also define the transformed weight error vector as

$$\mathbf{v}_i(n) = \mathbf{Q}^H \tilde{\mathbf{w}}_i(n) \tag{20}$$

and the transformed last term of equation (18) as

$$\mathbf{p}_i(n) = \mu_i \mathbf{Q}^H \mathbf{x} e_o^*(n). \tag{21}$$

Then we can rewrite the equation (18) after multiplying both sides by $\mathbf{Q}^H$ from the left as

$$\mathbf{v}_i(n) = \left(\mathbf{I} - \mu_i \mathbf{\Omega}\right)\mathbf{v}_i(n-1) - \mathbf{p}_i(n). \tag{22}$$

We note that the mean of $\mathbf{p}_i$ is zero by the orthogonality theorem and the crosscorrelation matrix of $\mathbf{p}_k$ and $\mathbf{p}_l$ equals

$$E[\mathbf{p}_k \mathbf{p}_l^H] = \mu_k \mu_l \mathbf{Q}^H E[\mathbf{x} e_o^*(n) e_o(n)\mathbf{x}^H]\mathbf{Q}. \tag{23}$$

We now invoke the Gaussian moment factoring theorem to write

$$E[\mathbf{x} e_o^*(n) e_o(n)\mathbf{x}^H] = E[\mathbf{x} e_o^*(n)]E[e_o(n)\mathbf{x}^H] + E[\mathbf{x}\mathbf{x}^H]E[|e_o|^2]. \tag{24}$$

The first term in the above is zero due to the principle of orthogonality and the second term equals $\mathbf{R}J_{min}$. Hence we are left with

$$E[\mathbf{p}_k\mathbf{p}_l^H] = \mu_k\mu_l J_{min}\mathbf{\Omega},\tag{25}$$

where $J_{min} = E[|e_o|^2]$ is the minimum mean square error produced by the corresponding Wiener filter. As the matrices $\mathbf{I}$ and $\mathbf{\Omega}$ in (22) are diagonal, it follows that the m-th element of vector $\mathbf{v}_i(n)$ is given by

$$v_{i,m}(n) = (1 - \mu_i\omega_m)\,v_{i,m}(n-1) - p_{i,m}(n)\tag{26}$$
$$= (1 - \mu_i\omega_m)^n\,v_m(0) + \sum_{i=0}^{n-1}(1 - \mu_i\omega_m)^{n-1-i}\,p_{i,m}(i),$$

where ω_m is the m-th eigenvalue of $\mathbf{R}_x$ and $v_{i,m}$ and $p_{i,m}$ are the m-th components of the vectors $\mathbf{v}_i$ and $\mathbf{p}_i$ respectively.

We immediately see that the mean value of $v_{i,m}(n)$ equals

$$E[v_{i,m}(n)] = (1 - \mu_i\omega_m)^n\,v_m(0)\tag{27}$$

as the vector $\mathbf{p}_i$ has zero mean.

The expected values of $v_{i,m}(n)$ exhibit no oscillations as the correlation matrix $\mathbf{R}$ is positive semidefinite with all its eigenvalues being nonnegative real numbers. In order the LMS algorithm to converge in mean, the weight errors need to decrease with time. This will happen if

$$|1 - \mu_i\omega_m| < 1\tag{28}$$

for all m. In that case all the natural modes of the algorithm die out as the number of iterations n approaches infinity. The condition needs to be fulfilled for all the eigenvalues ω_m and is obviously satisfied if the condition is met for the maximum eigenvalue. It follows that the individual step size μ_i needs to be selected such that the double inequality

$$0 < \mu_i < \frac{2}{\lambda_{max}}\tag{29}$$

is satisfied.

In some applications the input signal correlation matrix and its eigenvalues are not known *a priori*. In this case it may be convenient to use the fact that

$$tr\{\mathbf{R}_x\} = \sum_{i=0}^{N-1} r_x(i,i) = \sum_{i=0}^{N-1} \omega_i > \omega_{max},\tag{30}$$

where $r_x(i,i)$ is the i-th diagonal element of the matrix $\mathbf{R}$. Then we can normalize the step size with the instantaneous estimate of the trace of correlation matrix $\mathbf{x}^H(n)\mathbf{x}(n)$ to get so called normalized LMS algorithm. The normalized LMS algorithm uses the normalized step size

$$\mu_i = \frac{\alpha_i}{\mathbf{x}^H(n)\mathbf{x}(n)}$$

and is convergent if

$$0 < \alpha_i < 2.$$

The normalized LMS is more convenient for practical usage if the properties of the input signal are unknown or are varying in time like this is for example the case with speech signals.

To proceed with our development for the combination of two LMS filters we note that we can express the MSD and its individual components in (11) through the transformed weight error vectors as

$$E[\tilde{\mathbf{w}}_k^H(n)\tilde{\mathbf{w}}_l(n)] = E[\mathbf{v}_k^H(n)\mathbf{v}_l(n)] = \sum_{m=0}^{N-1} E[v_{k,m}(n)v_{l,m}^*(n)] \tag{31}$$

so we also need to find the auto– and cross correlations of v. Let us concentrate on the m-th component in the sum above corresponding to the cross term and denote it as $Y_m = E[v_{k,m}(n)v_{l,m}^*(n)]$. The expressions for the component filters follow as special cases. Substituting (26) into the expression of Y_m above, taking the mathematical expectation and noting that the vector $\mathbf{p}$ is independent of $\mathbf{v}(0)$ results in

$$Y_m = E\left[(1-\mu_k\omega_m)^n v_k(0)(1-\mu_l\omega_m)^n v_l^*(0)\right] \tag{32}$$

$$+E\left[\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}(1-\mu_k\omega_m)^{n-1-i}(1-\mu_l\omega_m)^{n-1-j}p_{k,m}(i)p_{l,m}^*(j)\right].$$

We now note that most likely the two component filters are initialized to the same value

$$v_{k,m}(0) = v_{l,m}(0) = v_m(0)$$

and that

$$E\left[p_{k,m}(i)p_{l,m}^*(j)\right] = \begin{cases} \mu_k\mu_l\omega_m J_{min}, & i = j \\ 0, & \text{otherwise} \end{cases}. \tag{33}$$

We then have for the m-th component of MSD

$$Y_m = (1-\mu_k\omega_m)^n(1-\mu_l\omega_m)^n |v_m(0)|^2 \tag{34}$$

$$+\mu_k\mu_l\omega_m J_{min}(1-\mu_k\omega_m)^{n-1}(1-\mu_l\omega_m)^{n-1}\sum_{i=0}^{n-1}(1-\mu_k\omega_m)^{-i}(1-\mu_l\omega_m)^{-i}.$$

The sum over i in the above equation can be recognized as a geometric series with n terms. The first term is equal to 1 and the geometric ratio equals $(1-\mu_k\omega_m)^{-1}(1-\mu_l\omega_m)^{-1}$. Hence we have

$$\sum_{i=0}^{n-1}(1-\mu_k\omega_m)^{-i}(1-\mu_l\omega_m)^{-i} \tag{35}$$

$$= \frac{(1-\mu_k\omega_m)(1-\mu_l\omega_m)}{\mu_k\mu_l\omega_m^2 - \mu_k\omega_m - \mu_l\omega_m} - \frac{(1-\mu_k\omega_m)^{-n+1}(1-\mu_l\omega_m)^{-n+1}}{\mu_k\mu_l\omega_m^2 - \mu_k\omega_m - \mu_l\omega_m}.$$

After substitution of the above into (34) and simplification we are left with

$$Y_m = E[v_{k,m}(n)v_{l,m}^*(n)] \tag{36}$$

$$= (1-\mu_k\omega_m)^n(1-\mu_l\omega_m)^n\left[|v_m(0)|^2 + \frac{J_{min}}{\omega_m^2 - \frac{\omega_m}{\mu_l} - \frac{\omega_m}{\mu_k}}\right]$$

$$-\frac{J_{min}}{\omega_m^2 - \frac{\omega_m}{\mu_l} - \frac{\omega_m}{\mu_k}},$$

which is our result for a single entry to the MSD crossterm vector. It is easy to see that for the terms involving a single filter we get an expressions that coincide with the one available in the literature Haykin (2002).

Let us now focus on the cross term

$$EMSE_{kl} = E\left[\tilde{\mathbf{w}}_k^H(n-1)\mathbf{x}(n)\mathbf{x}^H(n)\tilde{\mathbf{w}}_l(n-1)\right],$$

appearing in the EMSE equation (15). Due to the independence assumption we can rewrite this using the properties of trace operator as

$$
\begin{aligned}
EMSE_{kl} &= E\left[\tilde{\mathbf{w}}_k^H(n-1)\mathbf{R_x}\tilde{\mathbf{w}}_l(n-1)\right] \\
&= tr\left\{E\left[\mathbf{R_x}\tilde{\mathbf{w}}_l(n-1)\tilde{\mathbf{w}}_k^H(n-1)\right]\right\} \\
&= tr\left\{\mathbf{R_x}E\left[\tilde{\mathbf{w}}_l(n-1)\tilde{\mathbf{w}}_k^H(n-1)\right]\right\}.
\end{aligned}
\tag{37}
$$

Let us now recall that for any of the filters $\tilde{\mathbf{w}}_i(n) = \boldsymbol{\Omega}\mathbf{v}_i(n)$ to write

$$
\begin{aligned}
EMSE_{kl} &= tr\left\{\mathbf{R_x}E\left[\mathbf{Q}\mathbf{v}_l(n-1)\mathbf{v}_k^H(n-1)\mathbf{Q}^H\right]\right\} \\
&= tr\left\{E\left[\mathbf{v}_k^H(n-1)\mathbf{Q}^H\mathbf{R_x}\mathbf{Q}\mathbf{v}_l(n-1)\right]\right\} \\
&= tr\left\{E\left[\mathbf{v}_k^H(n-1)\boldsymbol{\Omega}\mathbf{v}_l(n-1)\right]\right\} \\
&= \sum_{i=0}^{N-1}\omega_i E\left[v_{k,i}^*(n-1)v_{l,i}(n-1)\right].
\end{aligned}
\tag{38}
$$

The EMSE of the combined filter can now be computed as

$$EMSE = \sum_{i=0}^{N-1}\omega_i E\left[|\lambda(n)v_{k,i}(n-1) + (1-\lambda(n)v_{l,i}(n-1)|^2\right],\tag{39}$$

where the components of type $E[v_{k,i}(n-1)v_{l,i}(n-1)]$ are given by (36). To compute $\lambda(n)$ we use (16) substituting (38) for its individual components.

4. Simulation results

A simulation study was carried out with the aim of verifying the approximations made in the previous Section. In particular we are interested in how well the small step-size theory applies to our combination scheme of two adaptive filters.

We have selected the sample echo path model number one shown in Figure 3 from *ITU-T Recommendation G.168 Digital Network Echo Cancellers* (2009), to be the unknown system to identify.

We have combined two 64 tap long adaptive filters. In order to obtain a practical algorithm, the expectation operators in both numerator and denominator of (8) have been replaced by exponential averaging of the type

$$P_u(n) = (1-\gamma)P_u(n-1) + \gamma u^2(n),\tag{40}$$

where $u(n)$ is the signal to be averaged, $P_u(n)$ is the averaged quantity and $\gamma = 0.01$. The averaged quantities were then used in (8) to obtain λ. With this design the numerator and
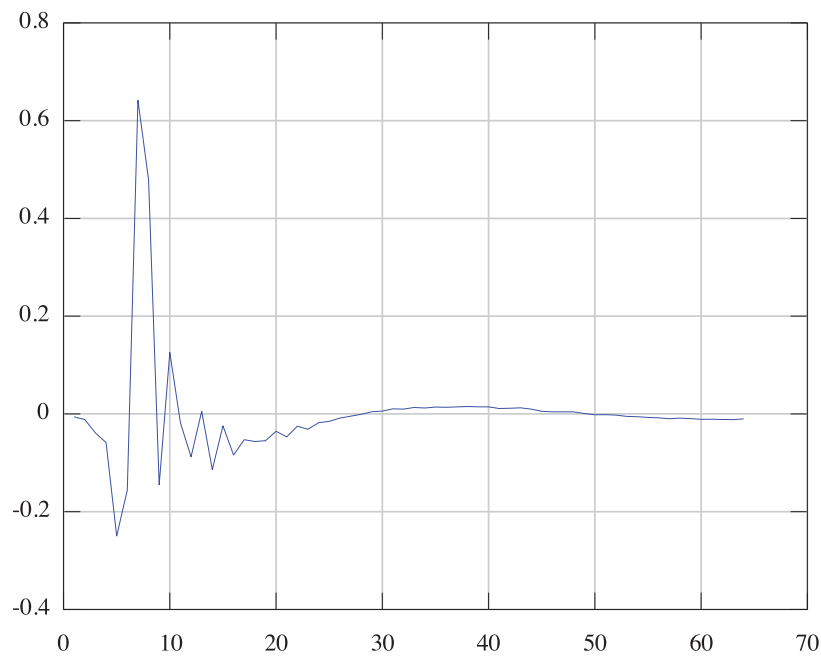
Fig. 3. The true impulse response.

denominator of in the λ expression (8) are relatively small random variables at the beginning of the test cases. For practical purposes we have therefore restricted λ to be less than unity and added a small constant to the denominator to avoid division by zero.

In the Figures below the noisy blue line represents the simulation result and the smooth red line is the theoretical result. The curves are averaged over 100 independent trials.

In our first simulation example we use Gaussian white noise with unity variance as the input signal. The measurement noise is another white Gaussian noise with variance $\sigma_v^2 = 10^{-3}$. The step sizes are $\mu_1 = 0.005$ for the fast adapting filter and $\mu_2 = 0.0005$ for the slowly adapting filter. Figure 4 depicts the evolution of EMSE in time. One can see that the system converges fast in the beginning. The fast convergence is followed by a stabilization period between sample times 1000 – 7000 followed by another convergence to a lower EMSE level between the sample times 8000 – 12000. The second convergence occurs when the mean squared error of the filter with small step size surpasses the performance of the filter with large step size. One can observe that the there is a good accordance between the theoretical and the simulated curves.

In the Figure 5 we show the time evolution of mean square deviation of the combination in the same test case. Again one can see that the theoretical and simulation curves fit well.

We continue with some examples with coloured input signal. In those examples the input signal x is formed from the Gaussian white noise with unity variance by passing it through the filter with transfer function

$$H(z) = \frac{1}{1 - 0.5z^{-1} - 0.1z^{-2}}$$

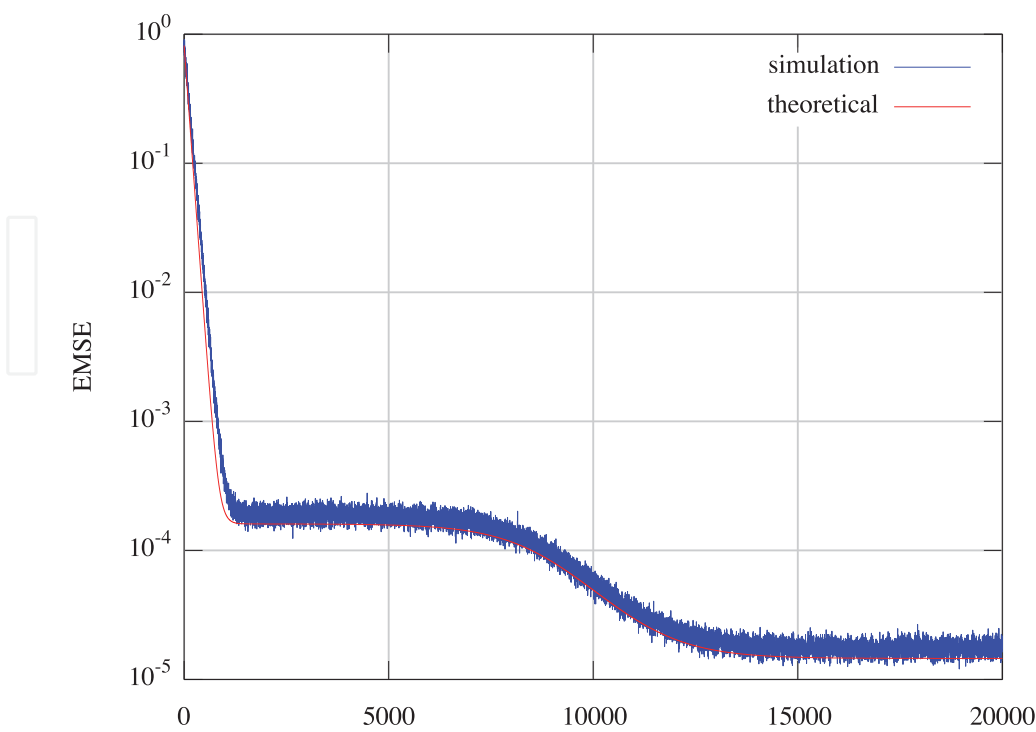to get a coloured input signal. The measurement noise is Gaussian white noise, statistically independent of x.

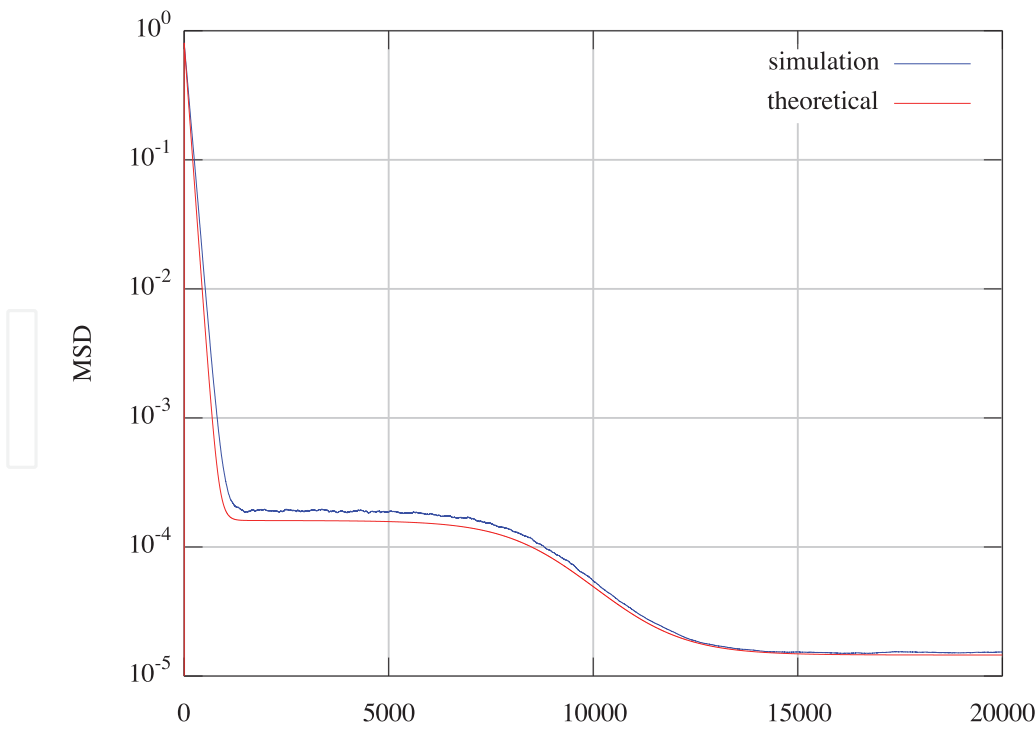Fig. 4. Time–evolutions of EMSE with $\mu_1 = 0.005$ and $\mu_2 = 0.0005$ and $\sigma_v^2 = 10^{-3}$.

Fig. 5. Time–evolutions of EMSE with $\mu_1 = 0.005$ and $\mu_2 = 0.0005$ and $\sigma_v^2 = 10^{-3}$.

In our first simulation example with coloured input we have used observation noise with variance $\sigma_v^2 = 10^{-4}$. The step size of the fast filter is $\mu_1 = 0.005$ and the step size of the slow filter $\mu_2 = 0.001$. As seen from Figure 6 there is a a rapid convergence, determined by the fast converging filter in the beginning followed by a stabilization period. When the EMSE of the slowly adapting filter becomes smaller than that of the fast one, between sample times 10000 and 15000, a second convergence occurs. One can observe a good resemblance between simulation and theoretical curves.
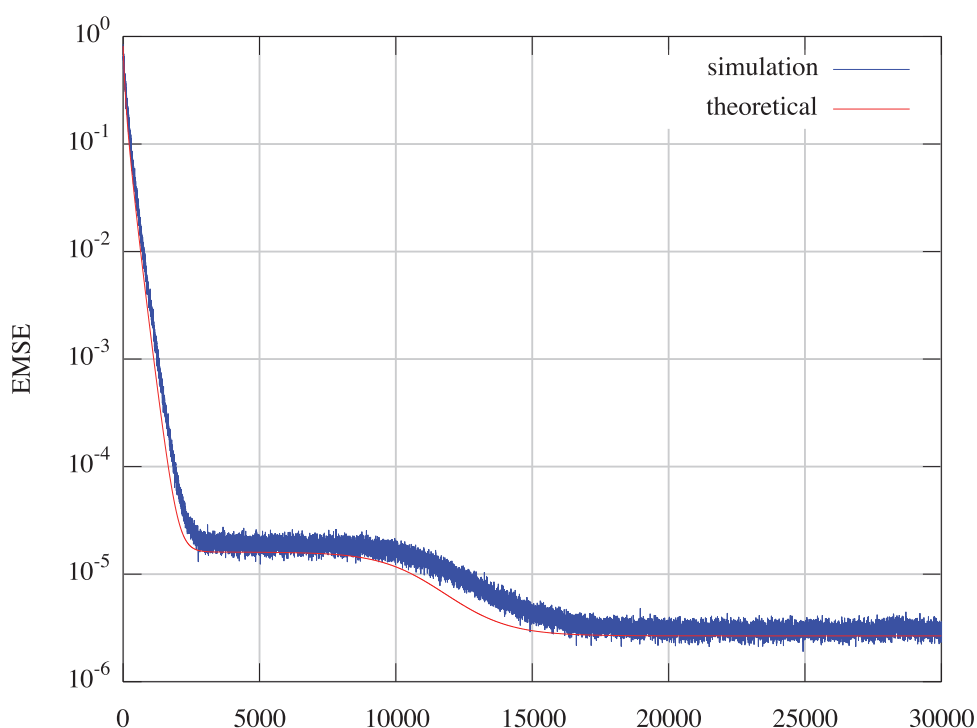
Fig. 6. Time–evolutions of EMSE with $\mu_1 = 0.005$ and $\mu_2 = 0.001$ and $\sigma_v^2 = 10^{-4}$.

In Figure 7 we have made the difference between the step sizes small. The step size of the fast adapting filter is now $\mu_1 = 0.003$ and the step size of the slowly adapting filter is $\mu_2 = 0.002$. One can see that the characteristic horizontal part of the learning curve has almost disappeared. We have also increased the measurement noise level to $\sigma_v^2 = 10^{-2}$. The simulation and theoretical curves show a good match.

In Figure 8 we have increased the measurement noise level even more to $\sigma_v^2 = 10^{-1}$. The step size of the fast adapting filter is $\mu_1 = 0.004$ and the step size of the slowly adapting filter is $\mu_2 = 0.0005$. One can see that the theoretical simulation results agree well.

Figure 9 depicts the time evolution of the combination parameter λ in this simulation. At the beginning of the test case the combination parameter is close to one. Correspondingly the output signal of the fast filter is used as the output of the combination. After a while, when the slow filter catches up the fast one and becomes better, λ changes toward zero and eventually becomes a small negative number. In this state the slow but more accurate filter determines the combined output. Again one can see that there is a clear similarity between the lines.
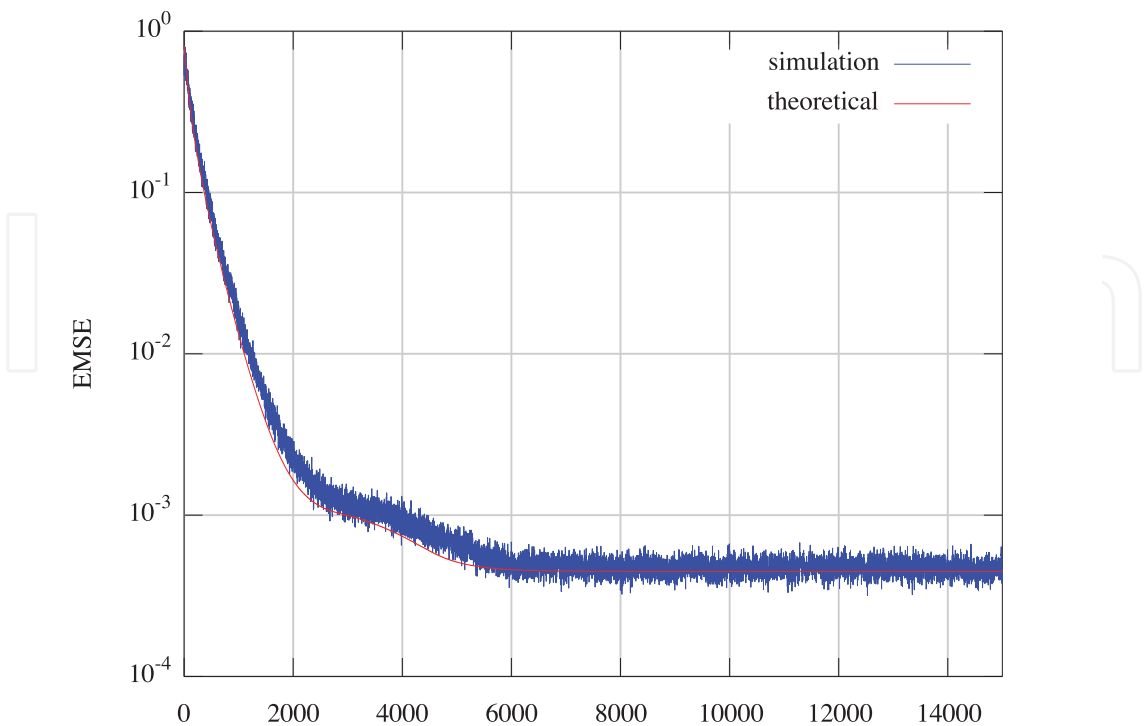
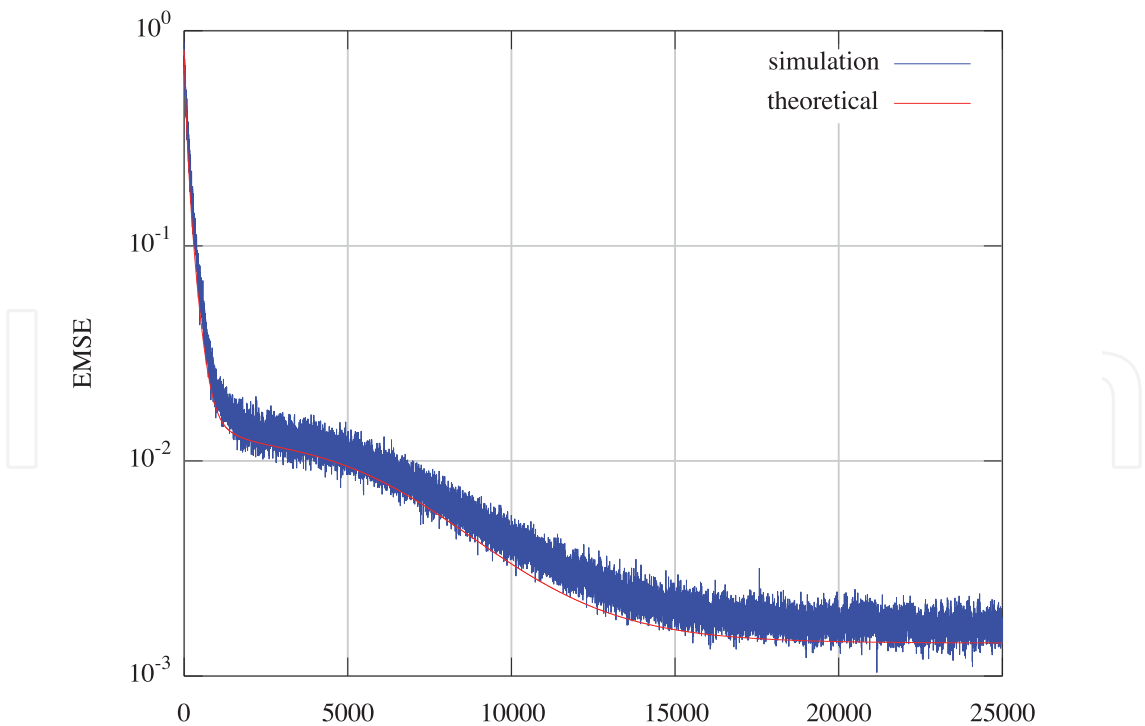Fig. 7. EMSE with $\mu_1 = 0.003$ and $\mu_2 = 0.002$. and $\sigma_v^2 = 10^{-2}$.

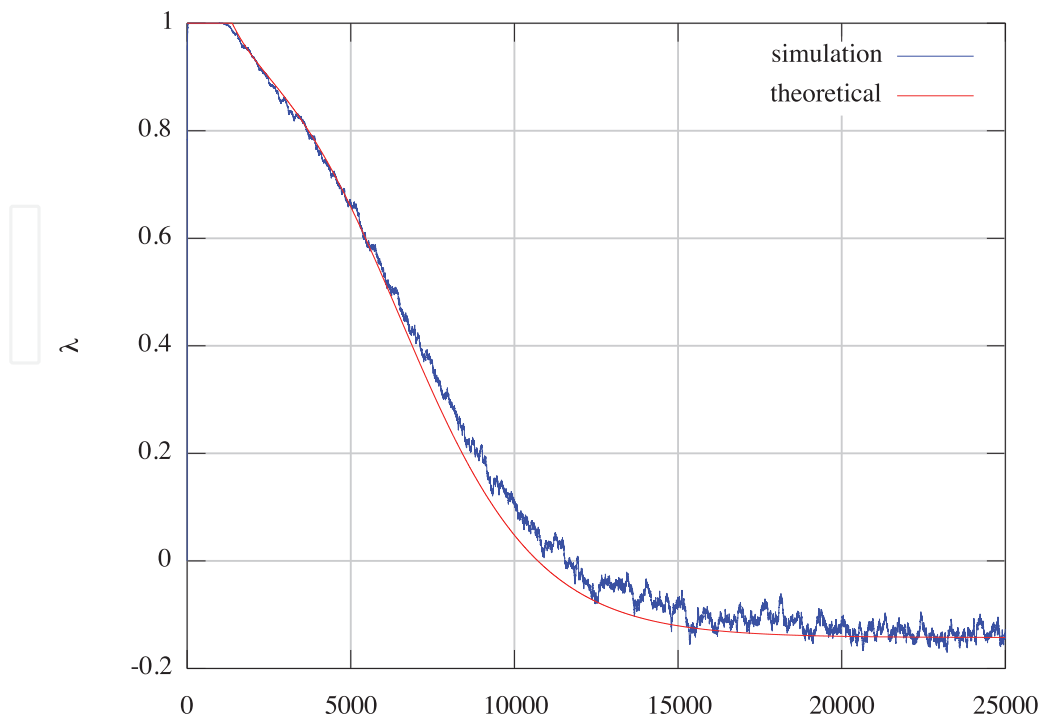Fig. 8. EMSE with $\mu_1 = 0.004$ and $\mu_2 = 0.0005$ and $\sigma_v^2 = 10^{-1}$.

Fig. 9. Time–evolutions of λ with $\mu_1 = 0.004$ and $\mu_2 = 0.0005$. and $\sigma_v^2 = 10^{-1}$.

5. Conclusions

In this chapter we have investigated a combination of two LMS adaptive filters that are simultaneously applied to the same input signals. The output signals of the two adaptive filters are combined together using an adaptive mixing parameter. The mixing parameter λ was computed using the output signals of the individual filters and the desired signal. The transient behaviour of the algorithm was investigated using the assumption of small step size and the expressions for evolution of $EMSE(n)$ and $\lambda(n)$ were derived. Finally it was shown in the simulation study that the derived formulae fit the simulation results well.

6. References

Arenas-Garcia, J., Figueiras-Vidal, A. R. & Sayed, A. H. (1997). A robust variable step–size lms–type algorithm: Analysis and simulations, *IEEE Transactions on Signal Processing* 45: 631–639.

Arenas-Garcia, J., Figueiras-Vidal, A. R. & Sayed, A. H. (2006). Mean-square performance of convex combination of two adaptive filters, *IEEE Transactions on Signal Processing* 54: 1078–1090.

Armbruster, W. (1992). *Wideband Acoustic Echo Canceller with Two Filter Structure, Signal Processing VI, Theories and Applications, J Vanderwalle, R. Boite, M. Moonen and A. Oosterlinck ed.*, Elsevier Science Publishers B.V.

Azpicueta-Ruiz, L. A., Figueiras-Vidal, A. R. & Arenas-Garcia, J. (2008a). A new least squares adaptation scheme for the affine combination of two adaptive filters, *Proc. IEEE*

International Workshop on Machine Learning for Signal Processing, Cancun, Mexico, pp. 327–332.

Azpicueta-Ruiz, L. A., Figueiras-Vidal, A. R. & Arenas-Garcia, J. (2008b). A normalized adaptation scheme for the convex combination of two adaptive filters, *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Las Vegas, Nevada, pp. 3301–3304.

Bershad, N. J., Bermudez, J. C. & Tourneret, J. H. (2008). An affine combination of two lms adaptive filters – transient mean–square analysis, *IEEE Transactions on Signal Processing* 56: 1853–1864.

Candido, R., Silva, M. T. M. & Nascimento, V. H. (2010). Transient and steady-state analysis of the affine combination of two adaptive filters, *IEEE Transactions on Signal Processing* 58: 4064–4078.

Fathiyan, A. & Eshghi, M. (2009). Combining several pbs-lms filters as a general form of convex combination of two filters, *Journal of Applied Sciences* 9: 759–764.

Harris, R. W., Chabries, D. M. & Bishop, F. A. (1986). Variable step (vs) adaptive filter algorithm, *IEEE Transactions on Acoustics, Speech and Signal Processing* 34: 309–316.

Haykin, S. (2002). *Adaptive Filter Theory, Fourth Edition,*, Prentice Hall.

ITU-T Recommendation G.168 Digital Network Echo Cancellers (2009). ITU-T.

Kim, K., Choi, Y., Kim, S. & Song, W. (2008). Convex combination of affine projection filters with individual regularization, *Proc. 23rd International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, Shimonoseki, Japan, pp. 901–904.

Kwong, R. H. & Johnston, E. W. (1992). A variable step size lms algorithm, *IEEE Transactions on Signal Processing* 40: 1633–1642.

Mandic, D., Vayanos, P., Boukis, C., Jelfs, B., Goh, S. I., Gautama, T. & Rutkowski, T. (2007). Collaborative adaptive learning using hybrid filters, *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Honolulu, Hawaii, pp. 901 – 924.

Martinez-Ramon, M., Arenas-Garcia, J., Navia-Vazquez, A. & Figueiras-Vidal, A. R. (2002). An adaptive combination of adaptive filters for plant identification, *Proc. 14th International Conference on Digital Signal Processing*, Santorini, Greece, pp. 1195–1198.

Matthews, V. J. & Xie, Z. (1993). A stochastic gradient adaptive filter with gradient adaptive step size, *IEEE Transactions on Signal Processing* 41: 2075–2087.

Ochiai, K. (1977). Echo canceller with two echo path models, *IEEE Transactions on Communications* 25: 589–594.

Sayed, A. H. (2008). *Adaptive Filters*, John Wiley and sons.

Shin, H. C. & Sayed, A. H. (2004). Variable step–size nlms and affine projection algorithms, *IEEE Signal Processing Letters* 11: 132–135.

Silva, M. T. M., Nascimento, V. H. & Arenas-Garcia, J. (2010). A transient analysis for the convex combination of two adaptive filters with transfer of coefficients, *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Dallas, TX, USA, pp. 3842–3845.

Trump, T. (2009a). An output signal based combination of two nlms adaptive algorithms, *Proc. 16th International Conference on Digital Signal Processing*, Santorini, Greece.

Trump, T. (2009b). Steady state analysis of an output signal based combination of two nlms adaptive filters, *Proc. 17th European Signal Processing Conference*, Glasgow, Scotland.

Trump, T. (2009c). Tracking performance of a combination of two nlms adaptive filters, *Proc. IEEE Workshop on Statistical Signal Processing*, Cardiff, UK.

Widrow, B. & Hoff, M. E. J. (1960). Adaptive switching circuits, *IRE WESCON Conv. Rec.*, pp. 96–104.

Zhang, Y. & Chambers, J. A. (2006). Convex combination of adaptive filters for a variable tap–length lms algorithm, *IEEE Signal Processing Letters* 13: 628–631.

Adaptive Filtering

Edited by Dr Lino Garcia

Adaptive filtering is useful in any application where the signals or the modeled system vary over time. The configuration of the system and, in particular, the position where the adaptive processor is placed generate different areas or application fields such as prediction, system identification and modeling, equalization, cancellation of interference, etc., which are very important in many disciplines such as control systems, communications, signal processing, acoustics, voice, sound and image, etc. The book consists of noise and echo cancellation, medical applications, communications systems and others hardly joined by their heterogeneity. Each application is a case study with rigor that shows weakness/strength of the method used, assesses its suitability and suggests new forms and areas of use. The problems are becoming increasingly complex and applications must be adapted to solve them. The adaptive filters have proven to be useful in these environments of multiple input/output, variant-time behaviors, and long and complex transfer functions effectively, but fundamentally they still have to evolve. This book is a demonstration of this and a small illustration of everything that is to come.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Tõnu Trump (2011). Transient Analysis of a Combination of Two Adaptive Filters, Adaptive Filtering, Dr Lino Garcia (Ed.), ISBN: 978-953-307-158-9, InTech, Available from: http://www.intechopen.com/books/adaptive-filtering/transient-analysis-of-a-combination-of-two-adaptive-filters

INTECH
open science | open minds