

A Time-Domain Feedback Analysis of Filtered-Error Adaptive Gradient Algorithms

Markus Rupp and Ali H. Sayed, *Member, IEEE*

Abstract—This paper provides a time-domain feedback analysis of gradient-based adaptive schemes. A key emphasis is on the robustness performance of the adaptive filters in the presence of disturbances and modeling uncertainties (along the lines of H^∞ -theory and robust filtering). The analysis is carried out in a purely deterministic framework and assumes no prior statistical information or independence conditions.

It is shown that an intrinsic feedback structure can be associated with the varied adaptive schemes. The feedback structure is motivated via energy arguments and is shown to consist of two major blocks: a time-variant lossless (i.e., energy preserving) feedforward path and a time-variant feedback path. The configuration is further shown to lend itself to analysis via a so-called small gain theorem, thus leading to stability and robustness conditions that require the contractivity of certain operators. Choices for the step-size parameter in order to guarantee faster rates of convergence are also derived, and simulation results are included to demonstrate the theoretical findings.

In addition, the time-domain analysis provided in this paper is shown to extend the so-called transfer function approach to a general time-variant scenario without any approximations.

I. INTRODUCTION

THE last decade has seen increasing interest in the fields of adaptive filtering, robust estimation, and robust control. Adaptive and robust filters (which are also known as H^∞ -filters) are increasingly being considered in numerous applications to help cope with time variations of system parameters and to compensate for the lack of *a priori* knowledge of the statistical properties of the input data and/or exogenous signals. In the control community, and especially over the past several years, considerable research has been conducted on robust control and filtering [1]–[5]. A major motivation for this work has been the need to design stabilizing controllers for uncertain plants, namely, plants that operate in the presence of disturbances and modeling uncertainties. The ideas developed in these contexts have been recently encountering useful counterparts in signal processing and communication problems.

Manuscript received January 21, 1995; revised November 16, 1995. This material was based on work supported by the National Science Foundation under Award MIP-9409319. The work of M. Rupp was also supported by a scholarship from DAAD (German Academic Exchange Service) as well as the scientific division of NATO. The associate editor coordinating the review of this paper and approving it for publication was Dr. Stephen M. McLaughlin.

M. Rupp was with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA. He is currently with the Wireless Technology Research Department, Lucent Technologies, Holmdel, NJ 07733-0400 USA.

A. H. Sayed is with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106-9560 USA.

Publisher Item Identifier S 1053-587X(96)03970-0.

In this paper, we pursue these ideas for the important class of gradient-based adaptive filters and study their robustness (or l_2 -stability) properties along the lines of H^∞ theory. In particular, one of the contributions of this work is to show how to choose the step-size parameter of an adaptive filter in order to result in a robust performance and in order to improve the convergence speed.

A. Robust Adaptive Filters

Intuitively, a robust filter is one for which the estimation errors are consistent with the disturbances in the sense that “small” disturbances would lead to “small” estimation errors. This is not generally true for any adaptive filter: The estimation errors may still be relatively large even in the presence of small disturbances.

The robustness issue is addressed in this paper in a purely deterministic framework and without assuming prior knowledge of noise and signal statistics or independence conditions. This is especially useful in situations where prior statistical information is missing since a robust design would guarantee a desired level of robustness independent of the statistical nature of the noise and signals. In loose terms, robustness implies that the ratio of estimation error energy to disturbance energy is guaranteed to be bounded by a positive constant, say, the constant one

$$\frac{\text{estimation error energy}}{\text{disturbance energy}} \leq 1. \quad (1)$$

Here, the term “disturbance energy” refers to the combined energies of measurement noise, modeling uncertainties, error in the initial weight guess, etc. From a practical point of view, a relation of the form (1) is desirable since it guarantees that the resulting estimation error energy will be at most equal to the disturbance energy no matter what the nature and the statistics of the disturbances are. In this sense, the algorithm will not unnecessarily magnify the disturbance energy, and consequently, small estimation errors will result when small disturbances occur.

In this work, we show that gradient adaptive schemes can be designed to be robust with respect to disturbances by imposing suitable conditions on the step-size parameter. This may be contrasted with results in a stochastic setting where stability (or convergence) statements are often given in the mean and mean-square sense. In such settings, even for the simple LMS algorithm, a constant step-size μ that is bounded by twice

the inverse of the maximal eigenvalue of the autocorrelation matrix can still lead to blow up in a practical experiment [6].

B. A Time-Domain Feedback Analysis

To address the robustness and convergence issues, this paper develops a time-domain approach that proves to be useful in both the analysis and design of robust estimators. It highlights and exploits an intrinsic feedback structure that can be associated with the gradient adaptive schemes.

Although the feedback nature of adaptive filters has been exploited in earlier places in the literature [7]–[9], the feedback configuration studied in this paper has a different emphasis. It does not only refer to the fact that the update equations can be put into a feedback form (as explained in [10]) but is instead motivated by energy arguments that also explicitly take into consideration both the effect of the measurement noise and the effect of the uncertainty in the initial guess for the weight vector. These extensions are incorporated into the feedback arguments of this paper because the derivation here is specifically interested in a study of the robustness properties of the adaptive schemes.

The feedback interconnection exhibits several features: Its feedforward mapping is lossless (i.e., energy preserving) while its feedback mapping is either memoryless or dynamic. Moreover, both mappings are time variant, and their interconnection lends itself to stability analysis via a so-called small gain theorem, which is a very useful tool in system theory [11], [12].

An interesting fallout of the time-domain analysis of this paper is that it can be regarded as an extension of the transfer function approach that is often used in the analysis of gradient-based recursions (e.g., [13], [14]). The time-domain analysis, however, is shown to avoid the restrictions and limitations that are characteristic of the transfer-function domain.

C. Notation

We use small boldface letters to denote vectors and capital boldface letters to denote matrices. In addition, the symbol “*” denotes Hermitian conjugation (complex conjugation for scalars). The symbol **I** denotes the identity matrix of appropriate dimensions, and the boldface letter **0** denotes either a zero vector or a zero matrix. Finally, the notation $\|\mathbf{x}\|$ denotes the Euclidean norm of a vector. All vectors are column vectors except for the input data vector denoted by \mathbf{u}_i , which is taken to be a row vector.

II. THE LEAST-MEAN-SQUARE ALGORITHM

One of the most widely used adaptive algorithms is the least-mean-squares (LMS) algorithm [15]. It starts with an initial guess \mathbf{w}_{-1} for an unknown $M \times 1$ weight vector \mathbf{w} and updates it as follows:

$$\begin{aligned} \mathbf{w}_i &= \mathbf{w}_{i-1} + \mu(i) \mathbf{u}_i^* [d(i) - \mathbf{u}_i \mathbf{w}_{i-1}] \\ &= \mathbf{w}_{i-1} + \mu(i) \mathbf{u}_i^* \tilde{e}_a(i) \end{aligned} \quad (2)$$

where the $\{\mathbf{u}_i\}$ are given nonzero row vectors, the $\{d(i)\}$ are given noisy (or disturbed) measurements of the terms $\{\mathbf{u}_i \mathbf{w}\}$, viz., $d(i) = \mathbf{u}_i \mathbf{w} + v(i)$, and \mathbf{w}_i is the weight estimate at iteration i . The factor $\mu(i)$ is the step-size parameter (which is allowed to be time variant), and the quantity $v(i)$ may account for both measurement noise and modeling errors. The difference $[d(i) - \mathbf{u}_i \mathbf{w}_{i-1}]$ is denoted by $\tilde{e}_a(i)$.

The following error measures are also useful for our later analysis: $\tilde{\mathbf{w}}_i$ denotes the difference between the true weight \mathbf{w} and its estimate \mathbf{w}_i , $\tilde{\mathbf{w}}_i = \mathbf{w} - \mathbf{w}_i$, $e_a(i)$ denotes the *a priori* estimation error, $e_a(i) = \mathbf{u}_i \tilde{\mathbf{w}}_{i-1}$, and $e_p(i)$ denotes the *a posteriori* estimation error $e_p(i) = \mathbf{u}_i \tilde{\mathbf{w}}_i$.

It is straightforward to verify that $\tilde{e}_a(i) = e_a(i) + v(i)$. Moreover, it follows from (2) that $\tilde{\mathbf{w}}_i$ satisfies

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1} - \mu(i) \mathbf{u}_i^* \tilde{e}_a(i). \quad (3)$$

If we further multiply (3) by \mathbf{u}_i from the left, we obtain the following relation among $\{e_p(i), e_a(i), v(i)\}$:

$$e_p(i) = [1 - \mu(i) \|\mathbf{u}_i\|^2] e_a(i) - \mu(i) \|\mathbf{u}_i\|^2 v(i). \quad (4)$$

A. Transfer Function Description of the LMS Algorithm

Before proceeding to the time-domain analysis of this paper, we first review a well-known approach to the analysis of LMS-type recursions that employs the concept of transfer functions [13], [14].

In this method, the input vector \mathbf{u}_i is assumed to have a shift structure, say, $\mathbf{u}_i = [u(i), \dots, u(i - M + 1)]$, where the individual entries are further assumed to arise from a sinusoidal excitation $u(i) = C \cos(\Omega i)$. Assuming a constant step-size μ and neglecting the initial condition $\tilde{\mathbf{w}}_{-1}$, the transfer function from the disturbance $v(\cdot)$ to the *a priori* estimation error $e_a(\cdot)$ can be shown to be approximately (see Appendix A, where $E_a(z)$ is the z -transform of $e_a(\cdot)$)

$$\frac{E_a(z)}{V(z)} \cong \frac{\frac{\mu}{\bar{\mu}} [1 - z \cos(\Omega)]}{z^2 - 2z \cos(\Omega) \left(1 - \frac{\mu}{2\bar{\mu}}\right) + \left(1 - \frac{\mu}{\bar{\mu}}\right)} \quad (5)$$

where we have defined $\bar{\mu} = \frac{2}{C^2 M}$. Although easily available in the literature, a derivation of the above result is included in Appendix A in order to highlight some of the restrictions and approximations that are needed to establish (5). These approximations will be avoided when the time-domain analysis is introduced in later sections. For now, however, we stress the fact that an interesting feedback structure is implied by (5). To clarify this, we define $\bar{V}(z)$,

$$\bar{V}(z) = \frac{\mu}{\bar{\mu}} V(z) - \left(1 - \frac{\mu}{\bar{\mu}}\right) E_a(z) \quad (6)$$

and use (5) to conclude that

$$\frac{E_a(z)}{\bar{V}(z)} = \frac{z^{-1} - \cos(\Omega)}{z - \cos(\Omega)}. \quad (7)$$

That is, the transfer function from $\bar{v}(\cdot)$ to $e_a(\cdot)$ is allpass. Consequently, the transfer function (5), from $v(i)$ to $e_a(\cdot)$, can be expressed as a feedback structure with an allpass filter in the forward path and a constant gain in the feedback loop.

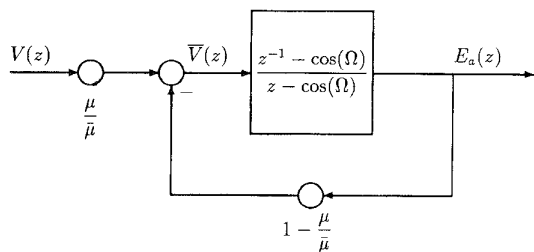


Fig. 1. Transfer function description for LMS: An allpass mapping with gain feedback.

This is depicted in Fig. 1. The feedback gain is $(1 - \mu/\bar{\mu})$, which is thus equal to zero if we choose $\mu = \bar{\mu}$. This is known to be the choice that results in the highest convergence speed. Further clarifications will be provided in later sections in the time domain.

B. Comments on the Transfer-Function Description

The transfer function derivation has some limitations that hinders its applicability to more general scenarios.

- i) The arguments explicitly assume that the input vectors exhibit shift structure, which restricts the analysis to transversal filter structures.
- ii) The input sequence is often assumed to be of a specific class (usually sinusoidal), but other choices have also been used in [14].
- iii) The effect of initial conditions is ignored. Although this may not be relevant to the steady-state performance of a stable filter, it is nevertheless useful in a robustness analysis of the filter.
- iv) A constant, rather than time-variant, step-size is assumed.
- v) Some nonlinear mixing terms are neglected as explained in Appendix A.

One of the contributions of this paper is to remove these limitations by employing a time-domain argument. We provide an *exact* derivation that circumvents the above restrictions and avoids any approximations. Once this is done, we then show how the results provide conclusions concerning the robustness and convergence behavior of gradient-type algorithms not only of the type (2) but also for general filtered-error variants. This is achieved by employing the energy-based arguments of [16] and [17].

III. A TIME-DOMAIN ANALYSIS

The analysis that follows highlights an important feedback structure that is implied by gradient-type recursions of the form (2). Our purpose in this section (and Section IV) is threefold:

- 1) to show that certain local and global energy relations can be associated with the LMS recursion (2)—these are also known as passivity relations;
- 2) to employ the passivity relations in order to highlight a feedback structure;

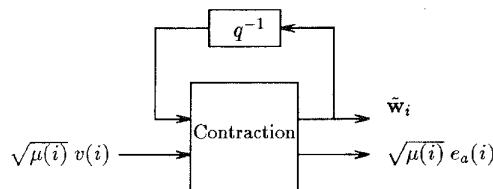


Fig. 2. Local contraction mapping.

- 3) to derive conditions on the step-size parameter in order to guarantee that the feedback structure will behave as a robust filter (cf. H^∞ theory).

Later, in Section V, we shall extend the analysis to filtered-error variants with update equations of the form (24).

A. Local Passivity Relations

To begin with, we compute the squared norm (i.e., energies) of both sides of (3) and use the relation $\tilde{e}_a(i) = e_a(i) + v(i)$ to conclude the following equality:

$$\|\tilde{\mathbf{w}}_i\|^2 + \mu(i)|e_a(i)|^2 + \mu(i)(1 - \mu(i)\|\mathbf{u}_i\|^2)|\tilde{e}_a(i)|^2 = \|\tilde{\mathbf{w}}_{i-1}\|^2 + \mu(i)|v(i)|^2. \quad (8)$$

Equality (8) allows us to study how the energies of the error terms $\{\tilde{\mathbf{w}}_i, e_a(i)\}$ propagate as the algorithm progresses. In particular, if we define $\bar{\mu}(i) = 1/\|\mathbf{u}_i\|^2$, then the following result is immediate from (8).

Lemma 1: Consider the gradient recursion (2). It always holds that

$$\frac{\|\tilde{\mathbf{w}}_i\|^2 + \mu(i)|e_a(i)|^2}{\|\tilde{\mathbf{w}}_{i-1}\|^2 + \mu(i)|v(i)|^2} \begin{cases} \leq 1, & \text{for } 0 < \mu(i) < \bar{\mu}(i) \\ = 1, & \text{for } \mu(i) = \bar{\mu}(i) \\ \geq 1, & \text{for } \mu(i) > \bar{\mu}(i) \end{cases}$$

B. Interpretation

The result of Lemma 1 has an interesting interpretation that was exploited in [16] in order to provide a minimax analysis of gradient recursions of the form (2) (see also [18] for a Krein space formulation and the last section of [19] for an argument based on the Cauchy–Schwarz inequality).

Here, we would like to stress that the first two cases of Lemma 1 establish a local error-energy bound (or passivity relation) that explains the robustness nature of the gradient recursion (2): They state that no matter what the value of the disturbance $v(i)$ is, and no matter how far the estimate \mathbf{w}_{i-1} is from the true vector \mathbf{w} , the sum $\|\tilde{\mathbf{w}}_i\|^2 + \mu(i)|e_a(i)|^2$ will always be at most equal to the sum $\|\tilde{\mathbf{w}}_{i-1}\|^2 + \mu(i)|v(i)|^2$.

This establishes, as shown in Fig. 2, the existence of a contractive map from the signals $\{\tilde{\mathbf{w}}_{i-1}, \sqrt{\mu(i)}v(i)\}$ to the signals $\{\tilde{\mathbf{w}}_i, \sqrt{\mu(i)}e_a(i)\}$.¹ The symbol q^{-1} denotes the unit delay operator.

¹A map that takes x to y , say $y = T[x]$, is said to be contractive if for all x , we have $\|T[x]\| \leq \|x\|$. That is, the output energy does not exceed the input energy.

C. A Global Contraction Mapping

Since the contractivity relation of Fig. 2 holds for each time instant i , it should also hold globally over an interval of time. Indeed, if $\mu(i) \leq \bar{\mu}(i)$ for all i in the interval $0 \leq i \leq N$, then for all such i (cf. Lemma 1)

$$\mu(i)|e_a(i)|^2 \leq \|\tilde{\mathbf{w}}_{i-1}\|^2 - \|\tilde{\mathbf{w}}_i\|^2 + \mu(i)|\bar{v}(i)|^2.$$

Summing over i , we conclude that

$$\|\tilde{\mathbf{w}}_N\|^2 + \sum_{i=0}^N \mu(i)|e_a(i)|^2 \leq \|\tilde{\mathbf{w}}_{-1}\|^2 + \sum_{i=0}^N \mu(i)|v(i)|^2. \quad (9)$$

This relation states that the map from the disturbances

$$\{\tilde{\mathbf{w}}_{-1}, \sqrt{\mu(0)}v(0), \dots, \sqrt{\mu(N)}v(N)\} \quad (10)$$

to the estimation errors

$$\{\sqrt{\mu(0)}e_a(0), \dots, \sqrt{\mu(N)}e_a(N), \tilde{\mathbf{w}}_N\} \quad (11)$$

is a contraction. In other words, assume we stack the entries of (10) into a column vector, the entries of (11) into a second column vector, and let \mathcal{T}_N denote the mapping that maps the first vector (10) to the second vector (11). The entries of this mapping can be determined from the update relation (3) and from the definition of the *a priori* estimation error $e_a(\cdot)$. The specific values of these entries are not of immediate interest here except to say that it can be verified that \mathcal{T}_N turns out to be a block lower triangular operator of the form

$$\begin{bmatrix} \sqrt{\mu(0)}e_a(0) \\ \sqrt{\mu(1)}e_a(1) \\ \vdots \\ \tilde{\mathbf{w}}_N \end{bmatrix} = \underbrace{\begin{bmatrix} x & & & \\ x & x & & \\ \vdots & & \ddots & \\ x & x & x & x \end{bmatrix}}_{\mathcal{T}_N} \begin{bmatrix} \tilde{\mathbf{w}}_{-1} \\ \sqrt{\mu(0)}v(0) \\ \vdots \\ \sqrt{\mu(N)}v(N) \end{bmatrix}.$$

(For example, the first block entry of \mathcal{T}_N , which relates $e_a(0)$ to $\tilde{\mathbf{w}}_{-1}$, can be easily seen to be $\sqrt{\mu(0)}\mathbf{u}_0$). The contractivity of \mathcal{T}_N means that its maximum singular value is at most one, $\bar{\sigma}(\mathcal{T}_N) \leq 1$.

Inequality (9) is a desirable robustness property in the sense that it guarantees that if the disturbance energy is small, then the resulting estimation error energy will be accordingly small.

We may add that other similar local, and global, passivity relations can be established by using *a posteriori* (rather than *a priori*) estimation errors [16]. This is useful in the study of (robust) adaptive IIR filters [20], but we forgo the details here.

IV. THE FEEDBACK STRUCTURE

The bounds in Lemma 1 can be described via an alternative form that will lead us to an interesting feedback structure. The structure will be shown to constitute the proper extension of the transfer function description of Fig. 1 to the general time-variant scenario, and it will further allow us i) to relax the condition on $\mu(i)$ in order to guarantee robustness and ii) to select $\mu(i)$ for faster speeds of convergence.

We have argued above that if the step sizes are chosen such that $\mu(i) \leq \bar{\mu}(i)$, then robustness (or contractivity) is

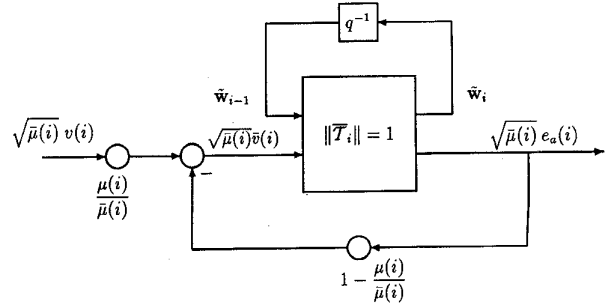


Fig. 3. Time-variant lossless mapping with gain feedback for gradient algorithms.

guaranteed in the sense that the weighted estimation-error energy will never exceed the weighted disturbance energy (cf. (9)). That is, for $\mu(i) \leq \bar{\mu}(i)$, the ratio of the energies of the signals in (11) and (10) will be bounded by one.

The condition on $\mu(i)$ can be relaxed at the expense of guaranteeing energy ratios that will be bounded by some other positive number, say

$$\frac{\text{weighted estimation error energy}}{\text{weighted disturbance energy}} \leq K \quad (12)$$

for some constant K (to be determined). This is still a desirable property since it means that the disturbance energy will be at most scaled by a factor of K .

This issue is addressed here in the context of a feedback analysis. For this purpose, we first rewrite (4) as

$$e_p(i) = \left(1 - \frac{\mu(i)}{\bar{\mu}(i)}\right)e_a(i) - \frac{\mu(i)}{\bar{\mu}(i)}v(i) \quad (13)$$

and use it to conclude that (2) can be rewritten in the equivalent form

$$\begin{aligned} \mathbf{w}_i &= \mathbf{w}_{i-1} + \bar{\mu}(i)\mathbf{u}_i^*[e_a(i) - e_p(i)] \\ &= \mathbf{w}_{i-1} + \bar{\mu}(i)\mathbf{u}_i^*[e_a(i) + \bar{v}(i)] \end{aligned} \quad (14)$$

where we have defined, for convenience, the signal $\bar{v}(i) = -e_p(i)$. Comparing with (2), expression (14) shows that (2) can be rewritten in terms of a new step size $\bar{\mu}(i)$ and a modified “noise” term $\bar{v}(i)$ (recall that in (2), we have $\tilde{e}_a(i) = e_a(i) + v(i)$). Therefore, if we follow arguments similar to those prior to Lemma 1, we readily conclude that for algorithm (2), the following equality holds for all $\{\mu(i), v(i)\}$:

$$\frac{\|\tilde{\mathbf{w}}_i\|^2 + \bar{\mu}(i)|e_a(i)|^2}{\|\tilde{\mathbf{w}}_{i-1}\|^2 + \bar{\mu}(i)|e_p(i)|^2} = 1. \quad (15)$$

This relation establishes a lossless map (which is denoted by $\bar{\mathcal{T}}_i$ from $\{\tilde{\mathbf{w}}_{i-1}, \sqrt{\bar{\mu}(i)}\bar{v}(i)\}$ to $\{\tilde{\mathbf{w}}_i, \sqrt{\bar{\mu}(i)}e_a(i)\}$). Correspondingly, using (13), the map from the original disturbance $\sqrt{\mu(i)}v(i)$ to $\sqrt{\bar{\mu}(i)}e_a(i)$ can be expressed in terms of a feedback structure, as in Fig. 3.

The similarities between Figs. 1 and 3 are clear. Both have a lossless feedforward path and a memoryless feedback path. However, it should be stressed that the analysis that led to Fig. 3 is *exact* and valid in a general time-variant setting. No approximations or assumptions were made on $\{\mathbf{u}_i, \mu(i)\}$.

A. l_2 -Stability and the Small Gain Theorem

Now that we have introduced the feedback structure of Fig. 3, we can discuss conditions on the step-size parameter $\mu(i)$ in order to guarantee robustness according to (12) for some K .

For this purpose, we start by noting that the feedback configuration of Fig. 3 lends itself rather immediately to stability analysis via tools that are by now standard in system theory, as we explain in the sequel.

It follows from (15) that for every i and for any $\mu(i)$, we have

$$\bar{\mu}(i)|e_a(i)|^2 = \|\tilde{\mathbf{w}}_{i-1}\|^2 - \|\tilde{\mathbf{w}}_i\|^2 + \bar{\mu}(i)|\bar{v}(i)|^2. \quad (16)$$

This allows us to conclude, under a suitable condition on $\mu(i)$, that the system in Fig. 3 is l_2 -stable, i.e., it maps a bounded energy sequence $\{\sqrt{\bar{\mu}(\cdot)} v(\cdot)\}$ to a bounded energy sequence $\{\sqrt{\bar{\mu}(\cdot)} e_a(\cdot)\}$ in a sense precised in (17) below. In fact, we shall also conclude that a similar result will hold even if we replace $\bar{\mu}(\cdot)$ with $\mu(\cdot)$ (cf. (18)).

Define

$$\Delta(N) = \max_{0 \leq i \leq N} \left| 1 - \frac{\mu(i)}{\bar{\mu}(i)} \right| \quad \text{and} \quad \gamma(N) = \max_{0 \leq i \leq N} \frac{\mu(i)}{\bar{\mu}(i)}.$$

That is, $\Delta(N)$ is the maximum absolute value of the gain of the feedback loop over the interval of time $0 \leq i \leq N$. Likewise, $\gamma(N)$ is the maximum value of the scaling factor $\mu(i)/\bar{\mu}(i)$ at the input of the feedback interconnection.

Theorem 1: Consider the gradient-recursion (2), define $\Delta(N)$ and $\gamma(N)$ as above, and let $\bar{\mu}(i) = 1/\|\mathbf{u}_i\|^2$. If $0 < \mu(i) < 2\bar{\mu}(i)$, then the map from $\{\tilde{\mathbf{w}}_{-1}, \sqrt{\bar{\mu}(\cdot)} v(\cdot)\}$ to $\{\sqrt{\bar{\mu}(\cdot)} e_a(\cdot)\}$ is l_2 -stable in the following sense:

$$\sqrt{\sum_{i=0}^N \bar{\mu}(i)|e_a(i)|^2} \leq \quad (17)$$

$$\frac{1}{1 - \Delta(N)} \left[\|\tilde{\mathbf{w}}_{-1}\| + \gamma(N) \sqrt{\sum_{i=0}^N \bar{\mu}(i)|v(i)|^2} \right].$$

Likewise, the map from $\{\tilde{\mathbf{w}}_{-1}, \sqrt{\mu(\cdot)} v(\cdot)\}$ to $\{\sqrt{\mu(\cdot)} e_a(\cdot)\}$ (i.e., with $\bar{\mu}(\cdot)$ replaced by $\mu(\cdot)$) is also l_2 -stable in the following sense:

$$\sqrt{\sum_{i=0}^N \mu(i)|e_a(i)|^2} \leq \quad (18)$$

$$\frac{\gamma^{1/2}(N)}{1 - \Delta(N)} \left[\|\tilde{\mathbf{w}}_{-1}\| + \gamma^{1/2}(N) \sqrt{\sum_{i=0}^N \mu(i)|v(i)|^2} \right].$$

Proof: The proof is based on the triangle inequality of norms and is given in Appendix B. ■

Note that the upper bound on $\mu(i)$ is now $2\bar{\mu}(i)$, which is equivalent to requiring $\Delta(N) < 1$. This is, in fact, a manifestation of the so-called small gain theorem in system analysis [11], [12]. In simple terms, the theorem states that the l_2 -stability of a feedback configuration (that includes Fig. 3 as a special case) requires that the product of the norms of the feedforward and the feedback operators be strictly bounded by one. Here, the feedforward map has (2-induced) norm equal to one (due to its losslessness), whereas the 2-induced norm of the feedback map is $\Delta(N)$.

The fact that the inequalities in Theorem 1 are valid even for $\mu(i)$ in the interval $\bar{\mu}(i) \leq \mu(i) < 2\bar{\mu}(i)$ suggests that a local bound, along the lines of Lemma 1, should also exist for this interval. In fact, this is also the case, as shown in the following statement (the derivation is given in Appendix C).

Lemma 2: Given recursion (2), the following holds for $\bar{\mu}(i) \leq \mu(i) < 2\bar{\mu}(i)$:

$$1 \leq \frac{\|\tilde{\mathbf{w}}_i\|^2 + \mu(i)|e_a(i)|^2}{\|\tilde{\mathbf{w}}_{i-1}\|^2 + \mu(i)|v(i)|^2} \leq \frac{\mu(i)}{2\bar{\mu}(i) - \mu(i)}.$$

Before proceeding further, it will be convenient to introduce a matrix notation for later use: Define

$$\mathbf{M}_N = \text{diag} \{ \mu(i) \}_{i=0}^N, \quad \bar{\mathbf{M}}_N = \text{diag} \{ \bar{\mu}(i) \}_{i=0}^N \quad (19)$$

and $\text{col}(\cdot)$ denotes a column vector of its arguments)

$$\mathbf{e}_{a:N} = \text{col}\{e_a(i)\}_{i=0}^N, \quad \mathbf{v}_N = \text{col}\{v(i)\}_{i=0}^N. \quad (20)$$

It is easy to see that $\Delta(N)$ and $\gamma(N)$ are equal to the 2-induced norms of \mathbf{M}_N and $\bar{\mathbf{M}}_N$, respectively. The condition $\Delta(N) < 1$ then amounts to requiring the (memoryless) feedback map $(\mathbf{I} - \mathbf{M}_N \bar{\mathbf{M}}_N^{-1})$ to be contractive.

B. A Deterministic Convergence Analysis

In order to further appreciate the significance of the (robustness) bounds of Theorem 1, we now exhibit a convergence analysis that is derived in a purely deterministic setting and without statistical assumptions.

It follows from (3) that $\tilde{\mathbf{w}}_i$ satisfies

$$\begin{aligned} \tilde{\mathbf{w}}_i &= \tilde{\mathbf{w}}_{i-1} - \sqrt{\mu(i)} \mathbf{u}_i^* \left[\sqrt{\mu(i)} (e_a(i) + v(i)) \right] \\ &= \tilde{\mathbf{w}}_{i-1} - \hat{\mathbf{u}}_i^* [\hat{e}_a(i) + \hat{v}(i)] \end{aligned} \quad (21)$$

where we have introduced the notation $\hat{\mathbf{u}}_i, \hat{e}_a(i), \hat{v}(i)$ in order to incorporate the factors $\sqrt{\mu(i)}$. We further introduce the following two deterministic conditions on $\{\hat{v}(i), \hat{\mathbf{u}}_i\}$:

- i) *Finite noise energy:* The sequence $\{\hat{v}(\cdot)\}$ is assumed to have finite energy, i.e., $\sum_{i=0}^{\infty} \mu(i)|v(i)|^2 < \infty$.
- ii) *Persistent excitation:* The rows $\{\hat{\mathbf{u}}_i\}$ are assumed to be persistently exciting. By this, we mean that there exists a finite integer $L \geq M$ such that the smallest singular value of $\text{col}\{\hat{\mathbf{u}}_i, \dots, \hat{\mathbf{u}}_{i+L}\}$ is uniformly bounded from below by a positive quantity for sufficiently large i .

In the proof of the next statement, we employ the quantities

$$\Delta = \sup_i |1 - \mu(i)/\bar{\mu}(i)|, \quad \gamma = \sup_i [\mu(i)/\bar{\mu}(i)].$$

Theorem 2: Assume $\mu(i)\|\mathbf{u}_i\|^2$ is uniformly bounded by 2 and $\|\tilde{\mathbf{w}}_{-1}\| < \infty$. If $\{\hat{v}(\cdot)\}$ has finite energy, then $\hat{e}_a(i) \rightarrow 0$. If $\{\hat{\mathbf{u}}_i\}$ is further persistently exciting, then $\mathbf{w}_i \rightarrow \mathbf{w}$.

Proof: If $\{\hat{v}(\cdot)\}$ has finite energy, then (18), for $N \rightarrow \infty$, implies that $\{\hat{e}_a(\cdot)\}$ also has finite energy. This is true since $\gamma < 2$, and $\Delta < 1$. We therefore conclude that $\{\hat{e}_a(\cdot)\}$ is a Cauchy sequence, and hence, $\hat{e}_a(i) \rightarrow 0$.

For the second statement of the theorem, we use (21) to write

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i+k-1} + \sum_{p=i+1}^{i+k-1} \hat{\mathbf{u}}_p^*(\hat{e}_a(p) + \hat{v}(p)).$$

Multiplying from the left by $\hat{\mathbf{u}}_{i+k}$ for $k = 1, \dots, L$, we get

$$\hat{\mathbf{u}}_{i+k} \tilde{\mathbf{w}}_i = \hat{e}_a(i+k) + \sum_{p=i+1}^{i+k-1} \hat{\mathbf{u}}_{i+k} \hat{\mathbf{u}}_p^*(\hat{e}_a(p) + \hat{v}(p)).$$

From the finite noise-energy assumption ($\hat{v}(i) \rightarrow 0$) and the fact that $\hat{e}_a(i) \rightarrow 0$, we conclude that the right-hand side vanishes as $i \rightarrow \infty$, and hence

$$\begin{bmatrix} \hat{\mathbf{u}}_{i+1} \\ \vdots \\ \hat{\mathbf{u}}_{i+L+1} \end{bmatrix} \tilde{\mathbf{w}}_i \rightarrow 0.$$

From the definition of persistent excitation, it follows that $\tilde{\mathbf{w}}_i \rightarrow 0$. ■

Another point of interest is to note that a related limiting result can be given for finite-power noise sequences $\{\hat{v}(\cdot)\}$ (rather than finite-energy), i.e., for $v(\cdot)$ satisfying

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=0}^N \mu(i) |v(i)|^2 = P_v < \infty.$$

For this purpose, we divide both sides of (18) by \sqrt{N} and take the limit as $N \rightarrow \infty$ to conclude that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=0}^N \mu(i) |e_a(i)|^2 \leq \frac{\gamma^2 P_v}{(1 - \Delta)^2}.$$

In other words, a bounded noise power leads to a bounded estimation error power. We may also add that the conclusion of Theorem 2 is in agreement with the result in [21, pp. 140–143] for the noiseless and constant step-size case.

C. Energy Propagation in the Feedback Cascade

More physical insights into the convergence behavior of the gradient recursion (2) can be obtained by studying the energy flow through the feedback configuration of Fig. 3.

Indeed, let us ignore the measurement noise $v(i)$ and assume that we have noiseless measurements $d(i) = \mathbf{u}_i \mathbf{w}$. It is known in the stochastic setting that for Gaussian processes [22], as well as for spherically invariant random processes [23], the maximal speed of convergence is obtained for $\mu(i) = \bar{\mu}(i)$,

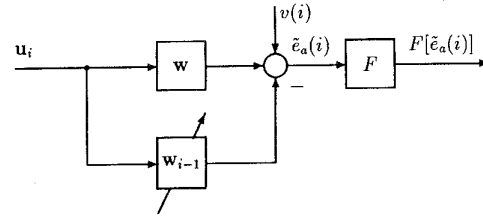


Fig. 4. Structure of filtered-error gradient algorithms.

i.e., for the so-called projection LMS algorithm. We now argue that this conclusion is consistent with the feedback configuration of Fig. 3.

Indeed, for $\mu(i) = \bar{\mu}(i)$, the feedback loop is disconnected. This means that there is no energy flowing back into the lower input of the lossless section from its lower output $e_a(\cdot)$. The losslessness of the feedforward path then implies that

$$E_w(i) = E_w(i-1) - E_e(i) \quad (22)$$

where $E_e(i) = \bar{\mu}(i) |e_a(i)|^2$, and $E_w(i) = \|\tilde{\mathbf{w}}_i\|^2$. Expression (22) implies that the weight-error energy is a nonincreasing function of time, i.e., $E_w(i) \leq E_w(i-1)$.

However, what if $\mu(i) \neq \bar{\mu}(i)$? In this case, the feedback path is active, and the convergence speed will be affected since the rate of decrease in the energy of the weight-error vector will be lowered. Indeed, for $\mu(i) \neq \bar{\mu}(i)$, we obtain

$$E_w(i) = E_w(i-1) - \left(1 - |1 - \mu(i)/\bar{\mu}(i)|^2\right) E_e(i) \quad (23)$$

where the coefficient multiplying $E_e(i)$ can be seen to be smaller than one. Hence, the rate of decrease in the energy of $\tilde{\mathbf{w}}_i$ is lowered.

V. FILTERED-ERROR GRADIENT ALGORITHMS

The feedback loop concept of the former sections applies equally well to gradient algorithms that employ filtered versions of $\tilde{e}_a(i) = d(i) - \mathbf{u}_i \mathbf{w}_{i-1}$. Such algorithms are useful when the error $\tilde{e}_a(i)$ cannot be observed directly but rather a filtered version of it, as indicated in Fig. 4. The operator F denotes the filter that operates on $\tilde{e}_a(i)$. It is assumed to be a finite-impulse response filter of order M_F

$$F(q^{-1})[x(i)] = F[x(i)] = \sum_{j=0}^{M_F-1} f_j x(i-j).$$

It may also be a time-variant filter, in which case, the coefficients f_j will vary with time, say, $f_j(i)$. Typical applications arise in the active control of noise (see, e.g., [13] and [24]–[27]) and in the context of adaptive IIR filtering (see [20] and [28] for a discussion along the lines of this paper).

Here, we limit ourselves to three classes of algorithms that employ filtered error measurements; the filtered-x LMS (FxLMS), the modified filtered-x LMS (MFxLMS), and the filtered error LMS (FELMS).

A. The Filtered-x LMS Algorithm

The filtered-x LMS algorithm employs an update of the form

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu(i)F[\mathbf{u}_i]^*F[\tilde{e}_a(i)] \quad (24)$$

where the input data \mathbf{u}_i is also processed by F . The linearity of F implies that $F[\tilde{e}_a(i)] = F[\mathbf{u}_i\tilde{\mathbf{w}}_{i-1}] + F[v(i)]$.

This algorithm has been analyzed in the literature under a slow adaptation assumption [13], [24], viz., when the variation in the weight estimates is slow over the length of the filter F , $\mathbf{w}_i \approx \mathbf{w}_{i-1} \approx \dots \approx \mathbf{w}_{i-M_F}$.² In this case, we can approximate $F[\mathbf{u}_i\mathbf{w}_{i-1}]$ by $F[\mathbf{u}_i]\mathbf{w}_{i-1}$, leading to the approximate update

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu(i)F[\mathbf{u}_i]^*(F[d(i)] - F[\mathbf{u}_i]\mathbf{w}_{i-1})$$

with $F[d(i)] = F[\mathbf{u}_i]\mathbf{w} + F[v(i)]$. This is of the same form as the standard LMS update (2) with $\{\mathbf{u}_i, d(i), v(i)\}$ replaced by their filtered versions $\{F[\mathbf{u}_i], F[d(i)], F[v(i)]\}$. Hence, our earlier conclusions concerning robustness conditions hold. In particular, the robustness condition now requires $0 < \mu(i) < 2/\|F[\mathbf{u}_i]\|^2$.

B. The Modified Filtered-x LMS Algorithm

Recently, an improvement has been proposed that avoids the slow adaptation assumption [29], [30]. This is achieved by adding two terms to the update recursion (24)

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu(i)F[\mathbf{u}_i]^*(F[\tilde{e}_a(i)] + F[\mathbf{u}_i\mathbf{w}_{i-1}] - F[\mathbf{u}_i]\mathbf{w}_{i-1}).$$

It can be easily verified that because of the additional terms, the above recursion is equivalent to the following:

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu(i)F[\mathbf{u}_i]^*(F[v(i)] + F[\mathbf{u}_i](\mathbf{w} - \mathbf{w}_{i-1})).$$

This is again of the same form as the LMS update (2) but with the filtered input sequence $F[\mathbf{u}_i]$ and the filtered noise sequence $F[v(i)]$. This time, though, no approximation has been employed. The results of the previous sections will then be immediately applicable with the proper change of variables. For example, the second bound of Theorem 1 will read as follows: Define $e_a(i) = F[\mathbf{u}_i]\tilde{\mathbf{w}}_{i-1}$, $\bar{\mu}(i) = 1/\|F[\mathbf{u}_i]\|^2$, and

$$\Delta(N) = \max_{0 \leq i \leq N} |1 - \mu(i)/\bar{\mu}(i)|, \gamma(N) = \max_{0 \leq i \leq N} \mu(i)/\bar{\mu}(i).$$

If $0 < \mu(i) < 2\bar{\mu}(i)$, then the map from $\{\tilde{\mathbf{w}}_{-1}, \sqrt{\mu(\cdot)}F[v(\cdot)]\}$ to $\{\sqrt{\mu(\cdot)}e_a(\cdot)\}$ is l_2 -stable in the following sense:

$$\sqrt{\sum_{i=0}^N \mu(i)|e_a(i)|^2} \leq \frac{\gamma^{1/2}(N)}{1 - \Delta(N)} \left[\|\tilde{\mathbf{w}}_{-1}\| + \gamma^{1/2}(N) \sqrt{\sum_{i=0}^N \mu(i)|F[v(i)]|^2} \right].$$

We may add that we have used the insights provided by the above analysis to propose two modifications to the FxLMS algorithm with improved computational requirements and convergence performance [27].

²This is often a reasonable assumption especially when the length M_F is considerably smaller than the adaptive filter length M .

C. The Filtered-Error LMS Algorithm

We now discuss the filtered-error LMS algorithm [31], which leads to a more involved feedback structure; it forces a dynamic system in the feedback loop rather than a memoryless transformation. Its update retains the input vector unchanged and takes the form

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu(i)\mathbf{u}_i^*F[\tilde{e}_a(i)]. \quad (25)$$

In contrast with the FxLMS algorithm and its modified form, the error-path filter F does not need to be known explicitly, and the algorithm also requires less computation. Similar update forms arise in the context of IIR modeling, such as Feintuch's algorithm [32] and the so-called simplified hyperstable adaptive recursive filter (SHARF) [33] (see, e.g., [28] for a discussion along the lines of this paper).

Before proceeding further, we may remark that the transfer function description can also be applied to the filtered-error LMS, thus leading to $E_a(z)/V(z) \cong$

$$\frac{\frac{\mu C^2 M}{2} F(z)(1 - z \cos(\Omega))}{z^2 - 2z \cos(\Omega) \left(1 - \frac{\mu C^2 M}{4} F(z)\right) + \left(1 - \frac{\mu C^2 M}{2} F(z)\right)}$$

where $e_a(i) = \mathbf{u}_i\tilde{\mathbf{w}}_{i-1}$. This can again be described in terms of an allpass feedforward path and a dynamic feedback loop given by $1 - \frac{\mu}{\bar{\mu}}F(z)$. That is, the feedback loop is not memoryless anymore. A similar structure also arises for a generalization of the filtered-x LMS algorithm to the case of an array of slowly-varying FIR adaptive filters [34].

1) *The Feedback Structure for FELMS:* Following the discussion that led to (14), it can be verified that (25) is equivalent to the following update:

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \bar{\mu}(i)\mathbf{u}_i^*[e_a(i) + \bar{v}(i)] \quad (26)$$

where $\bar{\mu}(i) = 1/\|\mathbf{u}_i\|^2$, $e_a(i) = \mathbf{u}_i\tilde{\mathbf{w}}_{i-1}$, and the modified "noise" sequence $\{\bar{v}(\cdot)\}$ is defined via³

$$\bar{\mu}(i)\bar{v}(i) = \mu(i)F[v(i)] - \bar{\mu}(i)e_a(i) + \mu(i)F[e_a(i)]. \quad (27)$$

Expression (26) is of the same form as (14), which readily implies that the following relation also holds:

$$\frac{\|\tilde{\mathbf{w}}_i\|^2 + \bar{\mu}(i)|e_a(i)|^2}{\|\tilde{\mathbf{w}}_{i-1}\|^2 + \bar{\mu}(i)|\bar{v}(i)|^2} = 1. \quad (28)$$

This establishes that the map from $\{\tilde{\mathbf{w}}_{i-1}, \sqrt{\bar{\mu}(i)}\bar{v}(i)\}$ to $\{\tilde{\mathbf{w}}_i, \sqrt{\bar{\mu}(i)}e_a(i)\}$, which is denoted by \bar{T}_i , is lossless and that the map from the original disturbance $\sqrt{\bar{\mu}(\cdot)}v(\cdot)$ to $\sqrt{\bar{\mu}(\cdot)}e_a(\cdot)$ can be expressed in terms of a feedback structure, as shown in Fig. 5. We remark that the notation

$$1 - \frac{\mu(i)}{\sqrt{\bar{\mu}(i)}} F[\cdot] \frac{1}{\sqrt{\bar{\mu}(i)}}$$

which appears in the feedback loop, should be interpreted as follows: We first divide $\sqrt{\bar{\mu}(i)}e_a(i)$ by $\sqrt{\bar{\mu}(i)}$, followed by

³This is different from the sequence $\{\bar{v}(\cdot)\}$ in (14), but we have opted to use the same symbol to always denote the modified noise sequence, both for uniformity of notation and in order to avoid a burden of symbols.

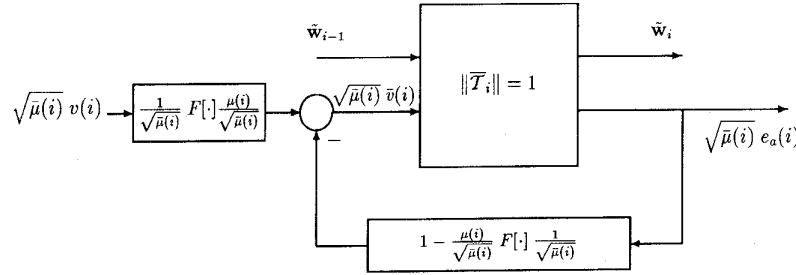


Fig. 5. Filtered-error LMS algorithm as a time-variant lossless mapping with dynamic feedback.

the filter F , and then by a subsequent scaling by $\mu(i)/\sqrt{\mu(i)}$. Likewise, the term $\sqrt{\mu(i)} v(i)$ is first divided by $\sqrt{\mu(i)}$, then filtered by F , and finally scaled by $\mu(i)/\sqrt{\mu(i)}$.

The feedback loop now consists of a dynamic system but we can still proceed to study the l_2 -stability of the overall configuration in much the same way as we did in Section IV. For this purpose, we use the vector and matrix quantities introduced in (19) and (20) and further define a vector $\bar{\mathbf{v}}_N$ similar to \mathbf{v}_N but with the entries $\bar{v}(\cdot)$ (of (27)) instead of $v(\cdot)$.

We also define the lower triangular matrix \mathbf{F}_N that describes the action of the filter F on a sequence at its input. This is generally a band matrix since $M_F \ll M$, as shown below for the special case $M_F = 3$

$$\mathbf{F}_N = \begin{bmatrix} f_0 & & & & \\ f_1 & f_0 & & & \\ f_2 & f_1 & f_0 & & \\ & f_2 & f_1 & f_0 & \\ & & \ddots & \ddots & \ddots \end{bmatrix}$$

It is thus immediate to verify that the successive outputs of $F[e_a(\cdot)]$ can be obtained by simply computing the matrix-vector product $\mathbf{F}_N \mathbf{e}_{a,N}$. In addition, if the filter F were time variant, all that changes is that the matrix \mathbf{F}_N will not be Toeplitz anymore. Instead, its first diagonal will consist of the values of the first coefficient $f_0(\cdot)$ at the successive time instants, and so on.

Define

$$\Delta(N) = \|\mathbf{I} - \bar{\mathbf{M}}_N^{-\frac{1}{2}} \mathbf{M}_N \mathbf{F}_N \bar{\mathbf{M}}_N^{-\frac{1}{2}}\|_{2,ind}$$

$$\gamma(N) = \|\bar{\mathbf{M}}_N^{-\frac{1}{2}} \mathbf{M}_N \mathbf{F}_N \bar{\mathbf{M}}_N^{-\frac{1}{2}}\|_{2,ind}$$

If we now follow the arguments of Section IV, we obtain the following result, which extends Theorem 1.

Theorem 2: Consider the filtered-error LMS recursion (25) and define $\Delta(N)$ and $\gamma(N)$ as above. If $\Delta(N) < 1$, then the map from $\{\tilde{\mathbf{w}}_{-1}, \sqrt{\mu(\cdot)} v(\cdot)\}$ to $\{\sqrt{\mu(\cdot)} e_a(\cdot)\}$ is l_2 -stable, i.e.,

$$\sqrt{\sum_{i=0}^N \mu(i) |e_a(i)|^2} \leq \frac{1}{1 - \Delta(N)} \left[\|\tilde{\mathbf{w}}_{-1}\| + \gamma(N) \sqrt{\sum_{i=0}^N \mu(i) |v(i)|^2} \right]$$

Moreover, the map from $\{\tilde{\mathbf{w}}_{-1}, \sqrt{\mu(\cdot)} v(\cdot)\}$ to $\{\sqrt{\mu(\cdot)} e_a(\cdot)\}$ is l_2 -stable with

$$\sqrt{\sum_{i=0}^N \mu(i) |e_a(i)|^2} \leq \frac{\gamma^{1/2}(N)}{1 - \Delta(N)} \left[\|\tilde{\mathbf{w}}_{-1}\| + \gamma^{1/2}(N) \sqrt{\sum_{i=0}^N \mu(i) |v(i)|^2} \right]$$

We thus see that the major requirement is for the feedback matrix $(\mathbf{I} - \bar{\mathbf{M}}_N^{-\frac{1}{2}} \mathbf{M}_N \mathbf{F}_N \bar{\mathbf{M}}_N^{-\frac{1}{2}})$ to be strictly contractive. We denote it by \mathbf{P}_N , which can be easily seen to have the following triangular form (it also has a band of width M_F):

$$\mathbf{P}_N = \begin{pmatrix} 1 - \frac{\mu(0)}{\bar{\mu}(0)} f_0 & & & \mathbf{O} \\ -\frac{\mu(1)}{\sqrt{\bar{\mu}(0)\bar{\mu}(1)}} f_1 & 1 - \frac{\mu(1)}{\bar{\mu}(1)} f_0 & & \\ -\frac{\mu(2)}{\sqrt{\bar{\mu}(0)\bar{\mu}(2)}} f_2 & -\frac{\mu(2)}{\sqrt{\bar{\mu}(1)\bar{\mu}(2)}} f_1 & 1 - \frac{\mu(2)}{\bar{\mu}(2)} f_0 & \\ \vdots & & & \ddots \end{pmatrix}$$

We see that the entries of \mathbf{P}_N depend on three parameters:

- 1) the step sizes $\mu(i)$;
- 2) the energies of the input sequence $\bar{\mu}(i)$;
- 3) the error filter F .

Several special cases may be of interest. For example, the special case $F = 1$ (i.e., no filter) immediately leads to the case we encountered earlier in Section IV. Another special case is $F = q^{-1}$ (i.e., a simple delay). The filtered-error LMS recursion (25) then collapses to the delayed-error LMS $\mathbf{w}_i = \mathbf{w}_{i-1} + \mu(i) \mathbf{u}_i^* \tilde{e}_a(i-1)$. The corresponding \mathbf{P}_N is given by (since $f_0 = 0$ and $f_1 = 1$)

$$\mathbf{P}_N = \begin{pmatrix} 1 & & & \mathbf{O} \\ -\frac{\mu(1)}{\sqrt{\bar{\mu}(0)\bar{\mu}(1)}} f_1 & 1 & & \\ & -\frac{\mu(2)}{\sqrt{\bar{\mu}(1)\bar{\mu}(2)}} f_1 & 1 & \\ \vdots & & & \ddots \end{pmatrix}$$

We see that \mathbf{P}_N cannot be a strictly contractive matrix since its leading (0,0) entry is not less than one. This is consistent with results in the literature, where it has been observed that the delayed-error LMS algorithm usually leads to unstable behavior. We also see from the general expression for \mathbf{P}_N

that a simple gain filter $F = f_0$ with a negative f_0 leads to a noncontractive \mathbf{P}_N .

2) *The Projection FELMS Algorithm:* An important special case of the FELMS algorithm is one that employs a step size of the form $\mu(i) = \alpha \bar{\mu}(i)$, $\alpha > 0$. That is, $\mu(i)$ is a scaled multiple of the reciprocal input energy. This leads to the projection FELMS algorithm

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \alpha \frac{\mathbf{u}_i^*}{\|\mathbf{u}_i\|^2} F[\tilde{e}_a(i)]. \quad (29)$$

In this case, it can be seen that the contractivity requirement collapses to requiring the strict contractivity of

$$\mathbf{P}_N = \begin{pmatrix} 1 - \alpha f_0 & & & \mathbf{0} \\ -\alpha \frac{\sqrt{\bar{\mu}(1)}}{\sqrt{\bar{\mu}(0)}} f_1 & 1 - \alpha f_0 & & \\ -\alpha \frac{\sqrt{\bar{\mu}(2)}}{\sqrt{\bar{\mu}(0)}} f_2 & -\alpha \frac{\sqrt{\bar{\mu}(2)}}{\sqrt{\bar{\mu}(1)}} f_1 & 1 - \alpha f_0 & \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

If we further assume that the energy of the input sequence \mathbf{u}_i does not change very rapidly over the filter length M_F , i.e., $\bar{\mu}(i) \approx \dots \approx \bar{\mu}(i - M_F)$, then \mathbf{P}_N collapses to

$$\mathbf{P}_N \approx \mathbf{I} - \alpha \mathbf{F}_N. \quad (30)$$

In this case, the strict contractivity of $(\mathbf{I} - \alpha \mathbf{F}_N)$ can be guaranteed by choosing α such that

$$\max_{\Omega} |1 - \alpha F(e^{j\Omega})| < 1 \quad (31)$$

where $F(z)$ is the transfer function of the error filter. This suggests, according to the energy arguments in Section IV-C, that for faster convergence (i.e., for smallest feedback gain), we should choose α optimally by solving the min-max problem:

$$\min_{\alpha} \max_{\Omega} |1 - \alpha F(e^{j\Omega})|. \quad (32)$$

If the resulting minimum is less than 1, then the corresponding optimum α will result in faster convergence and l_2 -stability (or robustness). Simulation results that confirm these conclusions are discussed in the next section.

Expression (32) provides a criterion for choosing the step-size parameter in the filtered-error case in order to speed up the convergence of the PELMS algorithm (for slow-varying input energy).

VI. SIMULATION RESULTS

The simulations in this section were carried out with the projection filtered-error algorithm (29) for the error-path filter:

$$F(q) = 1 - 1.2q^{-1} + 0.72q^{-2}.$$

The row vectors \mathbf{u}_i were taken with shift structure, with the individual entries $u(i)$ arising from a sinusoidal excitation of frequency Ω_0 . In this case, if we assume that the *a priori* error signal is dominated by the frequency component Ω_0 , then we can solve for the optimum α via the simpler expression (cf. (32)) $\min_{\alpha} |1 - \alpha F(e^{j\Omega_0})|$. This minimization can be solved

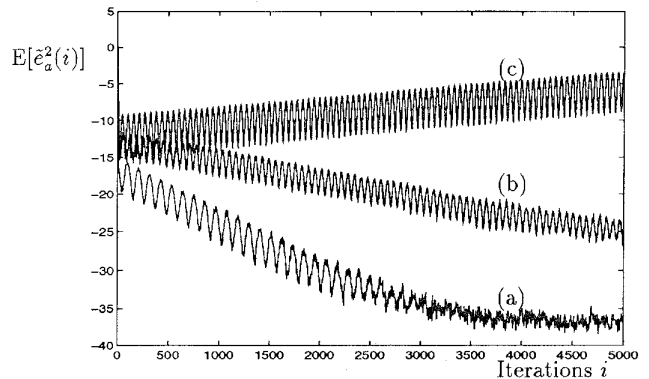


Fig. 6. Convergence behavior for FELMS algorithm with sinusoidal input sequence and various step-sizes (a) $\alpha = 0.085$; (b) 0.15 ; (c) 0.18 .

explicitly, and we get $\alpha_{opt} = \text{Real}(1/F(e^{-j\Omega_0}))$ to be the step size that causes faster convergence speed. Following the same procedure, the step size α_{lim} for which the stability limit is achieved (cf. (31)) can be calculated to be $\alpha_{lim} = 2\alpha_{opt}$.

To verify these statements, we created an input sequence of the form $u(i) = \sin[1.2i + \phi]$, where 50 different values for ϕ were uniformly chosen from the interval $[-\pi, \pi]$. The reason for adding a random phase ϕ is to obtain smoother learning curves after averaging. The optimal step size for F , α_{opt} , can be calculated to be $\alpha_{opt} = 0.085$, and the stability bound is obtained for $\alpha_{lim} = 0.17$. Fig. 6 shows three runs of the FELMS for the choices $\alpha = 0.085$, $\alpha = 0.15$ and $\alpha = 0.18$. As expected, the first value of α leads to the fastest convergence speed. In every simulation, we averaged over 50 trials. The additive noise $v(i)$ was assumed to be -40 dB below the input power during the experiments, and the order of the adaptive filter was set to $M = 10$. The algorithm was run for $N = 5000$ iterations. We also see that for the first two values of α , the sample average of $|\tilde{e}_a(i)|^2$ decreases with time, whereas for the last value, it increases.

The second part of our simulations is intended to measure the impact of the slow-energy variation assumption of the input sequence on the algorithm performance and, in particular, on the conditions (31) and (32).

For this purpose, the difference $(\bar{\mu}^{-1}(i) - \bar{\mu}^{-1}(i-1))^2$ can be used to measure the extent of variation in the input energy. This difference collapses to $(|u(i)|^2 - |u(i-M)|^2)^2$ due the assumed shift structure for \mathbf{u}_i . Since our data were generated randomly, the above measure can be approximated by $E(|u(i)|^2 - |u(i-M)|^2)^2$, which can be further reduced to $2(\kappa-1)E[|u(i)|^2]^2$, if we assume that $|u(i)|^2$ is uncorrelated with $|u(i-M)|^2$. This is a reasonable assumption for M relatively large. Hence, the kurtosis parameter κ gives a measure of the input energy variations. To describe the impact of these variations on the FELMS algorithm, we applied various random input sequences with different kurtosis, as listed in Table I.

All random sequences were white processes with variance one. Fig. 7 depicts the simulation results, where we again averaged over 50 trials. Unlike the previous simulations, the

TABLE I
KURTOSIS OF VARIOUS DENSITY FUNCTIONS

Distribution	Kurtosis
Bipolar	1
Uniform	1.8
Gaussian	3
K_0 (Modified Bessel)	9
Gamma	11.66

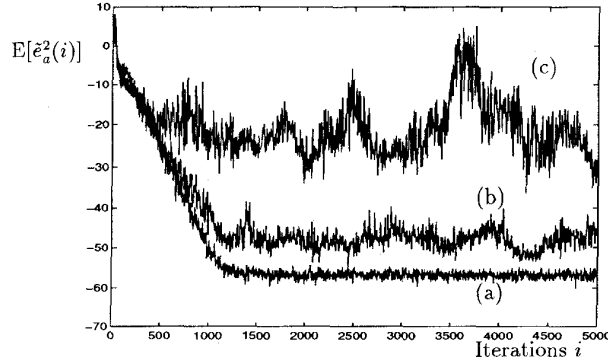


Fig. 7. Learning curves for FELMS algorithm with F and various random input processes with different pdf ($\alpha = 0.3$): (a) Bipolar, uniform, Gaussian; (b) K_0 ; (c) Gamma.

power of the additive noise $v(i)$ was set at -60 dB relative to the input sequence in order to better observe the effect of the various kurtosis. As can be seen from the figure, the higher the kurtosis, the more the variation in the steady-state value. The convergence speed and the stability bounds, however, remained practically unchanged.

VII. CONCLUDING REMARKS

We have provided a study of the robustness and convergence performance of gradient adaptive schemes in a purely deterministic framework by following a time-domain feedback analysis. In particular, conditions on the step-size parameters were derived in order to result in overall l_2 -stable structures and faster convergence. No prior statistical assumptions or independence conditions were employed. Instead, local and global passivity relations were emphasized along with energy propagation arguments within a feedback cascade. Simulation results were included to demonstrate the theoretical findings.

We may add that the analysis of this paper can also be carried out in a stochastic setting by determining conditions on the error path filter in order to guarantee an l_2 -stable map between variances of relevant stochastic variables. The details will be pursued elsewhere.

Moreover, the analysis also extends to other classes of algorithms, e.g., to block adaptive filters and to updates that involve nonlinear functionals (such as Perceptron training and sign algorithms) [35]. The results can also be shown to be related to developments in H^∞ -theory and can be extended to a class of nonlinear robust adaptive filters [20].

APPENDIX A

DERIVATION OF THE TRANSFER FUNCTION (5)

The k th entry of the vector \mathbf{u}_i is

$$u(i-k) = C \cos[\Omega(i-k)] = \frac{C}{2} (e^{j(i-k)\Omega} + e^{-j(i-k)\Omega}).$$

Correspondingly, the update for the k th entry of the weight-error vector $(\tilde{\mathbf{w}}_i)_k$ is given by

$$(\tilde{\mathbf{w}}_i)_k = (\tilde{\mathbf{w}}_{i-1})_k - \frac{\mu C}{2} [e^{j(i-k)\Omega} + e^{-j(i-k)\Omega}] \tilde{e}_a(i).$$

Ignoring initial conditions and using the z -transform, we obtain

$$\tilde{\mathbf{W}}_k(z) = -\frac{z}{z-1} \frac{\mu C}{2} (\tilde{E}_a(z e^{-j\Omega}) e^{-jk\Omega} + \tilde{E}_a(z e^{j\Omega}) e^{jk\Omega}).$$

Likewise, using $e_a(i) = \sum_{k=0}^{M-1} u(i-k)(\tilde{\mathbf{w}}_{i-1})_k$, we obtain $E_a(z) =$

$$\frac{C}{2} z^{-1} \sum_{k=0}^{M-1} [\tilde{\mathbf{W}}_k(z e^{-j\Omega}) e^{j\Omega(1-k)} + \tilde{\mathbf{W}}_k(z e^{j\Omega}) e^{j\Omega(k-1)}].$$

Substituting the expression for $\tilde{\mathbf{W}}_k(z)$ and ignoring the effects of the mixing terms $\tilde{E}_a(z e^{2j\Omega})$ and $\tilde{E}_a(z e^{-2j\Omega})$ for large enough M [13], we can write

$$\begin{aligned} E_a(z) &\cong -\frac{\mu C^2}{4} \sum_{k=0}^{M-1} \tilde{E}_a(z) \left(\frac{1}{z e^{-j\Omega} - 1} + \frac{1}{z e^{j\Omega} - 1} \right), \\ &= \frac{\mu C^2 M}{2} \frac{1 - z \cos(\Omega)}{z^2 - 2z \cos(\Omega) + 1} \tilde{E}_a(z). \end{aligned}$$

However, we know that $\tilde{E}_a(z) = E_a(z) + V(z)$, and consequently, the transfer function from $v(\cdot)$ to $e_a(\cdot)$ is approximately (5).

APPENDIX B

PROOF OF THEOREM 1

If we compute the sum of both sides of (16) and ignore $\|\tilde{\mathbf{w}}_N\|^2$, we obtain

$$\sqrt{\sum_{i=0}^N \bar{\mu}(i) |e_a(i)|^2} \leq \|\tilde{\mathbf{w}}_{-1}\| + \sqrt{\sum_{i=0}^N \bar{\mu}(i) |\bar{v}(i)|^2}.$$

However, it follows from the relation for $\bar{v}(i) = -e_p(i)$ in (13), and from the triangular inequality of norms, that

$$\sqrt{\sum_{i=0}^N \bar{\mu}(i) |\bar{v}(i)|^2} \leq$$

$$\sqrt{\sum_{i=0}^N \frac{\mu^2(i)}{\bar{\mu}(i)} |v(i)|^2} + \sqrt{\sum_{i=0}^N \left| 1 - \frac{\mu(i)}{\bar{\mu}(i)} \right|^2 \bar{\mu}(i) |e_a(i)|^2}.$$

Combining with the first inequality above and using the definitions of $\Delta(N)$ and $\gamma(N)$ prior to the statement of

Theorem 1, we conclude that

$$\sqrt{\sum_{i=0}^N \bar{\mu}(i) |e_a(i)|^2} \leq \|\tilde{\mathbf{w}}_{-1}\| + \gamma(N) \sqrt{\sum_{i=0}^N \bar{\mu}(i) |v(i)|^2} + \Delta(N) \sqrt{\sum_{i=0}^N \bar{\mu}(i) |e_a(i)|^2}.$$

When $(1 - \Delta(N)) > 0$, we conclude that (17) holds.

As for the second bound, we note that

$$\sum_{i=0}^N \mu(i) |e_a(i)|^2 \leq \gamma(N) \sum_{i=0}^N \bar{\mu}(i) |e_a(i)|^2$$

and use similar arguments to the above to conclude that (18) holds.

APPENDIX C

DERIVATION OF THE BOUND IN LEMMA 2

Define, for notational convenience, $\alpha(i) = \mu(i)/\bar{\mu}(i)$. If we substitute $\tilde{v}(i) = \alpha(i)v(i) - (1 - \alpha(i))e_a(i)$ into (15), we get the following equality:

$$\frac{(2 - \alpha(i)) \left[\|\tilde{\mathbf{w}}_i\|^2 + \mu(i) |e_a(i)|^2 \right]}{\text{den}(i)} = 1$$

where the expression for the denominator is

$$\begin{aligned} \text{den}(i) &= \alpha(i) \left(\|\tilde{\mathbf{w}}_{i-1}\|^2 + \mu(i) |v(i)|^2 \right) - \\ &\quad (1 - \alpha(i)) [\mu(i) (v(i) e_a^*(i) + v^*(i) e_a(i))] - \\ &\quad (1 - \alpha(i)) \left[\|\tilde{\mathbf{w}}_i\|^2 + \|\tilde{\mathbf{w}}_{i-1}\|^2 \right]. \end{aligned}$$

We now verify that for $1 \leq \alpha(i) < 2$

$$-\{ \mu(i) [v(i) e_a^*(i) + v^*(i) e_a(i)] - \|\tilde{\mathbf{w}}_i\|^2 - \|\tilde{\mathbf{w}}_{i-1}\|^2 \} \geq 0.$$

Indeed, if we substitute for $\tilde{\mathbf{w}}_i$, as given by (3), and use $\tilde{e}_a(i) = e_a(i) + v(i) = \mathbf{u}_i \tilde{\mathbf{w}}_{i-1} + v(i)$, we obtain that the above inequality is equivalent to verifying the nonnegativity of the following expression:

$$\begin{bmatrix} \tilde{\mathbf{w}}_{i-1}^* & v(i) \end{bmatrix} \mathbf{C} \begin{bmatrix} \tilde{\mathbf{w}}_{i-1} \\ v^*(i) \end{bmatrix}$$

where the central matrix is

$$\mathbf{C} = \begin{bmatrix} 2\mathbf{I} + \mu(i)(\alpha(i) - 2)\mathbf{u}_i^* \mathbf{u}_i & \mu(i)(\alpha(i) - 2)\mathbf{u}_i^* \\ \mu(i)(\alpha(i) - 2)\mathbf{u}_i & \mu(i)\alpha(i) \end{bmatrix}.$$

We now verify that the central matrix is positive semidefinite. Its top left-corner block is positive semidefinite, and its Schur complement S can be seen to be

$$S = \frac{2\mu(i)\alpha(i)(\alpha(i) - 1)}{2 + \alpha(i)(\alpha(i) - 2)}$$

which is larger than or equal to zero for $\alpha(i)$ between one and two. It then follows that we can write

$$\frac{(2 - \alpha(i)) \left[\|\tilde{\mathbf{w}}_i\|^2 + \mu(i) |e_a(i)|^2 \right]}{\alpha(i) \left[\|\tilde{\mathbf{w}}_{i-1}\|^2 + \mu(i) \alpha(i) |v(i)|^2 \right]} \leq 1$$

which establishes the result of Lemma 2.

REFERENCES

- [1] P. P. Khargonekar and K. M. Nagpal, "Filtering and smoothing in an H^∞ -setting," *IEEE Trans. Automat. Contr.*, vol. 36, pp. 151-166, 1991.
- [2] M. Green and D. J. N. Limebeer, *Linear Robust Control*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
- [3] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
- [4] I. Yaesh and U. Shaked, " H^∞ -optimal estimation: The discrete time case," in *Proc. Inter. Symp. Mathematical Theory Networks Syst.*, Kobe, Japan, June 1991, pp. 261-267.
- [5] B. Hassibi, A. H. Sayed, and T. Kailath, "Linear estimation in Krein spaces—Part II: Applications," *IEEE Trans. Automat. Contr.*, vol. 41, pp. 34-49, Jan. 1996.
- [6] M. Rupp, "Bursting in the LMS algorithm," *IEEE Trans. Signal Processing*, vol. 43, pp. 2414-2417, Oct. 1995.
- [7] B. Widrow, "Adaptive sampled-data systems," in *Proc. 1st IFAC Cong.*, Moscow, U.S.S.R., 1960.
- [8] I. D. Landau, *Adaptive Control—The Model Reference Approach*. New York: Dekker, 1979.
- [9] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press, 1983.
- [10] I. D. Landau, "A feedback system approach to adaptive filtering," *IEEE Trans. Inform. Theory*, vol. IT-30, Mar. 1984.
- [11] H. K. Khalil, *Nonlinear Systems*. New York: MacMillan, 1992.
- [12] M. Vidyasagar, *Nonlinear Systems Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1993, 2nd ed.
- [13] J. R. Glover, Jr., "Adaptive noise canceling applied to sinusoidal interferences," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, pp. 484-491, Dec. 1977.
- [14] P. M. Clarkson and P. R. White, "Simplified analysis of the LMS adaptive filter using a transfer function approximation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 987-993, July 1987.
- [15] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, Signal Processing Series, 1985.
- [16] A. H. Sayed and M. Rupp, "On the robustness, convergence, and minimax performance of instantaneous-gradient adaptive filters," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Oct. 1994, pp. 592-596.
- [17] ———, "A time-domain feedback analysis of adaptive gradient algorithms via the small gain theorem," in *Proc. SPIE Conf. Advanced Signal Processing: Algorithms, Architectures, Implementations*, vol. 2563, July 1995, San Diego, pp. 458-469.
- [18] B. Hassibi, A. H. Sayed, and T. Kailath, "LMS is H^∞ optimal," in *Proc. Conf. Decision Contr.*, vol. 1, San Antonio, Dec. 1993, pp. 74-79; also, *IEEE Trans. Signal Processing*, vol. 44, pp. 267-280, Feb. 1996.
- [19] A. H. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering," *IEEE Signal Processing Mag.*, vol. 11, no. 3, pp. 18-60, July 1994.
- [20] A. H. Sayed and M. Rupp, "A class of nonlinear adaptive H^∞ -filters with guaranteed l_2 -stability," *Proc. IFAC Symp. Nonlinear Contr. Syst. Des.*, vol. 1, Tahoe City, CA, June 1995, pp. 455-460; also, *Automatica*, to be published.
- [21] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms: Stability and Performance*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
- [22] N. J. Bershad, "Analysis of the normalized LMS algorithm with Gaussian inputs," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, no. 4, pp. 793-806, Aug. 1986.
- [23] M. Rupp, "The behavior of LMS and NLMS algorithms in the presence of spherically invariant processes," *IEEE Trans. Signal Processing*, vol. 41, pp. 1149-1160, Mar. 1993.
- [24] D. R. Morgan, "An analysis of multiple correlation cancellation loops with a filter in the auxiliary path," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 454-467, Aug. 1980.
- [25] L. J. Eriksson, "Development of the filtered-U algorithm for active noise control," *J. Acoust. Soc. Amer.*, vol. 89, no. 1, pp. 257-265, Apr. 1991.
- [26] S. Laugesen and S. Elliott, "Multichannel active control of random noise in a small reverberant room," *IEEE Trans. Speech, Audio Processing*, vol. 1, pp. 241-249, Apr. 1993.
- [27] M. Rupp and A. H. Sayed, "Modified FxLMS algorithms with improved convergence performance," in *Proc. 29th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Oct. 1995.
- [28] ———, "On the stability and convergence of Feintuch's algorithm for adaptive IIR filtering," in *Proc. ICASSP*, vol. 2, Detroit, MI, May 1995, pp. 1388-1391.
- [29] E. Bjarnason, "Noise cancellation using a modified form of the filtered-XLMS algorithm," in *Proc. Eusipco Signal Processing V*, Brussels, Belgium, 1992.

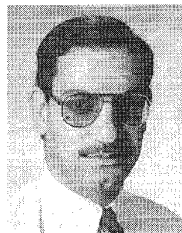
- [30] I. Kim, H. Na, K. Kim, and Y. Park, "Constraint filtered-x and filtered-u algorithms for the active control of noise in a duct," *J. Acoust. Soc. Amer.*, vol. 95, no. 6, pp. 3397-3389, June 1994.
- [31] S. Dasgupta, J. S. Garnett, and C. R. Johnson, Jr., "Convergence of an adaptive filter with signed filtered error," *IEEE Trans. Signal Processing*, vol. 42, pp. 946-950, Apr. 1994.
- [32] P. L. Feintuch, "An adaptive recursive LMS filter," *Proc. IEEE*, vol. 64, pp. 1622-1624, Nov. 1976.
- [33] M. G. Larimore, J. R. Treichler, and C. R. Johnson, "SHARF: An algorithm for adapting IIR digital filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 428-440, Aug. 1980.
- [34] S. J. Elliot, I. M. Stothers, and P. A. Nelson, "A multiple error LMS algorithm and its application to the active control of sound and vibration," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-35, pp. 1423-1434, Oct. 1987.
- [35] A. H. Sayed and M. Rupp, "A feedback analysis of Perceptron learning for neural networks, to appear in *Proc. 29th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Oct. 1995.



Markus Rupp received the Diploma in electrical engineering from FHS, Saarbruecken, Germany, and Universitaet of Saarbruecken in 1984 and 1988, respectively. During this time, he was also a lecturer in digital signal processing and high frequency techniques at the FHS. He received the Doctoral degree in 1993 *summa cum laude* at the TH Darmstadt in the field of acoustical echo cancellation.

He received the DAAD Post-Doctoral fellowship and spent November 1993 until September 1995 with the Department of Electrical and Computer

Engineering of the University of California, Santa Barbara. Since October 1995, he has been with the AT&T Wireless Technology Research Group, Holmdel, NJ, where he is currently working on new adaptive equalization schemes. His interests involve algorithms for speech applications, speech enhancement and recognition, echo compensation, adaptive filter theory, H_∞ -filtering, neural nets, classification algorithms, and signal detection.



Ali H. Sayed (M'92) was born in São Paulo, Brazil. In 1981, he graduated in first place in the National Lebanese Baccalaureat, with the highest score in the history of the examination. In 1987, he received the degree of engineer in electrical engineering from the University of São Paulo, Brazil, and was the first place graduate of the School of Engineering. In 1989, he received, with distinction, the M.S. degree in electrical engineering from the University of São Paulo, Brazil, and was awarded a FAPESP fellowship for overseas studies. In 1992, he received

the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA.

From September 1992 to August 1993, he was a Research Associate with the Information Systems Laboratory at Stanford University, after which he joined, as an Assistant Professor, the Department of Electrical and Computer Engineering at the University of California, Santa Barbara. His research interests are in the areas of adaptive and statistical signal processing, linear and nonlinear filtering and estimation, interplays between signal processing and control methodologies, interpolation theory, and structured computations in systems and mathematics.

Dr. Sayed is a member of SIAM and ILAS. He was awarded the Institute of Engineering Prize and the Conde Armando Alvares Penteado Prize, both in 1987 in Brazil. He is a recipient of a 1994 NSF Research Initiation Award and of the 1996 IEEE Donald G. Fink Prize.