TRACKING PERFORMANCE OF ADAPTIVELY BIASED ADAPTIVE FILTERS

Jerónimo Arenas-García[†] *and Miguel Lázaro-Gredilla*[‡]

[†] Dept. Signal Theory and Commun.
Univ. Carlos III de Madrid, Spain
jarenas@tsc.uc3m.es

[‡]Dept. of Communications Engineering
Univ. de Cantabria, Spain
miguellg@gtas.dicom.unican.es

ABSTRACT

Adaptive filters can improve their performance by exploiting the well-known tradeoff between bias and variance of the estimated solution. In a previous work, a scheme for adaptively biasing the filter weights was introduced, multiplying the output of a filter of any kind by a shrinking factor $\alpha \in [0, 1]$. With an appropriate value α, such a scheme can reduce the steady-state error, especially for low signal-to-noise ratio (SNR). Here, we extend such analysis for a tracking scenario in which the optimal solution follows a random walk-model. We briefly review a realizable scheme for learning α, based on recently proposed algorithms for adaptive filter combination. Our experiments validate the accurateness of the analysis, and illustrate the performance gains that can be expected from these biased configurations in stationary and tracking scenarios.

***Index Terms*—** Adaptive filters, biased estimation, tracking performance, bias-variance tradeoff, combination filters.

1. INTRODUCTION

Adaptive filters are nowadays crucial components in many signal processing applications, such as system identification, echo cancellation, or channel equalization. These tools are generally used in situations where the statistics of the involved stochastic processes are not (completely) known, so that an optimal filter design is not possible, and become especially useful when such statistics, and therefore the optimal solution, change over time [1].

No matter how sophisticated adaptive filters become, their adaptation speed is typically controlled by some parameter (e.g., a step size in stochastic gradient schemes, or a forgetting factor in recursive-least-squares (RLS) algorithms) whose selection imposes a compromise regarding the speed of convergence and the steady-state misadjustment in stationary and tracking conditions. Thus, increasing the adaptation speed of the filter results in faster adaptation, but also in larger gradient noise after convergence, increasing the estimation variance and the residual error of the filter.

In a recent paper [2], we studied how the modified filter configuration depicted in Fig. 1 can exploit a bias versus variance tradeoff, as is customary in the estimation theory literature [3], to reduce the mean-square error (MSE) of adaptive filters in stationary situations. The basic idea consists in multiplying the filter output, $y(n)$, by a shrinkage factor $\alpha \in [0, 1]$ to produce a modified estimator of the reference signal: $y_\alpha(n) = \alpha y(n)$. The resulting scheme can be considered as a new filter with weights $\mathbf{w}_\alpha(n) = \alpha\mathbf{w}(n)$ and output error $e_\alpha(n) = d(n) - y_\alpha(n)$, $d(n)$ being the reference signal to be estimated by the filter.

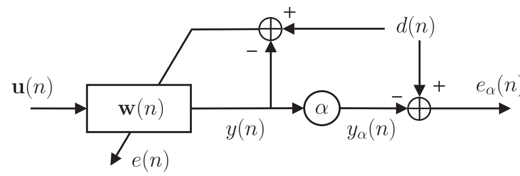


Fig. 1. Adaptive filter with biased weights estimation via output multiplication.

As was studied in [2], the introduction of the shrinkage factor biases the filter weights towards zero, but also reduces the variance of the error. As a results, if α is appropriately selected, the tradeoff between bias and variance can be exploited to reduce the MSE of the biased filter configuration. It should be clear that these advantages apply to any kind of adaptive filtering algorithm, and that more sophisticated schemes for biasing the filter could be considered, e.g., to benefit from structured inputs.

The steady-state excess MSE (EMSE) of the configuration in Fig. 1 was studied both analytically and empirically in [2] for stationary situations. In this paper, we extend such analysis to a more challenging tracking scenario where the optimal solution changes in each iteration according to a random-walk model. The analysis suggests that the filter can also reduce the steady-state tracking EMSE with respect to that of the unbiased filter, and those advantages are more significant for small signal-to-noise ratio (SNR). Since in practice the SNR and the degree of non-stationarity of the solution are not known, or can even be time varying, our scheme offers improved robustness to this lack of knowledge.

Since the optimal value of α can change with time, some practical schemes for adjusting the shrinkage factor were presented in [2], based on algorithms recently proposed in the literature for filter combination [4]–[8]. Among those schemes, the one using a sigmoid function (adapted from [7]) was shown to be very effective, providing both fast adaptation of α and reducing the additional gradient noise in some important situations (notably, for $\alpha = 1$, when the biased configuration should perform exactly like the original unbiased filter). For this reason, the sigmoid scheme will be briefly reviewed in Section 3, and used in the experiments of this paper.

The rest of the paper is organized as follows: in the next section we present the tracking data model, and carry out the analysis of the filter with bias. Section 3 reviews the practical scheme for adapting α, and Section 4 provides empirical evidence of both the validity of our analysis, and the ability of the real scheme to decrease the tracking error of NLMS and RLS filters. Finally, Section 5 states the most important conclusions of this work.

The work of J. Arenas-García was partly supported by MEC project TEC2008-02473. The work of M. Lázaro-Gredilla was partly supported by MICINN CONSOLIDER-INGENIO project CSD2008-00010 (COMONSENS).

2. TRACKING PERFORMANCE OF THE BIASED SCHEME

2.1. Tracking Data Model and Definitions

In the sequel, we adopt the following assumptions:

A1. $d(n)$ and $\mathbf{u}(n)$ are related via a linear regression model, $d(n) = \mathbf{w}_o^T(n)\mathbf{u}(n) + e_0(n)$, for some unknown time-varying weight vector $\mathbf{w}_o(n)$ of length M, and where $e_0(n)$ is an independent and identically distributed (i.i.d.) noise with zero mean and variance σ_0^2, and independent of $\mathbf{u}(m)$ for any n and m.

A2. First- and second-order moments of the input regressors are $E\{\mathbf{u}(n)\} = \mathbf{0}$ and $E\{\mathbf{u}(n)\mathbf{u}^T(n)\} = \boldsymbol{R}$.

A3. The unknown weight vector changes with time, according to the following tracking model taken from [1, Eq. (7.2.8)]:

$$\begin{cases} \mathbf{w}_o(n+1) &= \mathbf{w}_o + \boldsymbol{\theta}(n+1) \\ \boldsymbol{\theta}(n+1) &= \gamma\boldsymbol{\theta}(n) + \boldsymbol{q}(n) \end{cases} \tag{1}$$

with $0 \leq \gamma < 1$, and where $\boldsymbol{q}(n)$ is a sequence of i.i.d. perturbations with zero mean and covariance matrix $\boldsymbol{Q}$, independent of the input regressors and output noise at every time step. Using this model, it is easy to verify that the following expression, which will be used in the analysis, is satisfied:

$$\lim_{n\to\infty} E\{\mathbf{w}_o(n)\mathbf{w}_o^T(n)\} = \mathbf{w}_o\mathbf{w}_o^T + \frac{\boldsymbol{Q}}{1-\gamma^2}. \tag{2}$$

For the analysis, it will also be convenient to introduce some notation and additional variables. We define the weight error and *a priori* output error of the standard and biased filters as:

$$\boldsymbol{\varepsilon}(n) = \mathbf{w}_o(n) - \mathbf{w}(n), \qquad \boldsymbol{\varepsilon}_\alpha(n) = \mathbf{w}_o(n) - \mathbf{w}_\alpha(n),$$

$$e_a(n) = \mathbf{u}^T(n)\boldsymbol{\varepsilon}(n), \qquad e_{a,\alpha}(n) = \mathbf{u}^T(n)\boldsymbol{\varepsilon}_\alpha(n).$$

It can be easily checked that *a priori* errors are related to the total errors by $e_a(n) = e(n) - e_0(n)$ and $e_{a,\alpha}(n) = e_\alpha(n) - e_0(n)$.

To measure filter performance, we use the excess MSEs (EMSEs), which are defined as

$$J_{ex}(n) = E\{e_a^2(n)\}, \qquad J_{ex,\alpha}(n) = E\{e_{a,\alpha}^2(n)\},$$

whereas their limiting values as $n \to \infty$, which represent the steady-state tracking performance of the filters, will be denoted as $J_{ex}(\infty)$ and $J_{ex,\alpha}(\infty)$. Expressions for the steady-state EMSE of LMS (least mean squares) and NLMS (normalized LMS) filters under the tracking scenario described in A3 can be found in [1, pp. 396–397].

2.2. Mean-Square Peformance Analysis

We next study the tracking performance of filter $\mathbf{w}_\alpha(n)$. Similarly to what is done in [2, Eq. (5)] for the stationary case, the *a priori* error of filter with bias can be expressed as

$$e_{a,\alpha}(n) = \alpha e_a(n) + (1-\alpha)\mathbf{w}_o^T(n)\mathbf{u}(n). \tag{3}$$

Squaring this expression, and taking expectation, we obtain

$$\begin{aligned} J_{ex,\alpha}(n) = \alpha^2 J_{ex}(n) &+ (1-\alpha)^2 E\{\mathbf{w}_o^T(n)\mathbf{u}(n)\mathbf{u}^T(n)\mathbf{w}_o(n)\} \\ &+ 2\alpha(1-\alpha)E\{\boldsymbol{\varepsilon}^T(n)\mathbf{u}(n)\mathbf{u}^T(n)\mathbf{w}_o(n)\}. \end{aligned} \tag{4}$$

We can simplify this expression by taking into account that $\mathbf{w}_o(n)$ is independent of $\mathbf{u}(n)$:

$$\begin{aligned} J_{ex,\alpha}(n) = \alpha^2 J_{ex}(n) &+ (1-\alpha)^2 \mathrm{Tr}\left(\mathbf{R}E\{\mathbf{w}_o(n)\mathbf{w}_o^T(n)\}\right) \\ &+ 2\alpha(1-\alpha)\mathrm{Tr}\left(E\{\mathbf{u}(n)\mathbf{u}^T(n)\mathbf{w}_o(n)\boldsymbol{\varepsilon}^T(n)\}\right), \end{aligned} \tag{5}$$

where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix. Next, taking the limit as $n \to \infty$, and introducing also the common assumption that $\mathbf{w}(n)$ and $\mathbf{u}(n)$ are independent as $n \to \infty$, we obtain the following expression for the tracking EMSE of the biased filter with shrinkage factor α:

$$J_{ex,\alpha}(\infty) = \alpha^2 J_{ex}(\infty) + (1-\alpha)^2 A + 2\alpha(1-\alpha)B, \tag{6}$$

where we have used the following definitions:

$$\begin{aligned} A &= \mathrm{Tr}\left[\mathbf{R}\left(\mathbf{w}_o\mathbf{w}_o^T + \tfrac{\boldsymbol{Q}}{1-\gamma^2}\right)\right], \\ B &= \mathrm{Tr}\left[\mathbf{R}\lim_{n\to\infty} E\{\mathbf{w}_o(n)\boldsymbol{\varepsilon}^T(n)\}\right]. \end{aligned}$$

The limiting value of the expectation that appears in the definition of B will be dependent on the algorithm used for adapting $\mathbf{w}(n)$. For the LMS and NLMS cases, it is shown in the Appendix that

$$\lim_{n\to\infty} E\{\mathbf{w}_o(n)\boldsymbol{\varepsilon}^T(n)\} = \frac{1}{1+\gamma}\boldsymbol{Q}\left[\boldsymbol{I} - \gamma\left(\boldsymbol{I} - \frac{\mu}{c}\boldsymbol{R}\right)\right]^{-1}, \tag{7}$$

where μ is the step size, and $c = 1$ for LMS and $c = E\|\mathbf{u}(n)\|^2$ for NLMS.

Finally, taking derivatives of (6) with respect to α, and setting the result to zero, we obtain the value for the optimal shrinkage factor,

$$\alpha^* = \frac{B - A}{J_{ex}(\infty) + A - 2B}, \tag{8}$$

which substituting into (6) provides the optimal tracking EMSE of the biased configuration.

3. ADJUSTING THE SHRINKAGE FACTOR

In this section we review a practical algorithm for adapting the shrinkage factor α. It should be clear that in most practical situations we will lack the necessary knowledge for an optimal selection; furthermore, in many adaptive filtering applications the statistics change with time. This makes evident the need for algorithms that adapt uninterruptedly the value of $\alpha(n)$ to the current characteristics of the filtering scenario. By doing so, the biased configuration will show improved performance over the original filter, especially when the SNR or $\mathrm{Tr}(\boldsymbol{RQ})$ are unknown or time-varying.

In [2], different schemes for learning $\alpha(n)$ were investigated based on algorithms for adaptively combining adaptive filters that have appeared in the literature over recent years [4]– [8]. Among these algorithms, in this paper we consider the proposal from [7], which employs a sigmoid activation. According to our experiments in [2], the introduction of a non-linearity reduces the gradient noise near $\alpha = 1$, a case in which the biased scheme should work like the original filter.

To be more specific, $\alpha(n)$ is defined as the output of a sigmoid function,

$$\alpha(n) = \mathrm{sigmoid}[a(n)] = [1 + e^{-a(n)}]^{-1},$$

where the variable $a(n)$ is adapted at each iteration to minimize the overall error according to

$$\begin{aligned} a(n+1) &= a(n) - \frac{\mu_a}{p(n)} \cdot \frac{\partial e_\alpha^2(n)}{\partial a(n)} \\ &= a(n) + \frac{\mu_a}{p(n)} e_\alpha(n) y(n) \frac{\partial \alpha(n)}{\partial a(n)}, \end{aligned} \tag{9}$$

where

$$p(n) = \beta p(n-1) + (1-\beta)y^2(n)$$

is a low-pass filtered estimation of the power of $y(n)$, and from which $\alpha(n+1)$ can readily be recovered. Selection of β is not critical for appropriate performance of the algorithm, and we will simply set it to 0.9, as was done in [7], [8].

Following [2], we truncate $a(n)$ to the interval $[-a^+, a^+]$ to prevent adaptation rule (9) from freezing, and use the slightly modified definition

$$\alpha(n) = \frac{\text{sigmoid}[a(n)] - \text{sigmoid}[-a^+]}{\text{sigmoid}[a^+] - \text{sigmoid}[-a^+]}, \tag{10}$$

so that $\alpha(n)$ takes values 0 and 1 for the minimum and maximum values of $a(n)$, respectively.

The advantages of introducing the activation function (10) are twofold. First, it keeps $\alpha(n)$ within the reasonable interval $[0, 1]$ (note that, otherwise, the variance of the estimation error could be amplified). Second, since the derivative of (10) is small near $\alpha(n) = 0$ and $\alpha(n) = 1$, the adaptation speed of (9) is reduced near these points. Note that this is convenient, especially near $\alpha(n) = 1$, since it prevents adaptation of the shrinkage factor from degrading the performance of the original filter in this case.

4. EXPERIMENTS

In this section, we will investigate the tracking performance of the biased NLMS as compared with its standard version, and verify the quality of the theoretical expressions introduced in this work. Analogous behavior can be obtained with other filtering schemes; we will illustrate this empirically for the RLS case.

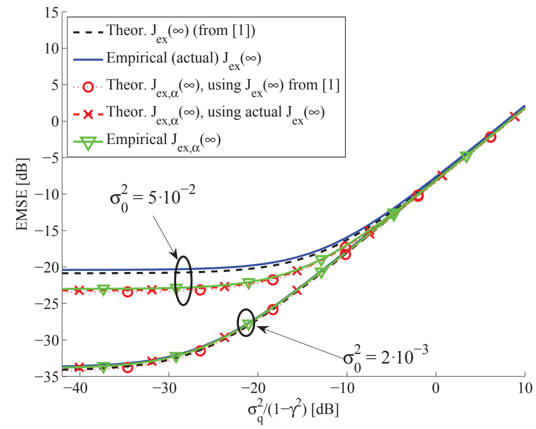
The tracking scenario for our experiments is as follows: Coefficients $\mathbf{w}_o$, of length $M = 16$, are generated i.i.d. from a zero-mean Gaussian distribution and scaled so that $||\mathbf{w}_o||^2 = 1$. Input regressors $\mathbf{u}(n)$ are obtained from a zero-mean i.i.d. Gaussian random process $u(n)$ with $\sigma_u^2 = 10^{-2}$, so that $\mathbf{u}(n) = [u(n), u(n-1), \ldots, u(n-M+1)]^T$.

We consider two different values for the output noise $\sigma_0^2 = \{2 \cdot 10^{-3}, 5 \cdot 10^{-2}\}$ to illustrate operation under different SNR conditions. The driving force for the random walk model (1), σ_q^2, is logarithmically swept between 10^{-5} and 2, with $\gamma = 0.9$ in all cases. Following [2], we fix the step size for adaptation of the shrinkage factor to $\mu_a = 0.1$. Each data point in the following plots corresponds to averaging 100 independent realizations, and for 10 000 time steps. An initial "burn-in" period of another 10 000 time steps is used to guarantee that the filter is in stable tracking operation.

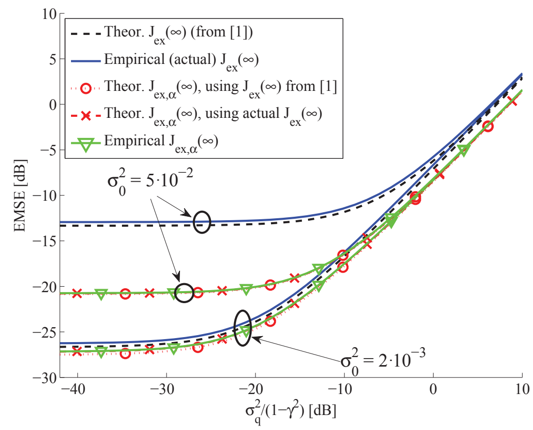
In addition to the experimental results obtained from the filter operation, it is also possible to obtain theoretical approximations for $J_{\text{ex},\alpha}(\infty)$ and α^* using (6) and (8), respectively. The quality of these approximations crucially depends on the quality of the required $J_{\text{ex}}(\infty)$. Therefore, we will present results based both on the actual $J_{\text{ex}}(\infty)$ as obtained from NLMS operation [theoretical, using actual $J_{\text{ex}}(\infty)$], and on the approximation provided in [1, p. 393] for the $J_{\text{ex}}(\infty)$ of the NLMS filter [theoretical, using $J_{\text{ex}}(\infty)$ from [1]].

4.1. NLMS tracking performance

Fig. 2 shows the EMSE of the standard NLMS filter [$J_{\text{ex}}(\infty)$] and of its biased counterpart [$J_{\text{ex},\alpha}(\infty)$] as a function of σ_q^2, for two different step sizes ($\mu \in \{0.3, 1\}$). Both empirical and theoretical values are plotted. Both panels show good agreement between the empirically obtained $J_{\text{ex},\alpha}(\infty)$ and the theoretical approximation (6) introduced in this work, independently of whether it is used with the actual $J_{\text{ex}}(\infty)$ or the approximate $J_{\text{ex}}(\infty)$. This latter approximate $J_{\text{ex}}(\infty)$, provided by [1], is also shown to be reasonably accurate.



(a) EMSE for $\mu = 0.3$



(b) EMSE for $\mu = 1$

Fig. 2. EMSE of the standard and biased NLMS for different values of μ and σ_0^2. In addition to the experimental results, we show the approximate $J_{\text{ex}}(\infty)$ from [1], and the novel approximate $J_{\text{ex},\alpha}(\infty)$ from (6) [using both the actual and approximate $J_{\text{ex}}(\infty)$].

Two operation modes, separated by a clear "knee", can be observed in these plots. In the left area (small σ_q^2), the tracking speed is so slow that the NLMS behaves as in the stationary case [2]. At the very "knee" point and to its right, the NLMS is actually in tracking operation. In all cases, and across a wide range of output noises and step sizes, the biased filter provides consistent improvement.

Note that the biased filter, just as the standard NLMS, does not have any information regarding the data-generating process. This implies that the advantages of biasing are still present in situations in which σ_0^2, σ_q^2, and/or γ are not known or change over time.

Fig. 3 shows α as estimated by (9) and its theoretically optimal value α^* for $\mu = 0.3$ (analogous results obtained for $\mu = 1$ are omitted). When α^* is computed using the actual $J_{\text{ex}}(\infty)$, an almost-perfect match between the theoretical and empirical value is obtained, thus proving the accuracy of (8). However, if the approximate $J_{\text{ex}}(\infty)$ from [1] is used in (8), this correspondence degrades as σ_q^2 grows and the NLMS enters tracking operation. Thus, the accuracy of the approximation for α^* is more dramatically affected by the quality of the

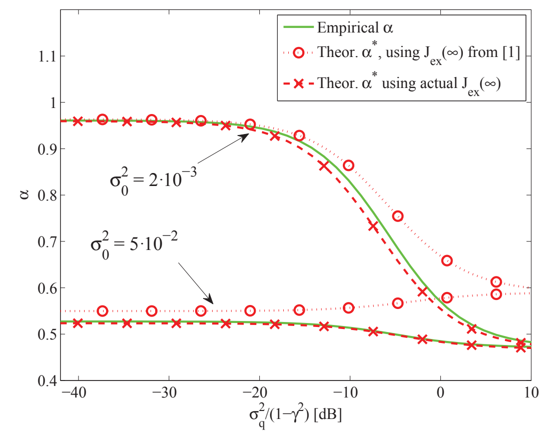


Fig. 3. Shrinkage factor α corresponding to $\mu = 0.3$. The experimental value is plotted together with the optimal value provided by (8) [using both the actual and approximate $J_{\text{ex}}(\infty)$].

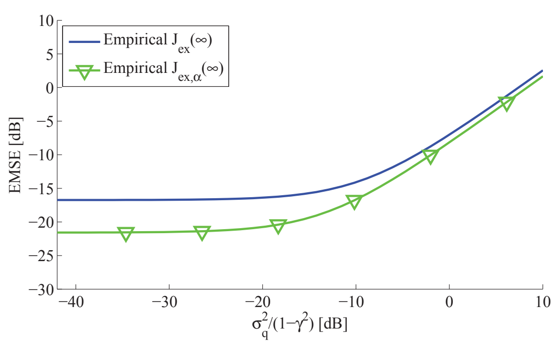


Fig. 4. EMSE of the standard and biased RLS with settings $\lambda = 0.95$ and $\sigma_0^2 = 5 \cdot 10^{-2}$ (analogous results obtained for other cases not shown here).

available $J_{\text{ex}}(\infty)$ than the accuracy of the theoretical approximation for $J_{\text{ex},\alpha}(\infty)$ was.

4.2. Tracking performance of other filtering schemes

Though we have investigated in detail a biased NLMS filter here, the performance of other filters can also be enhanced with the proposed biased configuration. For example, we have obtained analogous results for the RLS filter, where for a wide range of scenarios, the biased filter always showed improvement over the standard, unbiased version. A typical result is illustrated in Fig. 4.

5. CONCLUSIONS

Biasing the weights of adaptive filters is an interesting way of reducing their EMSE. In this paper, we have studied the tracking pefformance of a simple biased scheme, which multiplies the output of the filter by a shrinkage factor α. In addition to the already known advantages of such scheme in stationary situations [2], the new theoretical and simulation work allows us to conclude that biased configurations can also contribute to a lower tracking error. The bottom line is that biased filters can outperform standard ones in certain situations, and thus become useful in situations where the statistics are not known *a priori* or change unpredictably.

APPENDIX

In this Appendix we analyze the value of $\lim_{n\to\infty} E\{\mathbf{w}_o(n)\boldsymbol{\varepsilon}^T(n)\}$ for the particular case of LMS and NLMS filters using the data model and definitions in Subsection 2.1. We start by recalling that the adaptation rule for these filters is given by $\mathbf{w}(n+1) = \mathbf{w}(n) + \tilde{\mu}\mathbf{u}(n)e(n)$, where $\tilde{\mu} = \mu/c$, with $c = 1$ for LMS, and $c = \|\mathbf{u}(n)\|^2$ for NLMS, which will be approximated in the analysis by its expected value $c = E\|\mathbf{u}(n)\|^2$ for tractability reasons.

Subtracting both terms from $\mathbf{w}_o(n+1)$, and taking into account that according to the tracking model in A3 it can be shown that $\mathbf{w}_o(n+1) = \mathbf{w}_o(n) - (1-\gamma)\boldsymbol{\theta}(n) + \boldsymbol{q}(n)$, we arrive at

$$\boldsymbol{\varepsilon}(n+1) = [\boldsymbol{I} - \tilde{\mu}\mathbf{u}(n)\mathbf{u}^T(n)]\boldsymbol{\varepsilon}(n) - (1-\gamma)\boldsymbol{\theta}(n) + \boldsymbol{q}(n) + \tilde{\mu}\boldsymbol{\phi}(n),$$

where $\boldsymbol{\phi}(n) = -\mathbf{u}(n)e_0(n)$. Since the expression above will only be used in the limit as $n \to \infty$, and assuming independent $\mathbf{w}(n)$ and $\mathbf{u}(n)$ in steady-state, the term inside square brackets in the expression above can be approximated as $C \approx [\boldsymbol{I} - \tilde{\mu}\boldsymbol{R}]$. The above recursion allows us to obtain the general term for the weight error vector as

$$\boldsymbol{\varepsilon}(n) = C^n\boldsymbol{\varepsilon}(0) + \sum_{i=0}^{n-1} C^{n-1-i}[\boldsymbol{q}(i) - (1-\gamma)\boldsymbol{\theta}(i) + \tilde{\mu}\boldsymbol{\phi}(i)]. \quad (11)$$

Premultiplying the transpose of (11) by $\mathbf{w}_o(n) = \mathbf{w}_o + \boldsymbol{\theta}(n)$, and taking expectations of the result, leads to

$$E\{\mathbf{w}_o(n)\boldsymbol{\varepsilon}^T(n)\} = \sum_{i=0}^{n-1} C^{n-1-i} E\{\boldsymbol{\theta}(n)[\boldsymbol{q}^T(i) - (1-\gamma)\boldsymbol{\theta}^T(i)]\}.$$

If we now take into account that, according to the tracking model for $\mathbf{w}_o(n)$, we have $\boldsymbol{\theta}(n) = \sum_{j=0}^{n-1} \gamma^{n-1-j}\boldsymbol{q}(j)$, and after some algebra, we obtain

$$E\{\mathbf{w}_o(n)\boldsymbol{\varepsilon}^T(n)\} = \frac{1}{1+\gamma}\boldsymbol{Q}(\boldsymbol{I} - \gamma C)^{-1}(\boldsymbol{I} - \gamma^n C^n) + \frac{1}{1+\gamma}\boldsymbol{Q}(\boldsymbol{I} - \gamma^{-1}C)(\gamma^{2n-1}\boldsymbol{I} - \gamma^{n-1}C^n). \quad (12)$$

Finally, since $|\gamma| < 1$ and for convergent algorithms C^n vanishes as $n \to \infty$, the limiting value of the above expression gives (7).

6. REFERENCES

[1] A. H. Sayed, *Fundamentals of Adaptive Filtering*, Wiley, 2003.

[2] M. Lázaro-Gredilla, L. A. Azpicueta-Ruiz, A. R. Figueiras-Vidal, and J. Arenas-García, "Adaptively biasing the weights of adaptive filters," *IEEE Trans. Signal Process.*, vol. 58, pp. 3890–3895, Jul. 2010.

[3] S. Kay and Y. C. Eldar, "Rethinking biased estimation," *IEEE Signal Process. Mag.*, vol. 25, pp. 133–136, Mar. 2008.

[4] S. S. Kozat and A. C. Singer, "Multi-stage adaptive signal processing algorithms," in *Proc. of SAM Signal Proc. Workshop*, 2000, pp. 380–384.

[5] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," *IEEE Trans. Signal Process.*, vol. 54, pp. 1078–1090, Mar. 2006.

[6] J. Arenas-García, M. Martínez-Ramón, A. Navia-Vázquez, and A. R. Figueiras-Vidal, "Plant identification via adaptive combination of transversal filters", *Signal Processing*, vol. 86, pp. 2430-2438, Sep. 2006.

[7] L. A. Azpicueta-Ruiz, A. R. Figueiras-Vidal, and J. Arenas-García, "A normalized adaptation scheme for the convex combination of two adaptive filters," in *Proc. ICASSP*, Las Vegas, NV, 2008, pp. 3301–3304.

[8] R. Candido, M. T. M. Silva, V. H. Nascimento, "Affine combination of adaptive filters," in *Proc. 42nd Asilomar Conf. on Signal, Systems and Computers,* Oct. 2008, pp. 236–240.