EUSIPCO 2013 1569741805

ON A COMBINATION OF M ADAPTIVE FILTERS

Tõnu Trump

Department of Radio and Telecommunication Engineering
Tallinn University of Technology, Tallinn, Estonia
tonu.trump@lr.ttu.ee

ABSTRACT

In this paper we consider a combination of arbitrary number of LMS adaptive filters. The filters are connected in parallel and use the same input and desired signals. They differ by the step sizes, which gives the structure ability to achieve fast initial convergence together with gradually diminishing steady state error level. In the paper we show that the straightforward problem statement leads to a necessity to solve a singular linear system of equations. We therefore propose a regularization approach to deal with the problem. Results of the transient analysis of the resulting algorithm are presented together with some simulation results.

Index Terms— Adaptive filtering, regularization, analysis

1. INTRODUCTION

Most popular of the adaptive filtering schemes, proposed over the years, is probably the least mean square (LMS) family [1]. The LMS filters are controlled by a step size parameter, that determines the convergence speed and the steady state error level achieved by the algorithm. Large step size leads to a fast initial convergence of the algorithm but the mean square error in steady state is large in this case. In contrary a small step size results in a small steady state error but the convergence of the algorithm is slow [2, 3].

A recently proposed solution to this problem is combining the outputs of a number of adaptive filtering branches that use different step sizes [4, 5, 6, 7]. The solution has two stages. At the first stage there are several adaptive filters running in parallel on the same task. At the second stage the outputs of the adaptive filters are linearly combined to form the output of the system. This solution, involving two filters and a convex combination of those was first proposed in [8]. Later on there has been significant interest in the scheme and both convex combinations of two filters, where the mixture weights are constrained to be positive and affine combinations where this constraint has been removed have been investigated. Most of the papers use some nonlinear schemes to compute the mixing parameter. More lately an output signal based scheme that simplifies computation of the branch weights was proposed in [9] and [10].

Recently the two filter scheme was generalized to a mixture of M filters in [6] and different mixing strategies in which the final outputs are formed as the weighted linear combination of the outputs of several constituent adaptive algorithms was studied. The steady state performance in terms of the final mean square error (MSE) of the adaptive mixtures was investigated. The scheme can again be seen as consisting of two adaptive stages. At the first stage there are a number of adaptive filters running the same task but having different step sizes. These parallel units can be considered as diversity branches that can be used to improve the overall performance. At the second stage the output signals of the adaptive branches are mixed together. It was demonstrated in [6] that the mixture approaches can greatly improve the performance of the constituent filters.

In this paper we examine the task of computing the weights required to combine the output signals from the parallel filters. We conclude that straightforward statement of the mixing weight computation problem is not well posed as we would need to invert a singular matrix. We therefore propose a regularized solution to the problem at hand and provide the results of its analysis. We will assume throughout the paper that the signals are complex–valued and that the combination scheme uses LMS adaptive filters. The length of the filters is assumed to be equal in all the branches.

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 symbol $\mathbf{1}$ is used to denote the column vector of all ones. The operator $E[\cdot]$ denotes mathematical expectation, $tr[\cdot]$ stands for trace of a matrix.

2. ALGORITHM

Let us consider M adaptive filters running in parallel on the same input signals, as shown in Figure 1. Each of the filters is updated using the LMS adaptation rule

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

$$e_i(n) = d(n) - \mathbf{w}_i^H(n-1)\mathbf{x}(n), \quad (2)$$

$$d(n) = \mathbf{w}_o^H\mathbf{x}(n) + e_o(n). \quad (3)$$

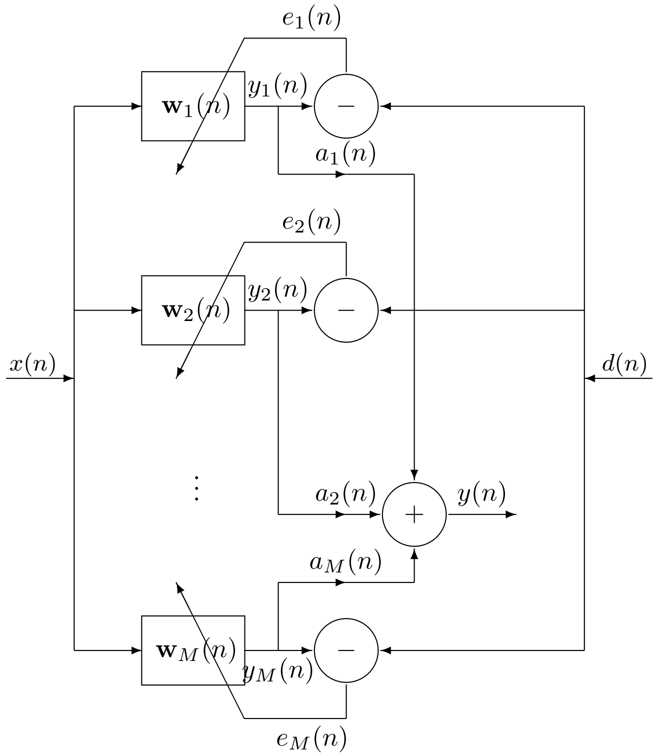

Fig. 1. The combined adaptive filter.

In the above equations the vector $\mathbf{w}_i(n)$ is the length N vector of coefficients of the i-th adaptive filter, with $i = 1, 2, \ldots, M$. 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 to all adaptive filters. 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 $e_o(n)$. 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.

The outputs of the adaptive filters of the first stage are combined according to

$$y(n) = \mathbf{a}^H(n)\mathbf{y}(n), \tag{4}$$

where $\mathbf{y}(n)$ is the M the vector of output signals of the individual filters with elements $y_i(n) = \mathbf{w}_i^H(n-1)\mathbf{x}(n)$ and the vector $\mathbf{a}(n)$ contains the mixing parameters $a_i(n)$. We require the sum of the mixing parameters to equal unity at each time instant n

$$\mathbf{a}^H(n)\mathbf{1} = 1. \tag{5}$$

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

$$e_a(n) = y_o(n) - \mathbf{a}^H(n)\mathbf{y}(n). \tag{6}$$

The mixing parameters $a_i(n)$ can be found by solving the following constrained optimization problem

$$\min_{\mathbf{a}} E\left[|e_a(n)|^2\right] \text{ subject to } \mathbf{a}^H\mathbf{1} = 1. \tag{7}$$

Alternatively we may rewrite the above as unconstrained minimization problem replacing $a_M^*(n)$ by $1 - a_1^*(n) - a_2^*(n) - \ldots - a_{M-1}^*(n)$.

Let us define a vector consisting of the first $M-1$ combination parameters $a_i(n)$ and denote the vector by $\bar{\mathbf{a}}(n)$ so that

$$\mathbf{a}(n) = \begin{bmatrix} \bar{\mathbf{a}}(n) \\ a_M(n) \end{bmatrix}. \tag{8}$$

Likewise we define the vector of $M-1$ first filter output signals $y_i(n)$-s as $\bar{\mathbf{y}}(n)$. Taking the conjugate derivative of the criterion we obtain

$$\frac{\partial E\left[|e_a(n)|^2\right]}{\partial \mathbf{a}^*(n)} = \mathbf{y}(n)\mathbf{y}^H(n)\mathbf{a}(n) - \mathbf{y}(n)y_o(n). \tag{9}$$

After setting the derivative to zero, substituting $a_M = 1 - \mathbf{1}^H\bar{\mathbf{a}}$ and simplifying we find that the cost function we wish to minimize, $E[|e_a(n)|^2]$, has in this problem no unique minimum point. Instead the minimum is reached on the entire hyperplane

$$E[y_M(n) - y_0(n)] = E[\bar{\mathbf{a}}^H(n)(y_M(n)\mathbf{1} - \bar{\mathbf{y}}(n))], \tag{10}$$

where $E[|e_a(n)|^2 = 0$. In an adaptive solution this would mean that the weights $a_i(n)$ can freely move around on the solution hyperplane, which would be inconvenient. In order to fix the minimum of the criterion to a certain point we need to add another condition. A natural additional condition would be a requirement that the Euclidean norm of the solution vector were minimal. This can be accomplished by adding the term $\beta\bar{\mathbf{a}}^H(n)\bar{\mathbf{a}}(n)$ to the mean squared error in our criterion function. The constant β shows the relative importance of the two requirements and is treated as a design parameter. We thus need to solve the following optimization problem

$$\mathbf{a} = \arg\min_{\mathbf{a}} E\left[|e_a(n)|^2 + \beta\bar{\mathbf{a}}^H(n)\bar{\mathbf{a}}(n)\right]. \tag{11}$$

The conjugate derivative of this criterion with respect to the vector of mixing weights $\bar{\mathbf{a}}(n)$ reads

$$\begin{aligned} &\frac{\partial E\left[|e_a(n)|^2 + \beta\bar{\mathbf{a}}^H(n)\bar{\mathbf{a}}(n)\right]}{\partial \bar{\mathbf{a}}^*(n)} \\ = \;& E\left[(-\bar{\mathbf{y}}(n) + \mathbf{1}y_M(n))(y_o(n) - y_M(n))^*\right] \\ &+E\left[(-\bar{\mathbf{y}}(n) + \mathbf{1}y_M(n))(-\bar{\mathbf{y}}(n) + \mathbf{1}y_M(n))^H\bar{\mathbf{a}}(n)\right. \\ &\left.+2\beta\bar{\mathbf{a}}(n)\right]. \end{aligned}$$

Setting the derivative to zero and solving the resulting system of equations with respect to $\bar{\mathbf{a}}(n)$ yields

$$\begin{aligned}\bar{\mathbf{a}}(n) &= E\left[(-\bar{\mathbf{y}}(n)+\mathbf{1}y_M(n))(-\bar{\mathbf{y}}(n)+\mathbf{1}y_M(n))^H\right.\\ &\quad \left.+2\beta\mathbf{I}\right]^{-1} \qquad (12)\\ &\quad E\left[(\bar{\mathbf{y}}(n)-\mathbf{1}y_M(n))(d(n)-y_M(n))^*)\right].\end{aligned}$$

Let us note that if $\beta = 0$ the determinant of the matrix in the above equation equals zero, which explains the need for regularization. Here 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 $e_o(n)$ are independent random processes, this can be done without introducing any error into our calculations.

3. ANALYSIS

In this section we provide expressions that characterize transient behaviour of the mixture of M adaptive filters i.e. we are interested in deriving 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 $\tilde{\mathbf{w}}_i(n) = \mathbf{w}_o - \mathbf{w}_i(n)$. The weight error vector of the entire mixture will then be

$$\tilde{\mathbf{w}}(n) = \sum_{i=1}^{M} a_i(n)\tilde{\mathbf{w}}_i(n). \qquad (13)$$

The mean square deviation of the mixture is given by

$$\begin{aligned}MSD &= E\left[\tilde{\mathbf{w}}^H(n)\tilde{\mathbf{w}}(n)\right] \qquad (14)\\ &= \sum_{k=1}^{M}\sum_{l=1}^{M} E\left[a_k^*(n)\tilde{\mathbf{w}}_k^H(n)\tilde{\mathbf{w}}_l(n)a_l(n)\right].\end{aligned}$$

The *a priori* estimation error of i–th adaptive filter of the first stage is defined as

$$e_{i,a}(n) = \tilde{\mathbf{w}}_i^H(n-1)\mathbf{x}(n). \qquad (15)$$

It follows from (6) that we can express the *a priori* error of the mixture as

$$e_a(n) = \sum_{i=1}^{M} a_i^*(n)e_{i,a}(n) \qquad (16)$$

and because $a_i(n)$ are according to (12) computable through mathematical expectations and are, hence, deterministic, we have the following expressions for the excess mean square error $EMSE$ of the combination

$$\begin{aligned}E[|e_a(n)|^2] &= \sum_{k=1}^{M}\sum_{l=1}^{M} a_k^*(n)a_l(n)\\ &\quad \cdot E\left[\tilde{\mathbf{w}}_k^H(n-1)\mathbf{x}(n)\mathbf{x}^H(n)\tilde{\mathbf{w}}_l(n-1)\right]\\ &= \sum_{k=1}^{M}\sum_{l=1}^{M} a_k^*(n)a_l(n)EMSE_{k,l} \qquad (17)\end{aligned}$$

Hence the $EMSE$ can be expressed as a linear combination of the terms

$$EMSE_{k,l} = E\left[\tilde{\mathbf{w}}_k^H(n-1)\mathbf{x}(n)\mathbf{x}^H(n)\tilde{\mathbf{w}}_l(n-1)\right] \qquad (18)$$

with weights being products of $a_i(n)$.

Next let us consider computation of the vector $\bar{\mathbf{a}}(n)$. Noting that $y_i(n) = \mathbf{w}_i^H(n-1)\mathbf{x}(n)$ we can rewrite the terms involved in (12) as follows. First, the i,j–th element of the matrix

$$(\mathbf{1}y_M(n)-\bar{\mathbf{y}}(n))(\mathbf{1}y_M(n)-\bar{\mathbf{y}}(n))^H \qquad (19)$$

equals

$$\tilde{\mathbf{w}}_i^H\mathbf{x}\mathbf{x}^H\tilde{\mathbf{w}}_j - \tilde{\mathbf{w}}_i^H\mathbf{x}\mathbf{x}^H\tilde{\mathbf{w}}_M - \tilde{\mathbf{w}}_M^H\mathbf{x}\mathbf{x}^H\tilde{\mathbf{w}}_j + \tilde{\mathbf{w}}_M^H\mathbf{x}\mathbf{x}^H\tilde{\mathbf{w}}_M \qquad (20)$$

and the vector

$$\begin{aligned}&(\mathbf{1}y_M(n)-\bar{\mathbf{y}}(n))(d(n)-y_M(n)) \qquad (21)\\ &= \begin{bmatrix}\tilde{\mathbf{w}}_1^H\mathbf{x}\mathbf{x}^H\tilde{\mathbf{w}}_M\\ \tilde{\mathbf{w}}_2^H\mathbf{x}\mathbf{x}^H\tilde{\mathbf{w}}_M\\ \vdots\\ \tilde{\mathbf{w}}_{M-1}^H\mathbf{x}\mathbf{x}^H\tilde{\mathbf{w}}_M\end{bmatrix} - \mathbf{1}\tilde{\mathbf{w}}_M^H\mathbf{x}\mathbf{x}^H\tilde{\mathbf{w}}_M.\end{aligned}$$

It thus turns out that in order to reveal the behaviour of both $EMSE(n)$ and $\mathbf{a}(n)$ we need to investigate the terms $EMSE_{k,l} = E[\tilde{\mathbf{w}}_k^H(n-1)\mathbf{x}(n)\mathbf{x}^H(n)\tilde{\mathbf{w}}_l(n-1)]$. Those terms have, however, been analysed in [11], where it has been shown that

$$\begin{aligned}EMSE_{kl} &\approx E\left[\tilde{\mathbf{w}}_k^H(n-1)\mathbf{R}_\mathbf{x}\tilde{\mathbf{w}}_l(n-1)\right] \qquad (22)\\ &= tr\left\{E\left[\mathbf{v}_k^H(n-1)\mathbf{\Lambda}\mathbf{v}_l(n-1)\right]\right\}\\ &= \sum_{i=0}^{N-1}\lambda_i E\left[v_{k,i}^*(n-1)v_{l,i}(n-1)\right],\end{aligned}$$

where λ_i is the i–th eigenvalue of the input signal covariance matrix $\mathbf{R}_\mathbf{x}$ and $v_{k,i}$ is the i–th component of transformed weight error vector of the k–th filter $\mathbf{v}_k(n) = \mathbf{Q}^H\tilde{\mathbf{w}}_k(n)$. The matrix $\mathbf{Q}$ here is the matrix that has the eigenvectors of the input signal covariance matrix as its columns. The matrix $\mathbf{\Lambda}$ is a diagonal matrix having eigenvalues associated with the corresponding eigenvectors on its main diagonal.

The individual expectation in the above equation equals

$$\begin{aligned}&E[v_{k,i}^*(n)v_{l,i}(n)] = (1-\mu_k\lambda_i)^n(1-\mu_l\lambda_i)^n \qquad (23)\\ &\cdot\left[|v_i(0)|^2 + \frac{J_{min}}{\lambda_i^2 - \frac{\lambda_i}{\mu_l} - \frac{\lambda_i}{\mu_k}}\right]\\ &-\frac{J_{min}}{\lambda_m^2 - \frac{\lambda_i}{\mu_l} - \frac{\lambda_i}{\mu_k}}.\end{aligned}$$

In the above J_{min} is the minimum possible error power produced by the Wiener filter for the problem at hand and we have assumed that the filters are all initialized to the same value $v_{k,i}(0) = v_i(0), \forall k$.

The EMSE of the combined filter can now be computed as

$$EMSE = \sum_{k=1}^{M}\sum_{l=1}^{M} a_k^*(n)a_l(n)EMSE_{k,l} \quad (24)$$
$$= \sum_{k=1}^{M}\sum_{l=1}^{M} a_k^*(n)a_l(n) \cdot \sum_{i=0}^{N-1} \lambda_i E\left[v_{k,i}^*(n-1)v_{l,i}(n-1)\right],$$

where the components of type $E\left[v_{k,i}^*(n-1)v_{l,i}(n-1)\right]$ are given by (23). To compute $\mathbf{a}(n)$ we use (12) with (19) and (21) substituting (23) for its individual components.

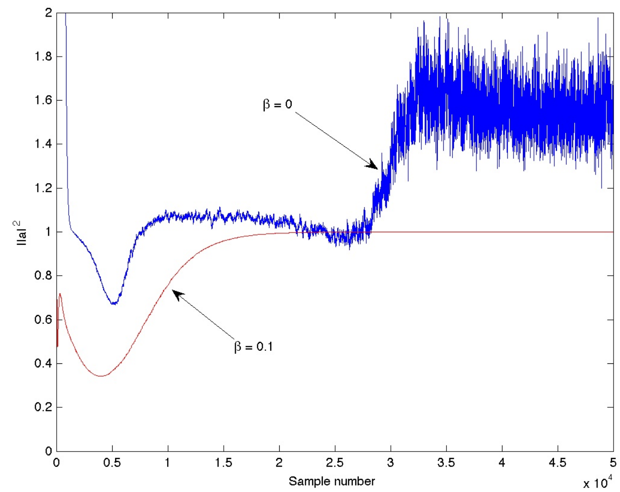

Fig. 2. Time–evolutions of $|\mathbf{a}|^2$ with $\beta = 0$ and $\beta = 0.1$.

4. SIMULATION RESULTS

In the simulation study we have combined three 64 tap long adaptive filters and selected the sample echo path model number one from [12] to be the unknown system to identify. The mathematical expectations have been in simulations replaced by exponential averaging

$$P_u(n) = (1-\gamma)P_u(n-1) + \gamma p(n), \quad (25)$$

where P is the averaged value, p is the value to be averaged and γ is a constant.

Let us first illustrate the problem we deal with in this paper. In Figure 2 we show the norm of weight vector $\mathbf{a}(n)$ with and without the term $2\beta\mathbf{I}$ in (12). If the regularization term is missing, the solution wanders around and never converges to a single value. This is seen observing that the Euclidean norm of the solution behaves randomly until the end of the simulation. The Euclidean norm of the solution with $2\beta\mathbf{I}$ added converges on the other hand to a determined value. The step sizes were selected $\mu_1 = 0.05$, $\mu_2 = 0.0005$ and $\mu_3 = 0.0002$ in this simulation and the input was white Gaussian noise.

In the second simulation example we use a coloured input signal created by passing a unity variance white Gaussian noise through a filter with transfer function

$$H_2(z) = \frac{1}{1 + 0.5z^{-1} + 0.2z^{-2} - 0.2z^{-3} - 0.1z^{-4}}. \quad (26)$$

The exponential averaging parameter $\gamma = 0.01$.

The resulting excess mean square error of the combination of three filters is shown in Figure 3. The simulation curve is averaged over 100 trials.The regularization parameter is $\beta = 5 \cdot 10^{-4}$ in this example. We see a period of rapid initial convergence followed by a short stabilization period. Starting from sample time 5000 another convergence of the system occurs ending at a lower error level around sample time 15000. Then there is another stabilization period and the third convergence happens between sample times 25000 and 40000. There is a good match between the simulation and theoretical curves.

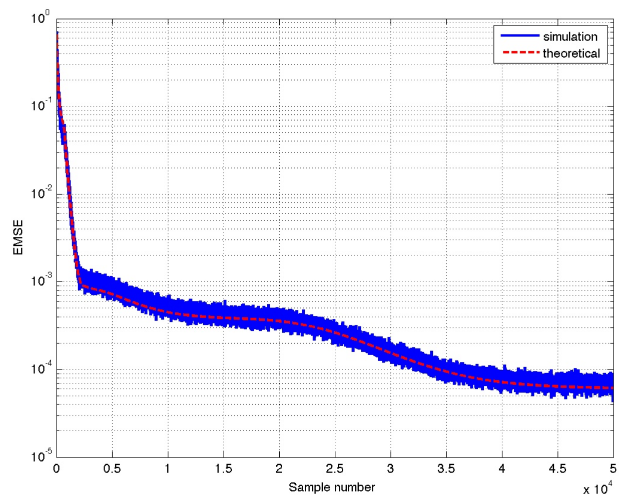

Fig. 3. Time–evolution of EMSE with $\mu_1 = 0.005$, $\mu_2 = 0.0008$, $\mu_3 = 0.0004$ and $\sigma_v^2 = 5 \cdot 10^{-3}$.

The time evolution of the weights a_i is shown in Figure 4. In the beginning of the simulation example the weight factor corresponding to the fastest adapting filter, a_1, is close to one and the other two are close to zero, meaning that the initial convergence of the combination is determined by the fastest adapting filter. After a while, when the fastest filter approaches its steady state, its weight starts to diminish and the second filter takes over. This happens around the sample

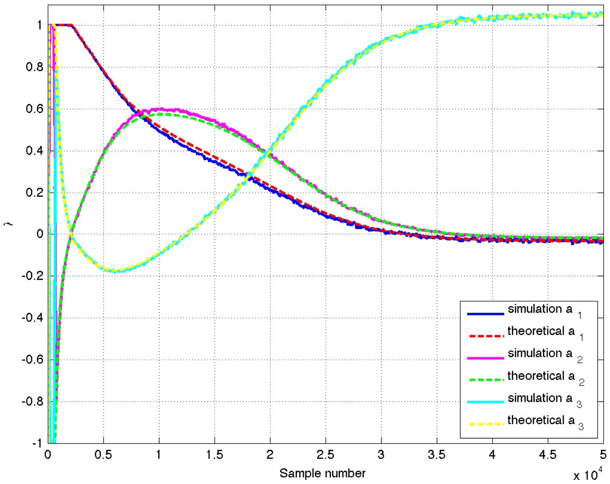

Fig. 4. Time–evolution of a_i with $\mu_1 = 0.005$, $\mu_2 = 0.0008$, $\mu_3 = 0.0004$ and $\sigma_v^2 = 5 \cdot 10^{-3}$.

time 10000. At the end of the simulation example the weight of the slowest adapting filter, a_3, approaches to unity while the other two approach to zero meaning that the steady state performance of the combination is determined by the most exact filter. The theoretical curves match the simulation results well.

Finally in Figure 5 we show the learning curves of the individual filters together with the learning curve of the combination. We see that the convergence speeds of the individual filters are all different and that the learning curve of the combination generally follows the individual filter with the smallest EMSE.

5. CONCLUSIONS

In this paper we investigated the combination of M adaptive filters. It has been shown that in order to solve for the optimal weights of the filter output signals we need to solve a singular linear system of equations. In order to cope with the problem a regularization approach was proposed. Analyses results of the resulting algorithm were provided.

6. REFERENCES

[1] B. Widrow and S.D. Stearns, *Adaptive Signal Processing*, Prentice Hall, 1985.

[2] S. Haykin, *Adaptive Filter Theory, Fourth Edition*, Prentice Hall, 2002.

[3] A. H. Sayed, *Adaptive Filters*, John Wiley and sons, 2008.

[4] R. Candido, M. T. M. Silva, and V. H. Nascimento, "Transient and steady-state analysis of the affine combination of two adaptive filters," *IEEE Transactions on Signal Processing*, vol. 58, pp. 4064–4078, August 2010.

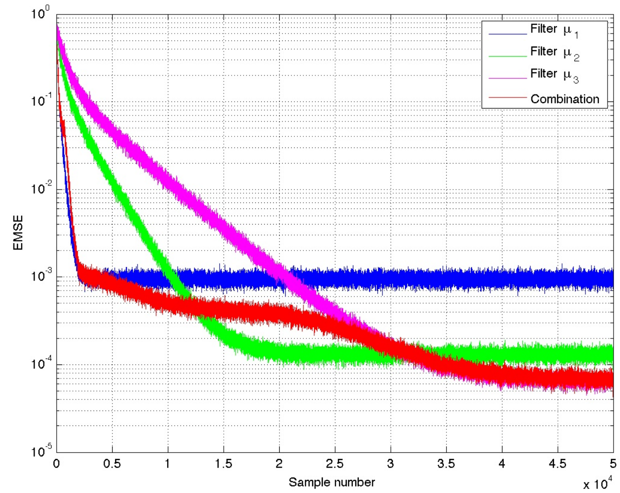

Fig. 5. Time–evolution of individual filter EMSEs with $\mu_1 = 0.005$, $\mu_2 = 0.0008$, $\mu_3 = 0.0004$ and $\sigma_v^2 = 5 \cdot 10^{-3}$.

[5] N. J. Bershad, J. C. Bermudez, and J. H. Tourneret, "An affine combination of two LMS adaptive filters – transient mean–square analysis," *IEEE Transactions on Signal Processing*, vol. 56, pp. 1853–1864, May 2008.

[6] S.S. Kozat, A.T Erdogan, A.C Singer, and A.H Sayed, "Steady state MSE performance analysis of mixture approaches to adaptive filtering," *IEEE Transactions on Signal Processing*, vol. 58, pp. 4050–4063, August 2010.

[7] T. Trump, "Steady state analysis of an output signal based combination of two NLMS adaptive filters," in *Proc. 17th European Signal Processing Conference*, Glasgow, Scotland, 2009, pp. 1720–1724.

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

[9] T. Trump, "An output signal based combination of two NLMS adaptive algorithms," in *Proc. 16th International Conference on Digital Signal Processing*, Santorini, Greece, 2009.

[10] L. A. Azpicueta-Ruiz, A. R. Figueiras-Vidal, and J. Arenas-Garcia, "A new least squares adaptation scheme for the affine combination of two adaptive filters," in *Proc. IEEE International Workshop on Machine Learning for Signal Processing*, Cancun, Mexico, 2008, pp. 327–332.

[11] T. Trump, "Output signal based combination of two NLMS adaptive filters - transient analysis," *Proceedings of the Estonian Academy of Sciences*, vol. 60, no. 4, pp. 258–268, 2011.

[12] *ITU-T Recommendation G.168 Digital Network Echo Cancellers*, ITU-T, 2009.