Analysis of the Desired-Response Influence on the Convergence of Gradient-Based Adaptive Algorithms

Luis Vicente, *Member, IEEE*, and Enrique Masgrau, *Member, IEEE*

***Abstract*—Although the convergence behavior of gradient-based adaptive algorithms, such as steepest descent and leas mean square (LMS), has been extensively studied, the influence of the desired response on the transient convergence has generally received little attention. However, empirical results show that this signal can have a great impact on the learning curve. In this paper we analyze the influence of the desired response on the transient convergence by making a novel interpretation, from the viewpoint of the desired response, of previous convergence analyses of SD and LMS algorithms. We show that, without prior knowledge that can be used to wisely select the initial weight vector, initial convergence is fast whenever there is high similarity between input and desired response whereas, on the contrary, when there is low similarity between these two signals, convergence is slow from the beginning.**

***Index Terms*—Adaptive filters, adaptive signal processing, convergence, gradient methods, least-mean-square (LMS) methods.**

I. Introduction

THE similarity between input and desired response plays a significant role in the adaptation process of gradient-based algorithms, such as steepest descent (SD) or least mean square (LMS). Empirical results support this observation, since clearly distinct convergence rates are obtained for different desired responses sharing the same input. However, the desired response signal has received little attention in most convergence analyses, even though it is present in all of them, through the cross-correlation vector. This is due to practical reasons, since it is not generally possible to choose the desired response and, also, a proper selection of the initial weight vector can mitigate any influence that this signal might play. Nevertheless, in order to wisely select the initial weight vector, prior information of the environment should be available, which is not always the case. On the other hand, the analysis proposed in this paper can provide a deeper understanding of the algorithms convergence. Moreover, this study provides a satisfactory explanation to why a system with an adaptive prewhitener of the input converges almost always faster than the system without it, which is indeed the topic that prompted this analysis.

So, in this paper, the effect of the desired response on the MSE transient behavior of SD and LMS algorithms is inspected, starting from previous and well-established convergence analyses [1], [2]. The main conclusion is that, without any prior information that can be used to wisely choose the initial weight vector, convergence is slow when there is low similarity between input and desired response, but it is fast, at least at the beginning, when the similarity between these two signals is high. Although this result might seem quite logical or intuitive, it has not been proven before, to the best of the authors knowledge. Also, the analysis is clearly supported by empirical observations.

In Section II, some results from previous convergence analyses of SD and LMS algorithms are recalled in order to make explicit the modes excitation dependence on the cross-correlation vector. For the LMS algorithm, the small-step-size theory is considered, avoiding the classical and restrictive independence assumption, and yielding more accurate results [2, Sec. 5.4], [3], [4]. A measure of the similarity between input and desired response, suitable for the finite-impulse response (FIR) adaptive filtering context, is also defined in Section II. In Section III, the simple case of a sinusoidal input with a two-tap adaptive filter is considered first. Two extreme desired responses, from the point of view of similarity with the input, are analyzed in detail and a geometrical interpretation of the results is provided via an example. Section IV deals with the more general case of wide-sense stationary stochastic processes as input and desired response. Again, two extreme desired responses are analyzed. The germ of the work introduced in this section was presented in [5]. Empirical results are also provided. Finally, the performance of a system with adaptive prewhitening of the input is considered in Section V, in order to explain why it converges almost always faster than the original system.

II. Convergence Dependence on Cross-Correlation Vector

According to most convergence analyses [1], [2], [6], the mean-square error (MSE) evolution with time, for SD and LMS algorithms, is given by the following general expression:

$$J[n] = E(|e[n]|^2) = J[\infty] + \sum_{k=1}^{N} A_k (1-\mu\lambda_k)^{2n} \qquad (1)$$

where μ is the step size of the algorithm, N is the length of the FIR adaptive filter, and λ_k are the eigenvalues of the correlation matrix $\mathbf{R} = E(\mathbf{x}[n]\mathbf{x}^H[n])$ of the input vector $\mathbf{x}[n]$. The superscript H denotes Hermitian transposition. The actual values for the final MSE $J[\infty]$ and the initial amplitude of each exponential function A_k depend on the adaptive algorithm being examined and, in the case of the LMS algorithm, also on the assumptions made to arrive at (1).

Manuscript received January 8, 2007; revised July 6, 2007, and September 13, 2007. This work was supported in part by the Ministerio de Educacion y Ciencia (Spain) under Project TIN2005-08660-C04-01. This paper was recommended by Associate Editor P. Regalia.

The authors are with the Aragon Institute for Engineering Research, University of Zaragoza, E50018 Zaragoza, Spain (e-mail: lvicente@unizar.es; masgrau@unizar.es).

Digital Object Identifier 10.1109/TCSI.2008.916690

Thus, in gradient-based adaptive systems the learning curve (1) consists of a sum of exponentials, each of which corresponds to a natural mode of the algorithm. The exponential decay for the kth natural mode has a time constant given by

$$\tau_{k,\mathrm{mse}} = \frac{-1}{2\ln(1-\mu\lambda_k)}. \tag{2}$$

Since the step size is the same for every convergence mode, it is its associated eigenvalue λ_k what determines whether a mode is relatively fast or slow. Thus, the statistical properties of the input to the adaptive filter are the one and only responsible for the detached convergence rate of the different modes.

Other mode feature as important as the time constant is the *excitation*, or initial amplitude A_k, since it determines whether and when a mode is dominant on global convergence. The natural mode being dominant at a specific time instant is the one whose contribution to the sum in (1) is the most important one. Thus, whenever the fast modes are much more excited than the slow ones the initial convergence of the algorithm is fast, even with great eigenvalue spread of the correlation matrix. Certainly, the slow modes will always become dominant in the end, but the point is how much convergence is left when this happens. So, when we are concerned with convergence rate of a gradient-based adaptive algorithm, we cannot obviate the relative excitation of natural modes.

The geometrical interpretation of the cost function (1) helps to clarify this point. It is evident that, depending on the departing point of the weight value track on the bowl-shaped error surface, convergence rate varies greatly. This topic has sometimes been referred to as the directionality of convergence for nonwhite inputs [2, Sec. 5.9]. Obviously, the actual weight value track depends on the initial weight vector, $\mathbf{w}[0]$, but also on the optimal solution for the weight vector, $\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p}$, which is clearly influenced by the desired response $d[n]$ through the cross-correlation vector $\mathbf{p} = E(\mathbf{x}[n]d^*[n])$.

A. Steepest Descent Modes Excitation

The SD method is described by the following adaptation formula:

$$\begin{aligned}\mathbf{w}[n+1] &= \mathbf{w}[n] - \frac{1}{2}\mu\nabla J[n] \\ &= \mathbf{w}[n] + \mu(\mathbf{p} - \mathbf{R}\mathbf{w}[n]).\end{aligned} \tag{3}$$

The stability analysis of this adaptive algorithm [1], [2, Sec. 4.3] establishes the following evolution with time for the MSE:

$$J[n] = J_{\min} + \sum_{k=1}^{N} \lambda_k |v_k[0]|^2 (1-\mu\lambda_k)^{2n} \tag{4}$$

where $J_{\min}$ is the minimum MSE, obtained with the optimum weight vector $\mathbf{w}_o$, and $v_k[0]$ is the kth component of the initial transformed weight-error vector

$$\mathbf{v}[0] = \mathbf{Q}^H(\mathbf{w}_o - \mathbf{w}[0]). \tag{5}$$

$\mathbf{Q}$ is the unitary matrix of eigenvectors resulting from the eigendecomposition of the correlation matrix, $\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H$, $\mathbf{\Lambda}$ being the diagonal matrix of associated eigenvalues.

From (4) we see that the initial amplitude or excitation of each mode in the SD method is

$$A_{k,\mathrm{SD}} = \lambda_k |v_k[0]|^2. \tag{6}$$

B. LMS Modes Excitation

The LMS algorithm adaptive recursion is given by

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \mu\mathbf{x}[n]e^*[n]. \tag{7}$$

Classical analyses of the LMS algorithm [6, Sec. 9.4],[7]–[9], are based on the *independence assumption*. Nevertheless, it is well-known that this assumption is clearly violated in many practical situations, such as the case of a tapped-delay line adaptive filter. More recent analyses have also been developed under the *small-step-size theory* [2, Sec. 5.4], [3], [4], whose assumptions, apart from the step size being small, $\mu \ll 2/\lambda_{\max}$, are much less restrictive. Although both theories lead to similar results, in the case of the small-step-size theory, they are more accurate and elegant.

Thus, according to the small-step-size theory, the LMS cost function can be expressed, when the step size μ is small, as

$$J[n] \approx J[\infty] + \sum_{k=1}^{N} \lambda_k \left(|v_k[0]|^2 - \frac{\mu J_{\min}}{2}\right)(1-\mu\lambda_k)^{2n} \tag{8}$$

where the final MSE is

$$J[\infty] = J_{\min} + \mu J_{\min} \sum_{k=1}^{N} \frac{\lambda_k}{2-\mu\lambda_k}. \tag{9}$$

Therefore, the amplitude of each natural mode in the LMS case is given by

$$A_{k,\mathrm{LMS}} = \lambda_k \left(|v_k[0]|^2 - \frac{\mu J_{\min}}{2}\right). \tag{10}$$

C. Cross-Correlation Vector Influence

The coefficients of the initial transformed weight-error vector, $v_k[0]$, which appear in both (6) and (10), are hiding the influence of the desired response of the adaptive filter on its convergence. Recalling (5) and taking into account the optimum Wiener filter solution,

$$\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p} = \mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}^H\mathbf{p} \tag{11}$$

and the unitary property of the eigenvectors matrix $\mathbf{Q}^{-1} = \mathbf{Q}^H$, it follows that

$$\mathbf{v}[0] = \mathbf{\Lambda}^{-1}\mathbf{Q}^H\mathbf{p} - \mathbf{Q}^H\mathbf{w}[0] \tag{12}$$

or, equivalently

$$v_k[0] = \lambda_k^{-1}\mathbf{q}_k^H\mathbf{p} - \mathbf{q}_k^H\mathbf{w}[0] \tag{13}$$

where $\mathbf{q}_k$ is the kth column of matrix $\mathbf{Q}$, that is, the eigenvector associated with the eigenvalue λ_k.

Using this last result in the expressions for the modes excitation, (6) and (10), the influence of the cross-correlation between input and desired response on convergence is patent. Obviously, from (13) we can also say that this influence can be counteracted by properly selecting the initial solution for the weight vector. However, some prior information is compulsory in order to choose a sensible initial weight vector. As our analysis is focused on the influence of the desired response, we assume there is no such prior knowledge, and thus, the initial weight vector can be thought of as being random.

Also, one particular but very common initialization of the weight vector, when there is no prior information, is the null one, $\mathbf{w}[0] = \mathbf{0}$, since it ensures no increment in the MSE level when turning the adaptive algorithm on. In this special case, the transformed weight-error vector $\mathbf{v}$ can be viewed from (12) as the result of a weighted projection of the cross-correlation vector $\mathbf{p}$ on each of the eigenvectors of the correlation matrix, with the inverse of the eigenvalues, λ_k^{-1}, as weighting factors.

D. Similarity Definition

At this point, we aim to define a suitable measure of the similarity between input and desired response. This is due to the fact that we want to inspect the differences in the algorithms convergence depending on this similarity.

Since the influence of the desired response on the SD or LMS convergence is channeled by the cross-correlation vector $\mathbf{p}$, it is clear that the similarity concept has to be closely related to the cross-correlation function between input and desired response. In fact, the cross-correlation function itself, $r_{xd}[k] = E(x[n]d^*[n-k])$, can be interpreted as a similarity measure. When we define the Euclidean distance between two stationary signals, with a possible time lag, we have

$$\begin{aligned} D[k] &= E(|x[n] - d[n-k]|^2) \\ &= P_x + P_d - 2\mathrm{Re}(r_{xd}[k]) \end{aligned} \tag{14}$$

where P_x and P_d are the mean-square values of both signals. Therefore, for fixed P_x and P_d values, the greater the real part of the cross-correlation function $\mathrm{Re}(r_{xd}[k])$, the lesser the distance between the two signals $D[k]$.

However, given to the role of these signals in the adaptive filtering problem, we should normalize them prior to measuring their similarity (or their distance). This is due to the fact that a constant scale change in one of the signals does not mean more or less difficulties for the convergence of the algorithms, as long as we can properly modify the step size. Therefore, a constant scale change in any of the two signals should not change their similarity measure.

We define the normalized signals, with unit power, as

$$\bar{x}[n] = \frac{x[n]}{\sqrt{P_x}} \tag{15}$$

$$\bar{d}[n] = \frac{d[n]}{\sqrt{P_d}}. \tag{16}$$

For the very common case of real signals, this amplitude normalization is enough. However, in the more general case of complex signals, the phase of the signals has also to be taken into account in the normalization process. Thus, the normalized signals are defined in this case by

$$\bar{x}[n] = \frac{x[n]}{\sqrt{P_x}} e^{-j\phi_x} \tag{17}$$

$$\bar{d}[n] = \frac{d[n]}{\sqrt{P_d}} e^{-j\phi_d} \tag{18}$$

where ϕ_x and ϕ_d are the mean phases[1] of the input and the desired response, respectively.

Once we have normalized the amplitude of the signals and aligned them with the real axis, their cross-correlation function can be seen, for our purposes, as a proper similarity function at a given time lag. However, one final aspect that we still have to consider is that for FIR adaptive filters, the cross-correlation vector $\mathbf{p}$ consists of several cross-correlation values, for different time lags. Hence, our similarity measure has also to take this into account by accumulating the corresponding values of the cross-correlation function of the normalized signals. Thus, the similarity measure between the input and the desired response that we consider for an FIR adaptive filter of order N is

$$S(x[n], d[n], N) = \sum_{k=0}^{N-1} |\mathrm{Re}(r_{\bar{x}\bar{d}}[-k])|. \tag{19}$$

III. Sinusoidal Case

Although the derivation of adaptive algorithms is in general based on wide-sense stationary stochastic signals, they are also applicable to deterministic environments. In this section, we analyze the very common case of real sinusoidal inputs and desired responses.

Let the input be

$$x[n] = C_x \cos(\omega n + \theta_x). \tag{20}$$

The desired response is composed of a sinusoid, of the same frequency as the input, plus any uncorrelated noise

$$d[n] = C_d \cos(\omega n + \theta_d) + u[n]. \tag{21}$$

In the following analysis, we consider a fixed phase difference:

$$\theta = \theta_d - \theta_x \tag{22}$$

[1]We define the *mean phase* of a signal $s[n]$ as

$$\phi_s = \frac{1}{2j} \ln E\left(\frac{s[n]}{s^*[n]}\right).$$

Note that this definition has an ambiguity of π, that is, $-\pi/2 < \phi_s \leq \pi/2$. For instance, in the particular case of a real signal, the mean phase is always 0, independently of whether the signal predominantly takes positive or negative values. The absolute value in the similarity measure (19) aims to cancel the effect of this ambiguity.

between the desired response and the input. This is a realistic situation, since it implies that there is a fixed filter relating both signals. Thus, the analysis corresponds to a deterministic situation, even in the case that the sinusoids in (20) and (21) were stochastic processes due to a possible unknown initial phase.

We only consider the case of an adaptive filter with two taps, $N = 2$, since it is the one with the greatest eigenvalue spread and, on the other hand, it also allows a graphical representation of the results, when we sketch the evolution of the weights on the error surface.

The behavior of adaptive algorithms when the input is deterministic may be quite different than when the input is stochastic, due to non-Wiener solutions of the algorithms [10]–[13]. However, under the assumption of small step size, (4) is still valid for the SD algorithm even for deterministic inputs. Moreover, as long as the uncorrelated noise component $u[n]$ in (21) is white, (8) is also applicable for the LMS algorithm in the previous scenario [2].

So, according to (6) and (10), we are now interested in the initial value of the transformed weight-error vector components for this sinusoidal case. For a tapped delay line configuration with two filter weights, it can be easily shown (see Appendix I) that

$$v_k[0] = \frac{C_d}{C_x\sqrt{2}}\left(\cos\theta \pm \sin\theta \sin\omega \frac{C_x^2}{2\lambda_k}\right) - \mathbf{q}_k^H \mathbf{w}[0]. \quad (23)$$

The negative sign in the term between brackets in (23) corresponds to the eigenvalue $\lambda_1 = C_x^2(1+\cos\omega)/2$ whereas the positive sign corresponds to the eigenvalue $\lambda_2 = C_x^2(1-\cos\omega)/2$.

The similarity measure in this simple sinusoidal case is

$$S = |\cos\theta| + |\cos(\omega+\theta)|. \quad (24)$$

In the following subsections, we analyze two extreme cases of similarity between desired response and input, with constant phase difference θ. This cases are the in-phase and quadrature sinusoidal signals, respectively. It is easy to check that the similarity (24) for the in-phase case is always greater than for the quadrature case (with the only exception of $\omega = \pi/2$, when both similarities are equal).

A. In-Phase (Or Counter-Phase) Desired Response

When the phase difference between input and desired response is $\theta = 0$ or $\pi, \cos\theta = \pm 1, \sin\theta = 0$ and (23) becomes

$$v_k[0] = \pm\frac{C_d}{C_x\sqrt{2}} - \mathbf{q}_k^H \mathbf{w}[0] \quad (25)$$

where the positive sign of the first term corresponds to a phase difference $\theta = 0$ and the negative sign to $\theta = \pi$. The similarity measure is, in this case,

$$S = 1 + |\cos\omega| \geq 1. \quad (26)$$

Using (25) in (6) and (10), we get the initial amplitudes of SD and LMS modes

$$A_{k,\mathrm{SD}} = \lambda_k \left|\pm\frac{C_d}{C_x\sqrt{2}} - \mathbf{q}_k^H \mathbf{w}[0]\right|^2 \quad (27)$$

$$A_{k,\mathrm{LMS}} = \lambda_k \left(\left|\pm\frac{C_d}{C_x\sqrt{2}} - \mathbf{q}_k^H \mathbf{w}[0]\right|^2 - \frac{\mu J_{\min}}{2}\right). \quad (28)$$

At first sight, we can say that, in this case of high similarity between input and desired response, the excitation of each SD and LMS mode is directly proportional to its own eigenvalue. Therefore, the fast mode is initially much more excited than the slow one and, so, fast initial convergence can be expected.

Of course, it can be argued that the term $\mathbf{q}_k^H \mathbf{w}[0]$, which is different for each eigenvalue, could as well turn convergence slow. However, we have assumed a random initial weight vector $\mathbf{w}[0]$, due to the lack of prior knowledge of the environment where the adaptive algorithm works. In this case, it seems fair to suppose that the expectation of the term $\mathbf{q}_k^H \mathbf{w}[0]$ is the same for both eigenvalues, and thus, the "expected" excitation for each mode is indeed directly proportional to its own eigenvalue.

B. Desired Response in Phase Quadrature

In the event of a phase shift $\theta = \pi/2$ between input and desired response, $\cos\theta = 0, \sin\theta = 1$ and (23) becomes

$$v_k[0] = \pm\frac{C_d C_x \sin\omega}{2\sqrt{2}\lambda_k} - \mathbf{q}_k^H \mathbf{w}[0] \quad (29)$$

where the sign of the first term is negative for the eigenvalue $\lambda_1 = C_x^2(1+\cos\omega)/2$ and positive for the eigenvalue $\lambda_2 = C_x^2(1-\cos\omega)/2$. When $\theta = -\pi/2$, (29) is also valid, but the signs corresponding to each eigenvalue are exchanged. The similarity measure is now

$$S = |\sin\omega| \leq 1. \quad (30)$$

Again, substituting (29) in (6) and (10), the initial amplitudes of the modes are

$$A_{k,\mathrm{SD}} = \lambda_k \left|\pm\frac{C_d C_x \sin\omega}{2\sqrt{2}\lambda_k} - \mathbf{q}_k^H \mathbf{w}[0]\right|^2 \quad (31)$$

$$A_{k,\mathrm{LMS}} = \lambda_k \left(\left|\pm\frac{C_d C_x \sin\omega}{2\sqrt{2}\lambda_k} - \mathbf{q}_k^H \mathbf{w}[0]\right|^2 - \frac{\mu J_{\min}}{2}\right). \quad (32)$$

For the following discussion, we think of a case with great eigenvalue spread, that is, the eigenvalue associated to the fast mode is very large while the one associated to the slow mode is very small. We can conclude from (31) and (32) that the excitation of the fast mode will be directly proportional to its own eigenvalue whereas the excitation of the slow mode will be inversely proportional to its own eigenvalue. Thus, it seems that both modes will be greatly excited.

However, to obtain fast initial convergence, the fast mode needs to be much more excited than the slow one in order to be dominant for a long period of time. So, in this low similarity

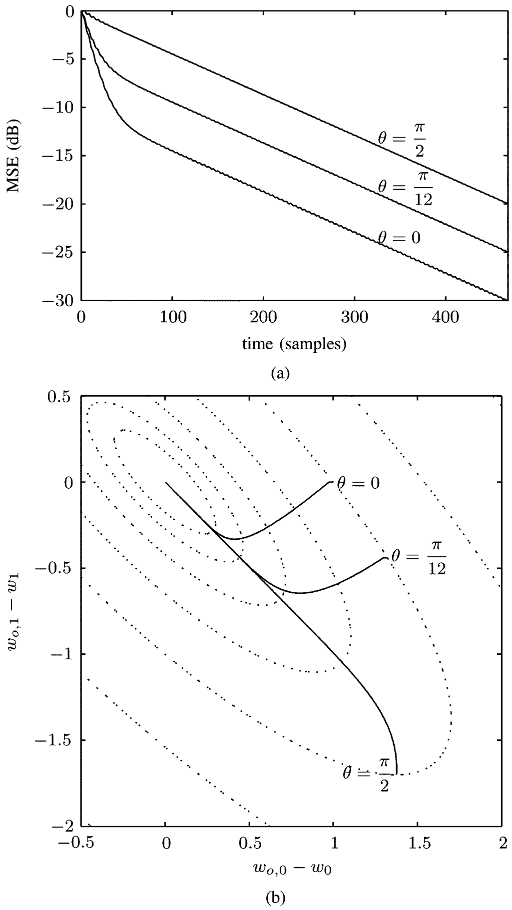

Fig. 1. LMS convergence for a two-tap adaptive filter with sinusoidal input, $\omega = \pi/5(\chi(\mathbf{R}) = 9.47)$, and different phase shifts between input and desired response. (a) Ensemble average learning curves. (b) Weight-value tracks on the error surface.

case, convergence will be slow, since the slow mode is very excited and can dominate convergence from the very beginning, making the fast mode go unnoticed.

In the same way as in the high similarity case, the term $\mathbf{q}_k^H \mathbf{w}[0]$ introduces some randomness in the excitation of both modes. Therefore, the previous discussion refers also to the "expected" excitation of natural modes. In the particular case of null initial filter weights there is no randomness. Furthermore, in this case the excitation of the SD fast mode is also inversely proportional to its own eigenvalue, and so slow convergence is even more evident.

C. Example

In the next example, we consider a two-tap LMS adaptive filter with sinusoidal input, $x[n] = \cos(\omega n + \theta_x)$, for different phase shifts between input and desired response. In order to see more clearly the two convergence modes in the learning curves, the desired responses do not contain any uncorrelated noise, $d[n] = \cos(\omega n + \theta_d)$. Thus, total cancellation, $J_{\min} = 0$ ($-\infty$ dB), is in this case possible, at least theoretically (but for machine precision).

Fig. 1(a) shows the ensemble average learning curves obtained from 1000 runs of the LMS algorithm for a frequency $\omega = \pi/5$ and three different phase shifts between input and desired response, $\theta = 0, \pi/12$, and $\pi/2$ (with respective similarities $S = 1.81, 1.60$ and 0.59). For each run, the initial phase of the input, θ_x, is a random variable with uniform distribution between 0 and 2π. However, the phase difference between input and desired response is fixed. The initial weight vector is $\mathbf{w}[0] = \mathbf{0}$. In this case, the two eigenvalues are $\lambda_{\max} = 0.9045$ and $\lambda_{\min} = 0.0955$, and therefore, the eigenvalue spread is $\chi(\mathbf{R}) = 9.47$. The step size is $\mu = 0.05$. Each mode individually converges at a rate of -0.4020 dB/sample and -0.0416 dB/sample, respectively. Both modes, with their different slopes, are evident in the learning curves in Fig. 1(a).

Analyzing the learning curves, we see that when $\theta = 0$, initial convergence is quite fast, until the slow mode becomes dominant. On the contrary, when $\theta = \pi/2$ the fast mode goes completely unnoticed, since the slow mode dominates convergence from the beginning. Last, when $\theta = \pi/12$ we get an intermediate situation of the two already commented, that is, initial convergence is as fast as when $\theta = 0$ but it lasts shorter.

When there are just two filter weights, it is possible to sketch a geometrical interpretation of the results. Fig. 1(b) shows the LMS mean weight-value tracks together with the elliptic error-surface contours for the three learning curves in Fig. 1(a). The components of the weight-error vector, $\mathbf{w}_o - \mathbf{w}$, are used in the plot to facilitate the comparison, since this way the minimum value of the error surface is located at point $(0, 0)$ of the plot for the three cases.

From Fig. 1(b), we see that the weight-value track for $\theta = 0$ departs from a point close to the optimum, that is, close to the ellipse minor axis, and, consequently, with high excitation of the fast mode, what justifies the fast convergence seen in Fig. 1(a). On the contrary, when $\theta = \pi/2$ the track departs from a point close to the ellipse major axis, exciting greatly the slow mode.

The results presented in this example, with $\omega = \pi/5$, can be transferred to any frequency, with the consideration that the greater the eigenvalue spread, the greater the difference in convergence speed when comparing different phase shifts θ. Also, it is remarkable that similar results to the ones presented here are obtained with a random initial weight vector for each run of the algorithm, as long as the mean of the random distribution is $E(\mathbf{w}[0]) = \mathbf{0}$, what seems sensible when prior knowledge of the environment is not available.

IV. Stochastic Case

In the previous section, the simple case of real sinusoidal signals with a two-tap filter has been analyzed. In this section we consider the more general case of wide-sense stationary stochastic processes for the input and the desired response, with any filter length. As in the previous section, we raise two extreme desired responses, from the viewpoint of similarity with the input, in order to draw conclusions concerning the influence of the desired response on the initial convergence of the algorithms.

A. High Similarity Between Input and Desired Response

Let the desired response be simply a scaled and delayed version of the input, with possibly an additive uncorrelated noise

$$d[n] = C^* x[n - m] + u[n] \tag{33}$$

(the complex conjugate in the scaling factor C^* is used only for notational convenience). In this case, the cross-correlation function $r_{xd}[n] = E(x[k]d^*[k-n])$ is also a scaled and time-shifted version of the autocorrelation function $r_{xx}[n] = E(x[k]x^*[k-n])$

$$r_{xd}[n] = Cr_{xx}[n+m]. \tag{34}$$

In the frequency domain, this condition means that the power spectral densities of both signals, without uncorrelated components, are directly proportional

$$S_{dd}(e^{j\omega}) = |C|^2 S_{xx}(e^{j\omega}). \tag{35}$$

The similarity measure is

$$S = \sum_{k=0}^{N-1} |\mathrm{Re}(r_{\bar{x}\bar{x}}[-k+m])|. \tag{36}$$

When the delay m is inside the filter span, that is, when $0 \le m < N$ (N being the length of the adaptive filter), the following condition is met

$$S \ge 1 \tag{37}$$

since $r_{\bar{x}\bar{x}}[0] = 1$ is one of the terms in the sum in (36). The equality in the previous condition happens only when the input (and also its correlated component in the desired response) is white.

Also, when $0 \le m < N$, the cross-correlation vector is proportional to one of the columns, $\mathbf{r}_m$, of the input correlation matrix, that is

$$\mathbf{p} = C\mathbf{r}_m = C\sum_{j=0}^{N-1} \lambda_j q_{j,m}^* \mathbf{q}_j \tag{38}$$

where $q_{j,m}$ is the mth component of the jth eigenvector. Using the orthogonality property of the eigenvectors, the projection of the cross-correlation vector on each of the eigenvectors is

$$\mathbf{q}_k^H \mathbf{p} = \mathbf{q}_k^H C \sum_{j=0}^{N-1} \lambda_j q_{j,m}^* \mathbf{q}_j = \lambda_k q_{k,m}^* C. \tag{39}$$

Making use of (39) in (6) and (10) we find the initial amplitudes of SD and LMS modes for this case

$$A_{k,\mathrm{SD}} = \lambda_k |q_{k,m}^* C - \mathbf{q}_k^H \mathbf{w}[0]|^2 \tag{40}$$

$$A_{k,\mathrm{LMS}} = \lambda_k \left(|q_{k,m}^* C - \mathbf{q}_k^H \mathbf{w}[0]|^2 - \frac{\mu J_{\min}}{2} \right). \tag{41}$$

As with the in-phase sinusoidal signals in the previous section, the excitation of each mode in this high-similarity case is directly proportional to its own eigenvalue. Therefore, fast initial convergence can be expected since faster modes are initially much more excited than slower ones.

Observe that the term $|q_{k,m}^* C - \mathbf{q}_k^H \mathbf{w}[0]|^2$ is now the one introducing some randomness in the modes excitation. So, this randomness exists now even with the null initial weight vector. Furthermore, the excitation uncertainty is now greater than in the sinusoidal case since the eigenvectors $\mathbf{q}_k$ are not generally known. Of course, it could be argued once more that, depending on the initial weight vector and as a consequence of the randomness in the excitation, convergence can also be slow even in this high similarity case. However, this unwanted possibility is remote and hard to find without exact knowledge of the eigenvalues and eigenvectors.

B. Low Similarity Between Input and Desired Response

In order to make possible some MSE reduction, it is mandatory to have some cross-correlation between input and desired response. Therefore, we do not consider here the trivial case of null cross-correlation, $r_{xd}[n] = 0$. Instead, we consider the situation where the cross-correlation function is just an impulse

$$r_{xd}[n] = C\delta[n+m]. \tag{42}$$

This low similarity condition can also be interpreted in the spectral domain, in terms of the power spectral density functions of both signals. In this case, their spectra will be inversely proportional

$$S_{dd}(e^{j\omega}) = \frac{|C|^2}{S_{xx}(e^{j\omega})} \tag{43}$$

if we consider only the correlated components of both signals (see Appendix II). Hence, now a low-pass input implies a high-pass desired response, and vice versa.

The delay m must be again inside the filter span, $0 \le m < N$, for the MSE reduction to be feasible. The condition met now by the similarity measure is

$$S = \left| \mathrm{Re}\left(\frac{C}{\sqrt{P_x P_d}} \right) \right| \le 1 \tag{44}$$

with the equality happening, once again, only when the correlated components of the input and the desired response are both white. The demonstration of the inequality in (44) is linked to the demonstration of the condition $|C|^2 \le P_x P_d$, that can be easily obtained in the frequency domain with the aid of (43) and Schwarz's inequality.

With $0 \le m < N$, all of the cross-correlation vector components are zero but for the mth one

$$\mathbf{p} = [0, \ldots, 0, C, 0, \ldots, 0]^T. \tag{45}$$

For this cross-correlation vector, we have the following projections on the eigenvectors

$$\mathbf{q}_k^H \mathbf{p} = q_{k,m}^* C. \tag{46}$$

Again, substituting (46) in (6) and (10), we get the initial amplitude of each mode in this case

$$A_{k,\mathrm{SD}} = \lambda_k \left| \frac{q_{k,m}^* C}{\lambda_k} - \mathbf{q}_k^H \mathbf{w}[0] \right|^2 \tag{47}$$

$$A_{k,\mathrm{LMS}} = \lambda_k \left(\left| \frac{q_{k,m}^* C}{\lambda_k} - \mathbf{q}_k^H \mathbf{w}[0] \right|^2 - \frac{\mu J_{\min}}{2} \right). \tag{48}$$

Inspecting the limiting cases for the excitations (47) and (48), we reach now the same conclusions as with sinusoidal signals in phase quadrature. Thus, the excitation of modes with very large eigenvalue will be, at best, directly proportional to its own eigenvalue whereas the excitation of modes with very small eigenvalue will be inversely proportional to it. Therefore, although fast and slow modes are greatly excited, convergence is now dominated by the slow modes from the beginning.

C. Example 1

To illustrate the preceding high versus low similarity analysis, let us consider the next example. The input is generated by filtering a unit-variance white Gaussian signal with the following second-order low-pass filter

$$H_x(z) = K \frac{r^2 + r\sqrt{3}z^{-1} + z^{-2}}{1 - r\sqrt{3}z^{-1} + r^2 z^{-2}}. \tag{49}$$

The gain factor, K, is such that the power of the input is $P_x = 1$.

The two desired responses from the preceding analysis are also generated by filtering the same unit-variance white Gaussian signal used to produce the input. The high-similarity desired response is generated with a delayed version of the low-pass input filter

$$H_{d1}(z) = z^{-16} H_x(z). \tag{50}$$

On the other hand, the low-similarity desired response is generated also by filtering the Gaussian signal but with this other high-pass filter

$$H_{d2}(z) = \frac{z^{-16}}{H_x^*(1/z^*)}. \tag{51}$$

Both filters, (50) and (51), are causal and stable whenever $r < 1$. Taking into account that $S_{xx}(e^{j\omega}) = |H_x(e^{j\omega})|^2$, it is easy to check that the power spectral density functions of the two desired responses $S_{d1d1}(e^{j\omega})$ and $S_{d2d2}(e^{j\omega})$, are directly and inversely proportional, respectively, to the power spectral density of the input, $S_{xx}(e^{j\omega})$, in accordance with (35) and (43) in our analysis.

Additionally, both desired responses contain also an additive uncorrelated white noise signal whose power level is 25 dB below that of the correlated component. Note that this uncorrelated component imposes a limit in the achievable MSE, $J_{\min} \geq -25$ dB.

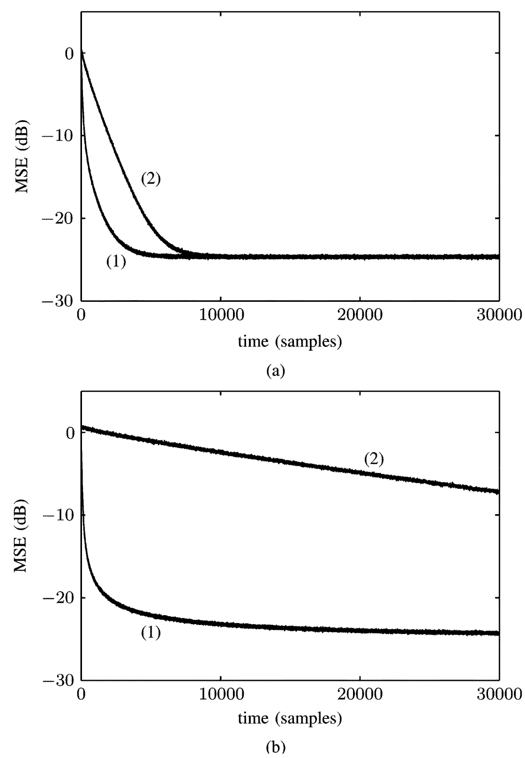

Fig. 2. LMS ensemble average learning curves with two different desired responses, (1) high similarity case, and (2) low similarity case, for two possible eigenvalue spreads. (a) $\chi(\mathbf{R}) = 31.39 (r = 0.25)$. (b) $\chi(\mathbf{R}) = 904.58 (r = 0.50)$.

Varying the distance r from the poles in $H_x(z)$ to the origin, we obtain different inputs with different eigenvalue spreads, $\chi(\mathbf{R}) = \lambda_{\max}/\lambda_{\min}$. Fig. 2(a) and (b) shows the learning curves obtained by averaging 5000 runs of the LMS algorithms for two different inputs, the ones obtained when $r = 0.25$ and $r = 0.50$, respectively. The initial weight vector is $\mathbf{w}[0] = \mathbf{0}$. The length of the adaptive filter and the adaptation step size used are $N = 33$ and $\mu = 0.15/N$. Thus, differences on convergence rate with the two desired responses are only due to the different excitation of natural modes. The similarity measure in the high similarity case is $S = 2.97$ when $r = 0.25$ and $S = 4.40$ when $r = 0.50$. In the low similarity case the similarities are $S = 0.53$ when $r = 0.25$ and $S = 0.14$ when $r = 0.50$.

It is clear that the high-similarity case converges significantly faster than the low-similarity one for both inputs, in accordance with our previous discussion. In addition, with greater eigenvalue spread ($r = 0.50$), the difference on convergence rate in the two situations becomes more evident, since the fast modes are faster and the slow modes are slower.

D. Example 2

Our previous analysis suggests that initial convergence is fast when there is great similarity between input and desired response and slow with little similarity. In order to check this point, in the next example we consider four different desired responses, none of them corresponding to the two extreme cases already analyzed. Each of the desired responses has different spectral similarity with the input.

The input is a unit-power autoregressive-moving-average (ARMA) process generated by filtering a unit-variance white Gaussian signal with the following low-pass filter:

$$H_x(z) = 0.4576 \frac{1 + 0.7391 z^{-1} + 0.1600 z^{-2}}{1 - 0.7391 z^{-1} + 0.1600 z^{-2}}. \tag{52}$$

The correlated component in the four desired responses is obtained by filtering this same unit-variance white Gaussian signal with different bandpass filters

$$H_{d_i}(z) = K_i \frac{(0.16 + 0.8\cos(\phi_i) z^{-1} + z^{-2})(1 - z^{-2})}{1 - 0.8\cos(\phi_i) z^{-1} + 0.16 z^{-2}} z^{-12}. \tag{53}$$

The gain factor K_i is selected in the four cases in order to get a unit-power desired response. In (53), the angles of the poles are $\pm\phi_i$, while the angles of two of the zeros are their supplementaries, $\pi \pm \phi_i$. The four desired responses are obtained with $\phi_i = \pi/8, 3\pi/8, 5\pi/8$, and $7\pi/8$, yielding the following similarities $S_i = 3.55, 2.58, 1.14$ and 0.66, respectively. There is also an uncorrelated white noise in the desired responses, whose power level is 25 dB below that of the correlated component. Again, this uncorrelated component imposes the limit, $J_{\min} \geq -25$ dB, in the achievable MSE.

The power spectral density functions of the input and the four desired responses are plotted in Fig. 3(a), in order to compare the spectral similarity between them. Fig. 3(b) shows the learning curves obtained in the four cases by averaging of 5000 runs of the LMS algorithm. The filter length is $N = 25$, which yields an eigenvalue spread $\chi(\mathbf{R}) = 394.39$. The step size is $\mu = 0.15/N$ and the initial weight vector is also $\mathbf{w}[0] = \mathbf{0}$.

The learning curves from Fig. 3(b) boost the main idea behind our analysis: the greater the spectral similarity between input and desired response, the faster convergence. That is, the fast modes are more excited and the slow modes are less excited in the case that the desired response is more similar to the input.

Therefore, according to our analysis and the empirical results, one criterion to be taken into account when selecting the desired response, whenever this is possible, should be obtaining the greatest spectral similarity between this signal and the input, in order to ensure a fast convergence for the adaptive algorithm. Also, when it is not possible to select the desired response, we can say that prior knowledge of the environment is more necessary, in order to select an adequate initial weight vector, when the similarity between desired response and input is low.

V. Adaptive Prewhitening of the Input

Prewhitening of the input is a well-known method for improving LMS convergence [14]–[17]: the input to the LMS filter is preconditioned, in order to obtain faster convergence, by filtering it with an inverse linear stochastic model of itself, that is, a whitener. Fig. 4 shows a block diagram of this alternative adaptive system where the whitener, implemented as a prediction error filter, is also made adaptive.

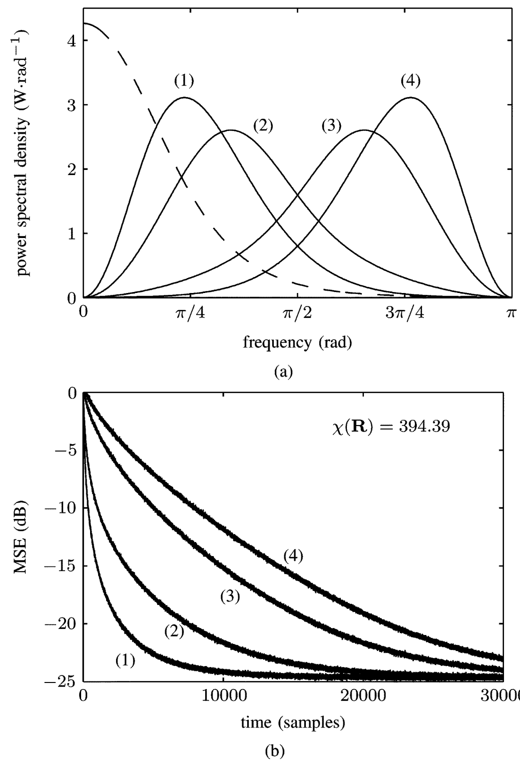

Fig. 3. LMS convergence with four different desired responses. (a) Power spectral density functions: input (dashed line) and four different desired responses (solid lines). (b) Ensemble average learning curves.

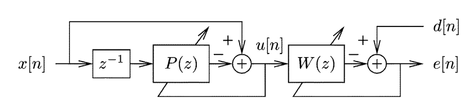

Fig. 4. System with adaptive prewhitening of the input.

In order to check the performance of this alternative adaptive system, we consider again the same data from the example in Subsection IV.D for the last of the four possible desired responses [$\phi_i = 7\pi/8$ in (53)]. That is, we have the same input, the filter length is $N = 25$, the step size is $\mu = 0.15/N$ and the initial weight vector is $\mathbf{w}[0] = \mathbf{0}$. On the other hand, for the prediction filter $P(z)$, the filter length is $N_p = 10$, the step size is $\mu_p = 0.01/N_p$ and the initial weight vector is $\mathbf{p}[0] = \mathbf{0}$. Fig. 5 shows the learning curves obtained by averaging of 5000 runs for the original LMS system, without the prewhitening, and for the alternative system with adaptive prewhitening of the input. The convergence improvement is evident in this example.

We have also considered the case of the first desired response in the example of Section IV-D [$\phi_i = \pi/8$ in (53)]. In this case, the two learning curves, with and without input prewhitening, overlap. So, the prewhitening stage is unnecessary, due to the fact that convergence was already fast without it [curve (1) in Fig. 3(b)].

To explain the improvement in convergence speed with an adaptive prewhitening of the input, we must consider that there are two adaptive filters in the system from Fig. 4. From the viewpoint of the main adaptive filter, $W(z)$, the new input (that is, the output of the whitener) is whiter than the previous one and

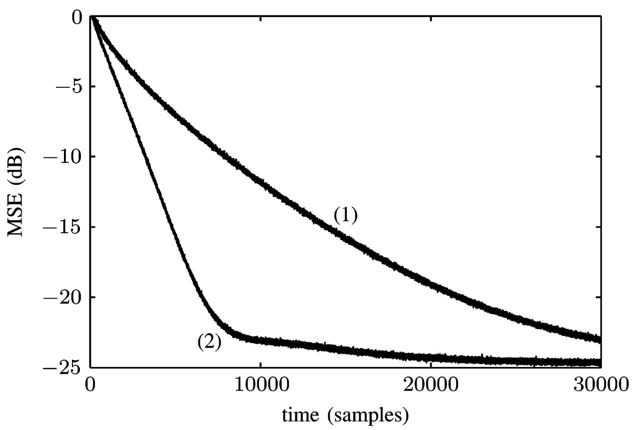

Fig. 5. LMS ensemble average learning curves for (1) the original system and (2) the system with adaptive prewhitening of the input.

hence there is an eigenvalue spread reduction. Thus, this reduction of the directionality of convergence justifies the improvement. However, a relatively fast whitener convergence is also needed for this argument to be true.

From the standpoint of the prediction filter, $P(z)$, its input is the same as that of the original system, but for a unit-sample delay. So, thinking only of the input, it seems that we have just transferred the convergence problems (i.e., eigenvalue spread) of the original system to the whitening stage. Nevertheless, according to our previous analysis, the prediction filter will generally exhibit fast convergence, due to the evident high similarity between its input, $x[n-1]$, and desired response, $x[n]$. Therefore, the superior performance of the system with the input prewhitening is also a consequence that can be extracted from our analysis.

Besides the adaptive prewhitening of the input and with the conclusions of our analysis in mind, one could also think of a novel adaptive system where the desired response, instead of the input, is preconditioned in order to get faster convergence. This can be done, for instance, by filtering $d[n]$ with a linear stochastic model of the input. This way, an improvement in convergence could be expected in relation to the basic adaptive system, due to the higher similarity between the input and the new desired response. However, from our tests we can say that when compared to the system that simply prewhitens the input, the convergence of this new system is clearly worse. In other words, the benefit on convergence speed of a reduction on the eigenvalue spread (i.e., in the directionality of convergence) is clearly greater than that of providing a more adequate excitation for the pre-existing modes. Therefore, in order to get faster convergence the input is the signal that should be preconditioned, by whitening it, and not the desired response.

VI. Conclusion

The influence of the desired response on the convergence of SD and LMS algorithms has been addressed in this paper. Starting from previous well-established convergence analyses, and considering two extreme cases, first for sinusoidal signals and later for wide-sense stationary stochastic processes, we have shown that the greater the similarity between input and desired response, the faster convergence. Thus, in the same way that very high coherence between input and desired response is required in order to get high cancellation levels, we can conclude that high similarity is a must in order to get the fastest convergence. This conclusion, that can seem quite logical or intuitive, is supported by experimental results and explained by our analysis. To the best of our knowledge, this was one of the few remaining unanalyzed aspects of the LMS convergence behavior.

From a practical perspective, we can say that, if it is possible to choose the input or the desired response in any sense, it would be wise to look for the greatest similarity between both signals, in order to ensure fast convergence. On the other hand, prewhitening of the input is a simple and effective way of speeding up convergence for stochastic processes when there is low similarity between desired response and input.

Appendix I
Derivation of (23)

Since the input in (20) is sinusoidal, so is its correlation function

$$r_{xx}[m] = \frac{C_x^2}{2}\cos(\omega m). \tag{54}$$

When the number of filter taps is $N = 2$, the correlation matrix of the input vector is

$$\mathbf{R} = \begin{pmatrix} r_{xx}[0] & r_{xx}[1] \\ r_{xx}[-1] & r_{xx}[0] \end{pmatrix} = \frac{C_x^2}{2}\begin{pmatrix} 1 & \cos\omega \\ \cos\omega & 1 \end{pmatrix} \tag{55}$$

which can be decomposed as $\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H$, where

$$\mathbf{Q} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{56}$$

and

$$\mathbf{\Lambda} = \frac{C_x^2}{2}\begin{pmatrix} 1+\cos\omega & 0 \\ 0 & 1-\cos\omega \end{pmatrix} \tag{57}$$

are the eigenvector and eigenvalue matrices, respectively.

The cross-correlation function between input and desired response is also sinusoidal, when the desired response is given by (21)

$$r_{xd}[m] = \frac{C_d C_x}{2}\cos(\omega m - \theta) \tag{58}$$

with θ as defined in (22). So, the cross-correlation vector is

$$\mathbf{p} = \begin{pmatrix} r_{xd}[0] \\ r_{xd}[-1] \end{pmatrix} = \frac{C_d C_x}{2}\begin{pmatrix} \cos\theta \\ \cos(\omega+\theta) \end{pmatrix}. \tag{59}$$

After some manipulation, we get

$$\mathbf{\Lambda}^{-1}\mathbf{Q}^H\mathbf{p} = \frac{C_d}{\sqrt{2}C_x}\begin{pmatrix} \cos\theta - \sin\theta\frac{\sin\omega}{1+\cos\omega} \\ \cos\theta + \sin\theta\frac{\sin\omega}{1-\cos\omega} \end{pmatrix}. \tag{60}$$

Using the previous result and recalling (12), it is straightforward to obtain (23).

Appendix II
Derivation of (43)

Since we consider only the correlated components in the wide-sense stationary processes $x[n]$ and $d[n]$, we can think that $d[n]$ could be obtained from $x[n]$ by filtering it with some unknown filter, with impulse response $h[n]$. In this case, the following relations hold:

$$r_{xd}[m] = h^*[-m] * r_{xx}[m] \tag{61}$$
$$r_{dd}[m] = h[m] * h^*[-m] * r_{xx}[m]. \tag{62}$$

Considering that the z-transform of $h^*[-m]$ is $H^*(1/z^*)$, the previous relations in the transformed domain are

$$S_{xd}(z) = H^*(1/z^*)S_{xx}(z) \tag{63}$$
$$S_{dd}(z) = H(z)H^*(1/z^*)S_{xx}(z). \tag{64}$$

From (42), we also know that

$$S_{xd}(z) = Cz^m. \tag{65}$$

Equating (63) and (65), we can obtain now the expression for the filter relating $x[n]$ and $d[n]$ in the low similarity case,

$$H(z) = \frac{C^* z^{-m}}{S_{xx}(z)}. \tag{66}$$

The relation $S_{xx}(z) = S_{xx}^*(1/z^*)$, that follows from the hermitian symmetry property of the autocorrelation function $r_{xx}[m] = r_{xx}^*[-m]$, has been used in the derivation of (66). Making use of (66) and this same relation in (64), it is straightforward to obtain (43).

Acknowledgment

The authors would like to thank Prof. S. Haykin and Dr. S. Olmos for their critical review of an early version of the paper and their helpful comments and suggestions, and the anonymous reviewers for their constructive comments, which undoubtedly served to improve this paper.

References

[1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
[2] S. Haykin, *Adaptive Filter Theory*. Upper Saddle River, NJ: Prentice-Hall, 2002.
[3] H. J. Butterweck, "A steady-state analysis of the LMS adaptive algorithm without use of the independence assumption," in *Proc. ICASSP*, 1995, pp. 1404–1407.
[4] H. J. Butterweck, "A wave theory of long adaptive filters," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 48, no. 6, pp. 739–747, Jun. 2001.
[5] L. Vicente and E. Masgrau, "Cross correlation p-vector influence on LMS convergence," in *Proc. EUSIPCO*, 2000.
[6] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.
[7] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, Jr., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 64, no. 8, pp. 1151–1162, Aug. 1976.
[8] J. E. Mazo, "On the independence theory of equalizer convergence," *Bell Syst. Tech. J.*, vol. 58, no. 5, pp. 963–993, May 1979.
[9] W. A. Gardner, "Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis and critique," *Signal Process.*, vol. 6, no. 2, pp. 113–133, Apr. 1984.
[10] J. R. Glover, Jr., "Adaptive noise canceling applied to sinusoidal interferences," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-25, no. 6, pp. 484–491, Dec. 1977.
[11] M. J. Shensa, "Non-Wiener solutions of the adaptive noise canceller with a noisy reference," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, no. 4, pp. 468–472, Aug. 1980.
[12] S. Olmos and P. Laguna, "Steady-state MSE convergence of LMS adaptive filters with deterministic reference inputs with applications to biomedical signals," *IEEE Trans. Signal Process.*, vol. 48, no. 8, pp. 2229–2241, Aug. 2000.
[13] L. Vicente and E. Masgrau, "Novel FxLMS convergence condition with deterministic reference," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3768–3774, Oct. 2006.
[14] S. Yamamoto, S. Kitayama, J. Tamura, and H. Ishigami, "An adaptive echo canceller with linear predictor," *Trans. IECE Jpn.*, vol. E 62, no. 12, pp. 851–857, Dec. 1979.
[15] P. C. W. Sommen and C. J. van Valburg, "Efficient realisation of adaptive filter using an orthogonal projection method," in *Proc. ICASSP*, 1989, pp. 940–943.
[16] M. Mboup, M. Bonnet, and N. Bershad, "LMS coupled adaptive prediction and system identification: A statistical model and transient mean analysis," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2607–2615, Oct. 1994.
[17] C. Breining, P. Dreiseitel, E. Hänsler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, "Acoustic echo control: An application of very-high-order adaptive filters," *IEEE Signal Process. Mag.*, vol. 16, no. 5, pp. 42–69, Jul. 1999.

Luis Vicente (M'05) was born in Zaragoza, Spain, in 1972. He received the M.Eng. and Ph.D. degrees in telecommunication engineering from the Engineering Faculty, University of Zaragoza, Zaragoza, Spain, in 1996 and 2005, respectively.

He is an Assistant Professor of Signal Processing and Communications in the Department of Electronics Engineering and Communications at the Engineering Faculty, and Researcher of the Aragon Institute for Engineering Research (I3A), both of the University of Zaragoza. His current research interests are in the field of adaptive signal processing, in particular, applied to active noise and vibration control, and vehicular technologies.

Enrique Masgrau (M'84) received the M.S. and Ph.D. degrees in electrical engineering from the Polytechnic University of Catalonia (UPC), Barcelona, Spain, in 1978 and 1983, respectively.

He was an Assistant Professor (1978 to 1992) at UPC. He joined the University of Zaragoza, Zaragoza, Spain, in 1992, as a Full Professor with the Department of Electronic Engineering and Communications. He is also a Member of the Aragon Institute of Engineering Research (I3A) where he is Manager of the Communications Technologies Group. His research interests include speech processing, acoustic noise cancellation, MIMO communication techniques, and ICT applications in automotive ("telematics"). In these areas, he has published over 100 technical papers in various international journals and conferences. He has also been serving as Reviewer of several international conferences and journals. He holds three international patents.