A Unified Approach to the Steady-State and Tracking Analyses of Adaptive Filters

Nabil R. Yousef, *Student Member, IEEE,* and Ali H. Sayed, *Fellow, IEEE*

Abstract—Most adaptive filters are inherently nonlinear and time-variant systems. The nonlinearities in the update equations tend to lead to difficulties in the study of their steady-state performance as a limiting case of their transient performance. This paper develops a unified approach to the steady-state and tracking analyses of adaptive algorithms that bypasses many of these difficulties. The approach is based on the study of the energy flow through each iteration of an adaptive filter, and it relies on a fundamental error variance relation.

Index Terms—Adaptive filter, feedback analysis, mean-square error, steady-state analysis, tracking analysis, transient analysis.

I. INTRODUCTION

THE performance of an adaptive filter is generally measured in terms of its transient behavior and its steady-state behavior. The former provides information about the stability and the convergence rate of an adaptive filter, whereas the latter provides information about the mean-square-error of the filter once it reaches steady state. Although the steady-state performance essentially corresponds to only one point on the learning curve of an adaptive filter, there are many situations where this information is of value by itself.

As is known, there have been numerous works in the literature on the performance of adaptive filters (see, e.g., [1]–[7] and the references therein). The prevailing approach to steady-state analysis has been to obtain steady-state results as the limiting case of a transient analysis. While this procedure is adequate for understanding both the steady-state and the transient behavior of an adaptive algorithm, it can encounter some difficulties. First, transient analyses tend to be laborious, especially for adaptive filters with nonlinear update equations. This is because they rely explicitly on a recursion for the weight-error variance, and recursions of this kind can become complicated for complex algorithms. This explains why more elaborate steady-state results exist for some adaptive filters than others. Second, transient analyses tend to require some simplifying assumptions, which at times can be restrictive, such as requiring the independence of certain vectors that are otherwise dependent. In this way, by obtaining steady-state results as a fallout of a transient

analysis, these results become limited by the same assumptions and restrictions. Third, it is common in the literature to perform transient and steady-state analyses of different adaptive filters separately by studying each nonlinear update form separately. Such distinct treatments generally obscure commonalities that exist among algorithms.

These points motivate the development in this paper of a unified approach to the steady-state performance of a large class of adaptive filters that bypasses several of the difficulties encountered in obtaining steady-state results as the limiting case of a transient analysis. The approach is based on studying the energy flow through each iteration of an adaptive filter [8]–[10], and it relies on a fundamental error variance relation that avoids the weight-error variance recursion altogether. This point of view has at least three merits. First, a steady-state analysis in its own right can complement an existing transient analysis. For instance, steady-state results can sometimes be obtained under weaker assumptions than those required to determine the steady-state behavior as a limiting case of the transient analysis. Thus, a steady-state analysis can be useful even when a transient analysis is available. Second, for algorithms for which there is limited transient analysis (due to excessive mathematical complexity, for example), having information about the algorithm's steady-state behavior is better than having limited or no information at all. Third, the proposed approach allows for a unified treatment of a large class of algorithms.

We may remark that although we focus in this paper on the steady-state performance of adaptive filters, the same approach can also be used to study the transient (i.e., convergence and stability) behavior of such filters. These details will be provided elsewhere.

A. Notation

Small boldface letters are used to denote vectors, and capital boldface letters are used to denote matrices, e.g., $\mathbf{w}$ and $\mathbf{C}$. In addition, the symbol "$*$" denotes Hermitian conjugation (complex conjugation for scalars). The symbol $\mathbf{I}$ denotes the identity matrix of appropriate dimensions, and the boldface letter $\mathbf{0}$ denotes either a zero vector or a zero matrix. The notation $\|\mathbf{x}\|$ denotes the Euclidean norm of a vector. All vectors are *column vectors*, except for a *single* vector, namely, the input data vector denoted by $\mathbf{u}_i$, which is taken to be a row vector for convenience of notation. The time instant is placed as a subscript for vectors and between parentheses for scalars, e.g., $\mathbf{w}_i$ and $v(i)$.

B. Problem Formulation

Consider noisy measurements $\{d(i)\}$ that arise from the linear model

$$d(i) = \mathbf{u}_i \mathbf{w}^o + v(i) \tag{1}$$

where

$\mathbf{w}^o$ unknown *column* vector we wish to estimate;
$v(i)$ accounts for measurement noise and modeling errors;
$\mathbf{u}_i$ *row* input (regressor) vector.

Both $\mathbf{u}_i$ and $v(i)$ are stochastic in nature. Many adaptive schemes have been developed in the literature for the estimation of $\mathbf{w}^o$ in different contexts (e.g., echo cancellation, system identification, blind and nonblind channel equalization). In this paper, we focus on the following general class of algorithms

$$\mathbf{w}_{i+1} = \mathbf{w}_i + \mu \mathbf{u}_i^* f_e(i) \tag{2}$$

where

$\mathbf{w}_i$ estimate for $\mathbf{w}^o$ at iteration i;
μ step-size;
$f_e(i)$ generic scalar function of the quantities $\{\mathbf{u}_i, \mathbf{w}_i, d(i)\}$.

Usually, $f_e(i)$ is a function of the so-called output estimation error, which is defined by

$$e(i) = d(i) - \mathbf{u}_i \mathbf{w}_i. \tag{3}$$

Different choices for $f_e(i)$ result in different adaptive algorithms. For example, Table I defines $f_e(i)$ for many famous special cases of (2) for both blind and nonblind modes of adaptation[1]. In the table, $0 \leq \delta \leq 1, R_1,$ and R_2 are positive constants, and $y(i) = \mathbf{u}_i \mathbf{w}_i$ is the adaptive filter output.

An important performance measure for an adaptive filter is its steady-state mean-square-error (MSE), which is defined as

$$\mathsf{MSE} = \lim_{i \to \infty} E(|e(i)|^2) = \lim_{i \to \infty} E(|v(i) + \mathbf{u}_i \tilde{\mathbf{w}}_i|^2)$$

where $\tilde{\mathbf{w}}_i = \mathbf{w}^o - \mathbf{w}_i$ denotes the weight error vector. Under the often realistic assumption that (see, e.g., [1]–[7])

    <u>A.1</u> The noise sequence $\{v(i)\}$ is iid and statistically independent of the regressor sequence $\mathbf{u}_i$

we find that the MSE is equivalently given by

$$\mathsf{MSE} = \sigma_v^2 + \lim_{i \to \infty} E(|\mathbf{u}_i \tilde{\mathbf{w}}_i|^2). \tag{4}$$

Now, the conventional way for evaluating (4), and which dominates most (if not all) derivations in the literature, is the following. First, one assumes, in addition to A.1, that the regression vector $\mathbf{u}_i$ is independent of $\tilde{\mathbf{w}}_i$. Then, the above MSE becomes

$$\mathsf{MSE} = \sigma_v^2 + \lim_{i \to \infty} \mathrm{Tr}(\mathbf{R} \mathbf{C}_i) \tag{5}$$

where $\mathbf{C}_i$ denotes the weight error covariance matrix, $\mathbf{C}_i = E(\tilde{\mathbf{w}}_i \tilde{\mathbf{w}}_i^*)$, and $\mathbf{R} = E(\mathbf{u}_i^* \mathbf{u}_i)$ is the input covariance matrix. As is evident from (5), this method of computation requires the determination of the steady-state value of $\mathbf{C}_i$, say, $\mathbf{C}_\infty$. Studying

[1]The list in the table assumes real-valued data. For complex-valued data, we replace e^3 by $e|e|^2$ and define $\mathrm{sign}[a + jb]$ by $(1/\sqrt{2})(\mathrm{sign}[a] + j\,\mathrm{sign}[b])$.

TABLE I
EXAMPLES FOR $f_e(i)$

ALGORITHM	$f_e(i)$
LMS	$e(i)$
NLMS	$e(i)/\|\mathbf{u}_i\|^2$
LMF	$e^3(i)$
LMMN	$\delta e(i) + (1 - \delta)e^3(i)$
SA	$\mathrm{sign}[e(i)]$
CMA1-2	$[R_1 \frac{y(i)}{|y(i)|} - y(i)]$
CMA2-2	$y(i)[R_2 - |y(i)|^2]$

$\mathbf{C}_i$ (which involves performing a transient analysis) and finding $\mathbf{C}_\infty$ can be a burden, especially for adaptive schemes with nonlinear update equations, which is the case for most of the algorithms listed in Table I. This explains why the steady-state analysis of these algorithms in the literature is more advanced in some cases than in others. It also explains why such analyses have often been carried out separately for each individual algorithm and under varied conditions of operations.

Thus, it would be useful to develop a framework that can handle a variety of algorithms in a unified manner and that can bypass several of the difficulties encountered in obtaining steady-state results as the limiting case of a transient analysis. The approach in this paper is a step in this direction. It is based on studying the energy flow through an adaptive filter, and it relies on a certain fundamental energy conservation relation originally developed in [8]–[10] in the context of robust analysis of adaptive filters.

The paper is organized as follows. In the next section, the energy relation is derived for a general class of adaptive algorithms. In Section III, this relation is used to derive expressions for the steady-state MSE of various algorithms. In Section IV, the arguments are extended to the nonstationary case. In addition, expressions for certain optimum parameter values that optimize the tracking performance of the algorithms are provided. In addition, a comparison is performed between the tracking abilities of several algorithms for various nonstationary environments. Conclusions of the paper are given in Section V. Several simulation results are included to demonstrate the theoretical results.

II. FUNDAMENTAL ENERGY RELATION

We start by noting that with any adaptive scheme of the form (2), we can associate the following so-called *a priori* and *a posteriori* estimation errors

$$e_a(i) = \mathbf{u}_i \tilde{\mathbf{w}}_i, \quad e_p(i) = \mathbf{u}_i \tilde{\mathbf{w}}_{i+1}.$$

Using the data model (1) and the definition (3), it is easy to see that the errors $\{e(i), e_a(i)\}$ are related via $e(i) = e_a(i) + v(i)$. If we further subtract $\mathbf{w}^o$ from both sides of (2) and multiply by $\mathbf{u}_i$ from the left, we also find that the three errors $\{e_p(i), e_a(i), e(i)\}$ are related via

$$\boxed{e_p(i) = e_a(i) - \mu\|\mathbf{u}_i\|^2 f_e(i)}. \tag{6}$$

Substituting (6) into (2), we obtain, for nonzero $\mathbf{u}_i$, the update relation

$$\tilde{\mathbf{w}}_{i+1} = \tilde{\mathbf{w}}_i - \frac{1}{\|\mathbf{u}_i\|^2}\mathbf{u}_i^*[e_a(i) - e_p(i)]. \tag{7}$$

By evaluating the energies of both sides of this equation, we obtain

$$\|\tilde{\mathbf{w}}_{i+1}\|^2 + \frac{1}{\|\mathbf{u}_i\|^2}|e_a(i)|^2 = \|\tilde{\mathbf{w}}_i\|^2 + \frac{1}{\|\mathbf{u}_i\|^2}|e_p(i)|^2. \tag{8}$$

When $\mathbf{u}_i = 0$, it is obviously true that

$$\|\tilde{\mathbf{w}}_{i+1}\|^2 = \|\tilde{\mathbf{w}}_i\|^2. \tag{9}$$

Both results (8) and (9) can be grouped together into a signle equation by defining

$$\bar{\mu}(i) = (\|\mathbf{u}_i\|^2)^{\dagger}$$

in terms of the pseudo-inverse of a scalar so that we obtain

$$\boxed{\|\tilde{\mathbf{w}}_{i+1}\|^2 + \bar{\mu}(i)|e_a(i)|^2 = \|\tilde{\mathbf{w}}_i\|^2 + \bar{\mu}(i)|e_p(i)|^2} . \tag{10}$$

This energy conservation relation holds for <u>all</u> adaptive algorithms whose recursions are of the form given by (2). *No approximations or assumptions are needed to establish* (10); it is an exact relation that shows how the energies of the weight error vectors at two successive time instants are related to the energies of the *a priori* and *a posteriori* estimation errors. Note also that (10) holds for a colored noise sequence $\{v(i)\}$. However, we will continue to focus on the case of white noise in the sequel.

A. Relevance to Steady-State Performance Analysis

Relation (10) has several ramifications. It was derived in [8]–[10] and used to study the robustness and l_2-stability of adaptive filters. Here, we will use it to perform steady-state and tracking analyses of such filters. Further applications of the energy relation to the study of blind adaptive equalizers can be found in [11] and [12].

Thus, note first that in steady state (i.e., as $i \to \infty$), we can assume that

$$E(\|\tilde{\mathbf{w}}_{i+1}\|^2) = E(\|\tilde{\mathbf{w}}_i\|^2). \tag{11}$$

This assumption is equivalent to assuming that the mean square deviation (MSD) converges to a steady-state value. This is a justifiable assumption since our aim is to study the performance of adaptive algorithms in steady state, i.e., after steady state is reached.[2] Now, observe that by using (11), and because of the energy-preserving relation (10), the effect of the weight-error vector is canceled out. By taking expectations of both sides of (10), we then get

$$E(\bar{\mu}(i)|e_a(i)|^2) = E(\bar{\mu}(i)|e_p(i)|^2).$$

[2]We may mention that by averaging analysis, and under some conditions, one can generally guarantee that there exists a small enough μ for which the filter reaches steady state (see, e.g., [13] and [14])—we do not expand on this stability issue here since the objective of this paper is to evaluate filter performance once steady state is reached.

Using (6), the above collapses to the following fundamental error variance relation in terms of $\{e_a(i), v(i)\}$ only (recall that $e(i) = e_a(i) + v(i)$):

$$\boxed{E(\bar{\mu}(i)|e_a(i)|^2) = E\left(\bar{\mu}(i)\left|e_a(i) - \frac{\mu}{\bar{\mu}(i)}f_e(i)\right|^2\right)} . \tag{12}$$

This equation can now be solved for the steady-state excess mean-square-error (EMSE), which is defined by

$$\zeta \triangleq \lim_{i\to\infty} E(|e_a(i)|^2).$$

Observe from (4) that the desired MSE is given by MSE $= \sigma_v^2 + \zeta$ so that finding ζ is equivalent to finding the MSE.

We emphasize again that (12) is an *exact* relation that holds without any approximations or assumptions, except for the assumption that the filter is in steady state. The procedure of finding the EMSE through (12) avoids the need for evaluating $E(\|\tilde{\mathbf{w}}_i\|^2)$ or its steady-state value $E(\|\tilde{\mathbf{w}}_\infty\|^2)$.

III. STEADY-STATE ANALYSIS

We now apply the above general procedure to various adaptive algorithms from Table I. Due to space limitations, we omit some trivial details and only highlight the main steps in the arguments. The reader will soon realize the convenience of working with (12).

A. LMS *Algorithm*

For LMS, we have $f_e(i) = e(i) = e_a(i) + v(i)$. Substituting into (12) and using the noise assumption A.1, it follows immediately that

$$2\mu\zeta^{\mathsf{LMS}} = \mu^2 E(\|\mathbf{u}_i\|^2|e_a(i)|^2) + \mu^2\sigma_v^2\,\mathrm{Tr}(\mathbf{R}). \tag{13}$$

To solve for ζ^{LMS}, we consider three cases.

1) For sufficiently small μ, we can assume that the term $\mu^2 E(\|\mathbf{u}_i\|^2|e_a(i)|^2)$ is negligible relative to the second term on the right-hand side of (13) so that

$$\boxed{\zeta^{\mathsf{LMS}} = \frac{\mu}{2}\sigma_v^2\,\mathrm{Tr}(\mathbf{R})} \quad (\text{small } \mu). \tag{14}$$

This is the same result obtained in [15] for small values of μ, but here, it is obtained more immediately.

2) For larger values of μ for which we cannot neglect the second term on the right-hand side of (13), we solve (13) by imposing the following assumption:[3]

   <u>A.2</u> At steady state, $\mu^2\|\mathbf{u}_i\|^2$ is statistically independent of $|e_a(i)|^2$.

   This assumption is realistic for long tapped-delay line filters[4]. Furthermore, it becomes *exact* for the case of con-

[3]By larger values of μ, we do not necessarily mean a large μ but, rather, step sizes that are not infinitesimally small and still guarantee filter stability.

[4]To guarantee convergence, the algorithm step-size μ is usually chosen to be inversely proportional to the filter length M [1]. Using the law of large numbers, $\|\mathbf{u}_i\|^2$ could be considered to be a random variable of variance proportional to M. Thus, $\mu^2\|\mathbf{u}_i\|^2$ has variance proportional to $1/M$, which decays with the filter length M. This means that for long-enough filters, the variations in $\mu^2\|\mathbf{u}_i\|^2$ are very small, and this term can be considered to be independent of $e_a(i)$.

stant modulus data that arises in some adaptive filtering applications (see, e.g., [16]). Using A.2, and (13), we directly obtain

$$\zeta^{\mathsf{LMS}} = \frac{\mu \sigma_v^2 \operatorname{Tr}(\mathbf{R})}{2 - \mu \operatorname{Tr}(\mathbf{R})} \quad \text{large } \mu. \qquad (15)$$

This is also a well-known result (see [1] and [3]) but is obtained here rather more directly and in a different manner.

3) For Gaussian white-input signals (with $\mathbf{R} = \sigma_u^2 \mathbf{I}$), (13) can be more accurately solved if one resorts to the widely used independence assumption [3]

    A.3 At steady state, $\tilde{\mathbf{w}}_i$ is statistically independent of $\mathbf{u}_i$.

In this case, it can be verified that

$$E(\|\mathbf{u}_i\|^2 |e_a(i)|^2) = (M + \lambda)\sigma_u^2 E(|e_a(i)|^2)$$

where M is the filter length, $\lambda = 1$ if the $\{\mathbf{u}_i\}$ are complex valued, and $\lambda = 2$ if the $\{\mathbf{u}_i\}$ are real valued. This leads to the well-known result

$$\zeta^{\mathsf{LMS}} = \frac{\mu \sigma_v^2 \operatorname{Tr}(\mathbf{R})}{2 - \mu(M + \lambda)\sigma_u^2} \quad \text{(Gaussian).} \qquad (16)$$

B. NLMS Algorithm

For the normalized LMS algorithm, $f_e(i) = e(i)/\|\mathbf{u}_i\|^2$. In this case, (12) and assumption A.1 lead to the equality

$$(2 - \mu)E\left(\frac{|e_a(i)|^2}{\|\mathbf{u}_i\|^2}\right) = \mu \sigma_v^2 E\left(\frac{1}{\|\mathbf{u}_i\|^2}\right). \qquad (17)$$

Again, this is an exact equality. We consider two cases.

1) Under assumption A.2, we have

$$E\left(\frac{|e_a(i)|^2}{\mu^2 \|\mathbf{u}_i\|^2}\right) = E\left(|e_a(i)|^2\right) \cdot E\left(\frac{1}{\mu^2 \|\mathbf{u}_i\|^2}\right)$$

so that (17) leads to the expression

$$\zeta^{\mathsf{NLMS}} = \frac{\mu \sigma_v^2}{(2 - \mu)}. \qquad (18)$$

This result becomes *exact* for constant modulus data. In addition, observe that it is independent of $\mathbf{R}$.

2) In some works (see, e.g., [3, p. 443]), the following approximation is sometimes called upon:

$$E\left(\frac{|e_a(i)|^2}{\|\mathbf{u}_i\|^2}\right) \approx \frac{E\left(|e_a(i)|^2\right)}{E\left(\|\mathbf{u}_i\|^2\right)}$$

in which case, (17) leads directly to

$$\zeta^{\mathsf{NLMS}} = \frac{\mu \sigma_v^2}{(2 - \mu)} E\left(\frac{1}{\|\mathbf{u}_i\|^2}\right) \operatorname{Tr}(\mathbf{R}). \qquad (19)$$

This is the same expression obtained in [17] in a different and less direct way.

C. Sign Algorithm

For the sign algorithm (SA), we have $f_e(i) = \operatorname{sign}[e(i)]$. In this case, (12) leads to the equality

$$E(e_a(i) \operatorname{sign}(e_a(i) + v(i))) = \frac{\mu}{2} \operatorname{Tr}(\mathbf{R}). \qquad (20)$$

By assuming that $e(i)$ and $v(i)$ are real-valued jointly Gaussian in steady state (as used in [18] and [19]), and by using A.1 and Price's theorem[5] [20], we obtain

$$E(e_a(i) \operatorname{sign}(e_a(i) + v(i))) = \sqrt{\frac{2}{\pi}} \frac{E(|e_a(i)|^2)}{\sqrt{\sigma_v^2 + E(|e_a(i)|^2)}}.$$

Substituting into (20) and solving for $E(|e_a(i)|^2)$, we find that

$$\zeta^{\mathsf{SA}} = \frac{\alpha}{2} \cdot \left(\alpha + \sqrt{\alpha^2 + 4\sigma_v^2}\right) \qquad (21)$$

where $\alpha = \sqrt{(\pi/8)}\mu \operatorname{Tr}(\mathbf{R})$. Expression (21) is the same result that was obtained in [18] by using the independence assumptions in addition to the above Gaussianity assumption. Here, the result is obtained without resorting directly to the independence assumptions—see [21] for more details.

D. LMF and LMMN Algorithms

For the least-mean mixed-norm (LMMN) algorithm with real-valued data, we have $f_e(i) = \delta e(i) + (1 - \delta)e^3(i)$ [22] (the case of complex-valued data is considered further ahead toward the end of this section). The least-mean fourth (LMF) algorithm corresponds to the special case $\delta = 0$ [23]. Introduce, for compactness of notation

$$\bar{\delta} = 1 - \delta, \quad E(|v(i)|^4) = \xi_4^v \quad E(|v(i)|^6) = \xi_6^v.$$

By making the reasonable assumption that in steady state $|e_a(i)|^2 \ll |v(i)|^2$ (see, e.g., [24]) and by using A.1, the energy equation (12) implies that

$$2\mu b \zeta^{\mathsf{LMMN}} = \mu^2 a \operatorname{Tr}(\mathbf{R}) + \mu^2 c E(\|\mathbf{u}_i\|^2 |e_a(i)|^2) \qquad (22)$$

where we introduced the constants

$$a = \delta^2 \sigma_v^2 + 2\delta\bar{\delta}\xi_4^v + \bar{\delta}^2 \xi_6^v \qquad (23)$$
$$b = \delta + 3\bar{\delta}\sigma_v^2 \qquad (24)$$
$$c = \delta^2 + 12\delta\bar{\delta}\sigma_v^2 + 15\bar{\delta}\xi_4^v. \qquad (25)$$

We again consider three cases.

1) For values of μ that are small enough so that the term $\mu^2 E(\|\mathbf{u}_i\|^2 |e_a(i)|^2)$ could be ignored, we obtain

$$\zeta^{\mathsf{LMMN}} = \frac{\mu a}{2b} \operatorname{Tr}(\mathbf{R}) \quad \text{(small } \mu\text{).} \qquad (26)$$

In [24], the same result was obtained for vanishingly small μ by using averaging analysis and the ODE method (see, e.g., [14]). For $\delta = 0$, the above expression collapses to

$$\zeta^{\mathsf{LMF}} = \frac{\mu}{2}\left(\frac{\xi_6^v}{3\sigma_v^2}\right) \operatorname{Tr}(\mathbf{R}) \quad \text{(small } \mu\text{)} \qquad (27)$$

---

[5]For two jointly Gaussian real-valued random variables x and y, we have $E(x \operatorname{sign}(y)) = \sqrt{2/\pi} \cdot 1/\sigma_y E(xy)$.

which is the same expression obtained in [23] by using the independence assumptions.

2) For larger values of μ, and using A.2 again, we get the following new expressions for the EMSE:

$$\zeta^{\mathsf{LMMN}} = \frac{\mu a \operatorname{Tr}(\mathbf{R})}{2b - \mu c \operatorname{Tr}(\mathbf{R})} \qquad \text{(large } \mu) \qquad (28)$$

$$\zeta^{\mathsf{LMF}} = \frac{\mu \xi_6^v \operatorname{Tr}(\mathbf{R})}{6\sigma_v^2 - 15\mu \xi_4^v \operatorname{Tr}(\mathbf{R})} \qquad \text{(large } \mu). \qquad (29)$$

Fig. 1 compares the theoretical MSE obtained from (26) and (28) with the experimental MSE. In the simulations, the unknown system weight vector $\mathbf{w}^o$ is of length 10, the input $\mathbf{u}_i$ is Gaussian of unit variance, and $\delta = 0.5$. The noise is chosen to be a linear combination of normally and uniformly distributed independent random variables of variances $\sigma_n^2 = 10^{-6}$ and $\sigma_c^2 = 10^{-4}/12$, respectively. Each simulation result is the steady-state statistical average of 100 runs, with 10^5 iterations in each run. The figure shows that both expressions are in good match with simulation results at small values of μ. However, (28) provides a better match with the simulation results for relatively larger values of μ, which validates the use of assumption A.2.

3) For Gaussian white-input signals $(\mathbf{R} = \sigma_u^2 \mathbf{I})$, (22) can be solved by imposing A.3 to yield

$$\zeta^{\mathsf{LMMN}} = \frac{\mu M \sigma_u^2 a}{2b - \mu(M+2)\sigma_u^2 c} \qquad \text{(Gaussian)} \qquad (30)$$

$$\zeta^{\mathsf{LMF}} = \frac{\mu M \sigma_u^2 \xi_6^v}{6\sigma_v^2 - 15\mu(M+2)\sigma_u^2 \xi_4^v} \qquad \text{(Gaussian)}. \qquad (31)$$

For the case of complex-valued data, we replace e^3 by $e|e|^2$ and assume the noise is circular, i.e., $E(v^2(i)) = 0$. Then, repeating the above arguments, we find that the three expressions (26), (28), and (30) are still valid but with b and c replaced by

$$b' = \delta + 2\bar{\delta}\sigma_v^2, \quad c' = \delta^2 + 8\delta\bar{\delta}\sigma_v^2 + 9\bar{\delta}\xi_4^v. \qquad (32)$$

Corresponding expressions for the LMF algorithm can be obtained by setting $\delta = 0$.

E. CM Algorithms

Similar analyses can be carried out for constant modulus (CM) algorithms. The details are provided in [11]. Here, we only briefly comment on one particular case for the sake of illustration. Assume $v(i) = 0$ (and, hence, $e_a(i) = e(i)$), and define

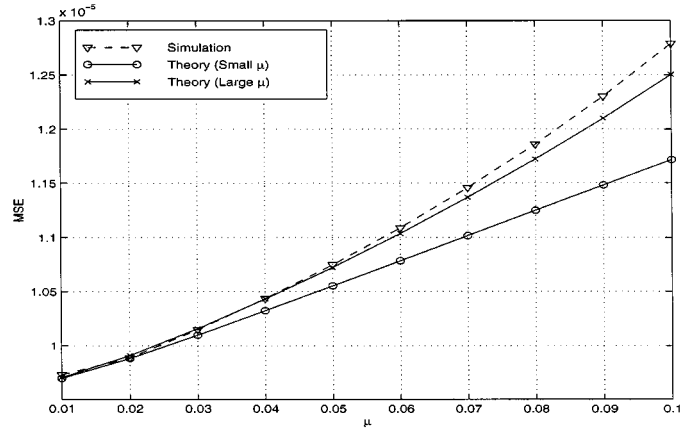$$\sigma_d^2 = E(|d(i)|^2), \quad \xi_4^d = E(|d(i)|^4), \quad \xi_6^d = E(|d(i)|^6).$$

Fig. 1.   Experimental and theoretical MSE versus μ for LMMN.

Let $R_2 = \xi_4^d/\sigma_d^2$, and assume also that all data are real-valued (the complex case is also studied in [11]). Define further, for compactness of notation, $z(i) = y(i)(R_2 - |y(i)|^2)$. Then, (12) yields, for CMA2-2

$$2\mu E(e_a(i)z(i)) = \mu^2 E(\|\mathbf{u}_i\|^2 |z(i)|^2). \qquad (33)$$

To solve this equation for $E|e_a(i)|^2$, we make the following reasonable (and common) assumption—for more motivation and explanation on this assumption, see [11] and [25]:

A.4 The signals $d(i)$ and $e_a(i)$ are independent in steady state so that $E(d(i)e_a(i)) = 0$ since the signal $d(i)$ is assumed zero mean.

Assumptions A.2 and A.4 yield for small enough μ

$$\zeta^{\mathsf{CMA2-2}} = \mu \left(\frac{\sigma_d^2 R_2^2 - 2R_2 \xi_4^d + \xi_6^d}{2(3\sigma_d^2 - R_2)} \right) \operatorname{Tr}(\mathbf{R})$$

$$\text{(small } \mu).$$

This expression is slightly different from the one obtained in [25]; it was shown in [11] that it leads to a better approximation for the MSE.

Table II summarizes the derived expressions for the steady-state EMSE for several of the algorithms of Table I.

IV. TRACKING ANALYSIS

In a nonstationary environment, the data $\{d(i)\}$ is assumed to arise from a linear model of the form $d(i) = \mathbf{u}_i \mathbf{w}_i^o + v(i)$, where the unknown system $\mathbf{w}_i^o$ is now time variant. It is often assumed that the variation in $\mathbf{w}_i^o$ is according to the model $\mathbf{w}_{i+1}^o = \mathbf{w}_i^o + \mathbf{q}_i$, where $\mathbf{q}_i$ denotes the random perturbation (see, e.g., [1], [3], and [19]).[6] The purpose of the tracking analysis of an adaptive filter is to study its ability to track such time variations.

We now show how to evaluate the tracking performance of an adaptive algorithm by the same feedback method proposed in this paper. For this purpose, we first redefine the weight error

---

[6]The approach of this paper can be applied to a more general model for $\mathbf{w}_i^o$, which takes into account colored system variations and carrier offsets. Details will be provided elsewhere. Preliminary results appear in [26].

vector as $\tilde{\mathbf{w}}_i = \mathbf{w}_i^o - \mathbf{w}_i$ and the *a posteriori* estimation error as $e_p(i) = \mathbf{u}_i(\tilde{\mathbf{w}}_{i+1} - \mathbf{q}_i)$. Then, $\tilde{\mathbf{w}}_i$ satisfies

$$\tilde{\mathbf{w}}_{i+1} = \tilde{\mathbf{w}}_i + \mu(i)\mathbf{u}_i^* f_e(i) + \mathbf{q}_i. \tag{34}$$

If we further multiply (34) by $\mathbf{u}_i$ from the left, we obtain that (6) and (7) still hold for the nonstationary case, whereas (10) becomes

$$\boxed{\|\tilde{\mathbf{w}}_{i+1} - \mathbf{q}_i\|^2 + \bar{\mu}(i)|e_a(i)|^2 = \|\tilde{\mathbf{w}}_i\|^2 + \bar{\mu}(i)|e_p(i)|^2}. \tag{35}$$

For mathematical tractability of the tracking analysis, we impose the following assumption, which is typical in the context of tracking analysis of adaptive filters (see, e.g., [19]).

> A.5 The sequence $\{\mathbf{q}_i\}$ is a stationary sequence of independent zero-mean vectors and positive definite autocorrelation matrix $\mathbf{Q} = E(\mathbf{q}_i\mathbf{q}_i^*)$, which is mutually independent of the sequences $\{\mathbf{u}_i\}$ and $\{v(i)\}$.

Using (6), (34), and A.5, it is straightforward to verify that the variance relation (12) should now be replaced by (36), shown at the bottom of the page. Comparing the above with (12), we see that evaluating the nonstationary EMSE is simply a straightforward extension of evaluating the stationary EMSE. The only addition is the steady-state contribution by the system nonstationarity, which is equal to $\text{Tr}(\mathbf{Q})$.

This is a useful observation in the context of the tracking analysis of adaptive algorithms since it allows us to arrive at tracking results almost by inspection from the stationary case results. In the literature, both cases have usually been studied separately. We will now show how to use (36) to solve for the nonstationary EMSE for the algorithms given in Table I.

The results for LMS and NLMS can be obtained in a straightforward manner, just by extending the arguments given in the stationary case. Hence, we will only state the resulting expressions in these two cases. Moreover, for space considerations, we omit the tracking analysis of the CM algorithms and refer instead to the related work [27]; we only reproduce the result of that paper here. For these reasons, in the sequel, we focus on the SA, LMMN, and LMF algorithms. The final expressions for the MSE in the nonstationary case for all algorithms are summarized in Tables III and IV; the latter contains expressions for the optimal parameters that result in the smallest MSE.

A. Sign Algorithm

Comparing (12) and (36) and using (22), we obtain

$$E(e_a(i)\text{sign}(e_a(i) + v(i))) = \mu^{-1}\text{Tr}(\mathbf{Q}) + \frac{\mu}{2}\text{Tr}(\mathbf{R}). \tag{37}$$

TABLE II
EXPRESSIONS FOR THE EMSE FOR DIFFERENT ADAPTIVE ALGORITHMS IN A STATIONARY ENVIRONMENT

ALGORITHM	EMSE
LMS (small μ)	$\frac{\mu}{2}\sigma_v^2\,\text{Tr}(\mathbf{R})$
LMS (large μ)	$\frac{\mu\sigma_v^2\,\text{Tr}(\mathbf{R})}{2-\mu\,\text{Tr}(\mathbf{R})}$
NLMS	$\frac{\mu\sigma_v^2}{(2-\mu)},\quad \frac{\mu\sigma_v^2}{(2-\mu)}\,\text{E}\left(\frac{1}{\|\mathbf{u}_i\|^2}\right)\text{Tr}(\mathbf{R})$
LMF (real, small μ)	$\frac{\mu}{2}\left(\frac{\xi_6^v}{3\sigma_v^2}\right)\text{Tr}(\mathbf{R})$
LMF (real, large μ)	$\frac{\mu\xi_6^v\,\text{Tr}(\mathbf{R})}{6\sigma_v^2-15\mu\xi_4^v\,\text{Tr}(\mathbf{R})}$
LMF (complex, small μ)	$\frac{\mu}{2}\left(\frac{\xi_6^v}{2\sigma_v^2}\right)\text{Tr}(\mathbf{R})$
LMF (complex, large μ)	$\frac{\mu\xi_6^v\,\text{Tr}(\mathbf{R})}{4\sigma_v^2-9\mu\xi_4^v\,\text{Tr}(\mathbf{R})}$
LMMN (real, small μ)	$\frac{\mu a}{2b}\,\text{Tr}(\mathbf{R})$
LMMN (real, large μ)	$\frac{\mu a}{2b-\mu c\,\text{Tr}(\mathbf{R})}\,\text{Tr}(\mathbf{R})$
LMMN (complex, small μ)	$\frac{\mu a}{2b'}\,\text{Tr}(\mathbf{R})$
LMMN (complex, large μ)	$\frac{\mu a}{2b'-\mu c'\,\text{Tr}(\mathbf{R})}\,\text{Tr}(\mathbf{R})$
SA	$\frac{\alpha}{2}\left(\alpha + \sqrt{\alpha^2 + 4\sigma_v^2}\right),\quad \alpha = \sqrt{\frac{\pi}{8}}\mu\,\text{Tr}(\mathbf{R})$
CMA2-2 (real, small μ)	$\mu\left(\frac{\sigma_d^2 R_2^2 - 2R_2\xi_4^d + \xi_6^d}{2(3\sigma_d^2 - R_2)}\right)\text{Tr}(\mathbf{R})$
CMA2-2 (complex, small μ)	$\mu\left(\frac{\sigma_d^2 R_2^2 - 2R_2\xi_4^d + \xi_6^d}{2(2\sigma_d^2 - R_2)}\right)\text{Tr}(\mathbf{R})$

TABLE III
EXPRESSIONS FOR THE EMSE IN A NONSTATIONARY ENVIRONMENT FOR DIFFERENT ALGORITHMS

ALGORITHM	EMSE
LMS (small μ)	$\frac{1}{2}\mu^{-1}\text{Tr}(\mathbf{Q}) + \frac{1}{2}\mu\sigma_v^2\,\text{Tr}(\mathbf{R})$
LMS (large μ)	$\frac{\mu^{-1}\text{Tr}(\mathbf{Q})+\mu\sigma_v^2\,\text{Tr}(\mathbf{R})}{2-\mu\,\text{Tr}(\mathbf{R})}$
NLMS	$\frac{\mu^{-1}\text{Tr}(\mathbf{Q})+\mu\sigma_v^2}{(2-\mu)},\quad \frac{\mu^{-1}\text{Tr}(\mathbf{Q})+\mu\sigma_v^2\,\text{E}\left(\frac{1}{\|\mathbf{u}_i\|^2}\right)\text{Tr}(\mathbf{R})}{(2-\mu)}$
LMF (small μ)	$\frac{\mu^{-1}\text{Tr}(\mathbf{Q})+\mu\xi_6^v\,\text{Tr}(\mathbf{R})}{6\sigma_v^2}$
LMF (large μ)	$\frac{\mu^{-1}\text{Tr}(\mathbf{Q})+\mu\xi_6^v\,\text{Tr}(\mathbf{R})}{6\sigma_v^2-15\mu\,\text{Tr}(\mathbf{R})\xi_4^v}$
LMMN (small μ)	$\frac{\mu^{-1}\text{Tr}(\mathbf{Q})+\mu a\,\text{Tr}(\mathbf{R})}{2b}$
LMMN (large μ)	$\frac{\mu^{-1}\text{Tr}(\mathbf{Q})+\mu a\,\text{Tr}(\mathbf{R})}{2b-\mu c\,\text{Tr}(\mathbf{R})}$
SA	$\frac{\alpha}{2}\left(\alpha + \sqrt{\alpha^2 + 4\sigma_v^2}\right),\quad \alpha = \sqrt{\frac{\pi}{8}}\left(\mu^{-1}\text{Tr}(\mathbf{Q}) + \mu\,\text{Tr}(\mathbf{R})\right)$
CMA2-2 (small μ)	$\frac{\mu^{-1}\text{Tr}(\mathbf{Q})+\mu(\sigma_d^2 R_2^2 - 2R_2\xi_4^d + \xi_6^d)\,\text{Tr}(\mathbf{R})}{2(3\sigma_d^2 - R_2)}$

TABLE IV
OPTIMUM ALGORITHM PARAMETERS AND MINIMUM EMSE IN NONSTATIONARY ENVIRONMENTS

ALGORITHM	μ_o	ζ_{min}
LMS	$\frac{1}{\sigma_v}\sqrt{\frac{\text{Tr}(\mathbf{Q})}{\text{Tr}(\mathbf{R})}}$	$\sigma_v\sqrt{\text{Tr}(\mathbf{Q})\,\text{Tr}(\mathbf{R})}$
NLMS	$\frac{1}{\sigma_v}\sqrt{\frac{\text{Tr}(\mathbf{Q})}{\text{E}\left(\frac{1}{\|\mathbf{u}_i\|^2}\right)\text{Tr}(\mathbf{R})}}$	$\sigma_v\sqrt{\text{Tr}(\mathbf{Q})\,\text{E}\left(\frac{1}{\|\mathbf{u}_i\|^2}\right)\text{Tr}(\mathbf{R})}$
LMF (small μ)	$\sqrt{\frac{\text{Tr}(\mathbf{Q})}{\text{Tr}(\mathbf{R})\xi_6^v}}$	$\frac{\sqrt{\text{Tr}(\mathbf{Q})\,\text{Tr}(\mathbf{R})\xi_6^v}}{3\sigma_v^2}$
LMMN (small μ)	$\sqrt{\frac{\text{Tr}(\mathbf{Q})}{a_o\,\text{Tr}(\mathbf{R})}}$	$\frac{\sqrt{a_o\,\text{Tr}(\mathbf{Q})\,\text{Tr}(\mathbf{R})}}{b_o}$
SA	$\sqrt{\frac{\text{Tr}(\mathbf{Q})}{\text{Tr}(\mathbf{R})}}$	$\frac{\pi}{4}\text{Tr}(\mathbf{Q})\,\text{Tr}(\mathbf{R}) + [\frac{\pi}{2}\sigma_v^2\,\text{Tr}(\mathbf{Q})\,\text{Tr}(\mathbf{R})$ $+\frac{\pi^2}{16}\text{Tr}^2(\mathbf{Q})\,\text{Tr}^2(\mathbf{R})]^{1/2}$
CMA2-2 (small μ)	$\sqrt{\frac{\text{Tr}(\mathbf{Q})}{(\sigma_d^2 R_2^2 - 2R_2\xi_4^d + \xi_6^d)\,\text{Tr}(\mathbf{R})}}$	$\frac{\sqrt{\text{Tr}(\mathbf{Q})(\sigma_d^2 R_2^2 - 2R_2\xi_4^d + \xi_6^d)\,\text{Tr}(\mathbf{R})}}{3\sigma_d^2 - R_2}$

$$E(\bar{\mu}(i)|e_a(i)|^2) = \text{Tr}(\mathbf{Q}) + E\left(\bar{\mu}(i)\left|e_a(i) - \frac{\mu}{\bar{\mu}(i)}f_e(i)\right|^2\right) \tag{36}$$

Using the same procedure used in the stationary case, it is straightforward to show that the EMSE is still given by (21)

$$\zeta^{\mathsf{SA}} = \frac{\alpha}{2} \cdot \left(\alpha + \sqrt{\alpha^2 + 4\sigma_v^2} \right) \tag{38}$$

where α is now given by

$$\alpha = \sqrt{\frac{\pi}{8}} (\mu^{-1} \operatorname{Tr}(\mathbf{Q}) + \mu \operatorname{Tr}(\mathbf{R})).$$

The optimum step size is then seen to be

$$\mu_o^{\mathsf{SA}} = \sqrt{\frac{\operatorname{Tr}(\mathbf{Q})}{\operatorname{Tr}(\mathbf{R})}}.$$

These are the same results obtained in [28] and [29]—for more details see [21].

B. LMF *and* LMMN *Algorithms*

We focus on real-valued data (the complex case only changes some coefficients, as we saw in the stationary analysis). Comparing (12) and (36) and using (22), we obtain

$$\begin{aligned} 2\mu b \zeta^{\mathsf{LMMN}} \\ = \operatorname{Tr}(\mathbf{Q}) + \mu^2 a \operatorname{Tr}(\mathbf{R}) + \mu^2 c E(\|\mathbf{u}_i\|^2 |e_a(i)|^2). \end{aligned} \tag{39}$$

To solve for ζ^{LMMN}, we consider three cases.

1) For sufficiently small μ, we can assume that the third term on the right-hand side of (39) is negligible with respect to the second term so that

$$\zeta^{\mathsf{LMMN}} = \frac{\mu^{-1} \operatorname{Tr}(\mathbf{Q}) + \mu a \operatorname{Tr}(\mathbf{R})}{2b} \quad \text{(small } \mu\text{)}. \tag{40}$$

At $\delta = 0$, (40) reduces to the EMSE of the LMF algorithm, which is given by

$$\zeta^{\mathsf{LMF}} = \frac{\mu^{-1} \operatorname{Tr}(\mathbf{Q}) + \mu \xi_6^v \operatorname{Tr}(\mathbf{R})}{6\sigma_v^2} \quad \text{(small } \mu\text{)}. \tag{41}$$

2) For larger values of μ, (39) can be solved by imposing A.2 to obtain

$$\zeta^{\mathsf{LMMN}} = \frac{\mu^{-1} \operatorname{Tr}(\mathbf{Q}) + \mu a \operatorname{Tr}(\mathbf{R})}{2b - \mu c \operatorname{Tr}(\mathbf{R})} \quad \text{((large } \mu\text{)} \tag{42}$$

and

$$\zeta^{\mathsf{LMF}} = \frac{\mu^{-1} \operatorname{Tr}(\mathbf{Q}) + \mu \xi_6^v \operatorname{Tr}(\mathbf{R})}{6\sigma_v^2 - 15\mu \operatorname{Tr}(\mathbf{R})\xi_4^v} \quad \text{(large } \mu\text{)}. \tag{43}$$

3) For Gaussian white-input signals ($\mathbf{R} = \sigma_u^2 \mathbf{I}$), imposing A.3, we get

$$\zeta^{\mathsf{LMMN}} = \frac{\mu^{-1} \operatorname{Tr}(\mathbf{Q}) + \mu M \sigma_u^2 a}{2b - \mu(M+2)\sigma_u^2 c} \quad \text{(Gaussian)} \tag{44}$$

$$\zeta^{\mathsf{LMF}} = \frac{\mu^{-1} \operatorname{Tr}(\mathbf{Q}) + \mu M \sigma_u^2 \xi_6^v}{6\sigma_v^2 - 15\mu(M+2)\sigma_u^2 \xi_4^v} \quad \text{(Gaussian)}. \tag{45}$$

Expressions (40)–(45) are new results that describe the ability of the LMF and LMMN algorithms to track system nonstationarities. The following conclusions follow from these results. We can see that the steady-state EMSE for both of the LMF and LMMN algorithms is composed of two terms. The first term decreases with μ and increases with the system nonstationarity variance $\operatorname{Tr}(\mathbf{Q})$. The second term increases with μ and the received signal variance $E(\|\mathbf{u}_i\|^2) = \operatorname{Tr}(\mathbf{R})$. Thus, unlike the stationary case [see (26)–(29)], the steady-state EMSE is not a monotonically increasing function of μ. We can also see that there exists an optimal value of the step size μ_o that minimizes the steady-state MSE in the nonstationary case. This is established in Appendix A.

Fig. 2 shows the theoretical and simulated EMSE versus μ for the optimal value of δ calculated from (53) to be $\delta_o = 0.5432$. Here, we are using a noise sequence that is a mixture of Gaussian and uniform noises with variances $\sigma_n^2 = \sigma_c^2 = 0.1$. Moreover, $\mathbf{Q} = \sigma_q^2 \mathbf{I}$ and $\mathbf{R} = \mathbf{I}$ with $\sigma_q = 10^{-3}$. Fig. 3 shows theoretical and simulated results versus δ for the optimal value of μ calculated from (54) to be $\mu_o = 0.0029$. Both simulations show that optimal parameter values obtained from simulations $\{\delta_o, \mu_o\} = \{0.59, 0.003\}$ are a good match with the values $\{0.5432, 0.0029\}$, given by (53) and (54), respectively.

In Appendix A, we use the above EMSE expressions to investigate the existence of optimum design parameters $\{\delta_o, \mu_o\}$ that minimize the steady-state EMSE of LMF and LMMN, as given by (40) and (41). We also compare the tracking performance of these algorithms with LMS for different noise distributions (Gaussian, uniform, and a mixture of Gaussian and uniform).

V. CONCLUSION

This paper develops an approach for the steady-state analysis of adaptive filters that bypasses the need for considering the limiting case of a transient analysis. One of the main features of the new framework is that its starting point is the fundamental energy (or variance) relation (12) [or (36) in the nonstationary case]. This relation is fundamental in that it is exact, and it holds for any adaptive scheme of the general form (2), irrespective of any approximations. By expanding both sides of the relation, and by imposing certain conditions or assumptions, one obtains an equation in the desired EMSE. This equation is rather trivial to solve when the step size is assumed to be sufficiently small. For larger step sizes, on the other hand, the equation leads to tighter expressions for the EMSE.

We may add that the approach can be extended in a rather straightforward manner to other scenarios as well, such as the study of the performance of adaptive schemes in finite-precision implementations and the study of adaptive filters of the RLS
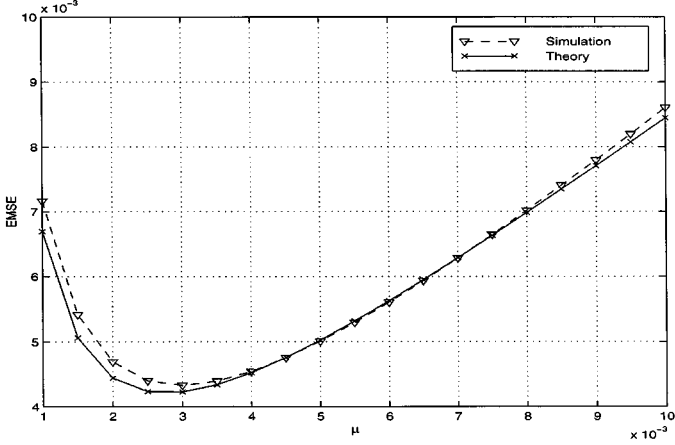
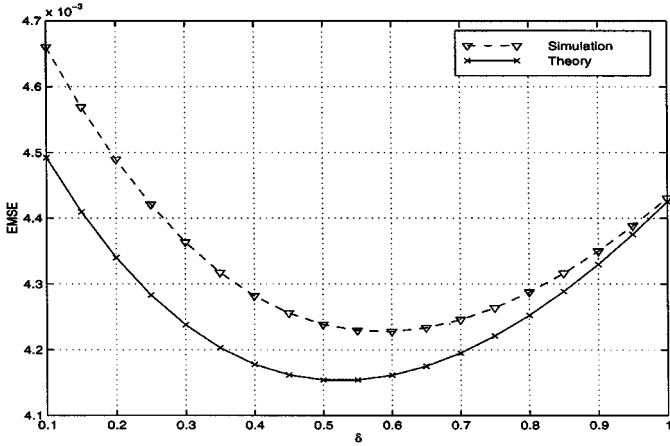Fig. 2.   Experimental and theoretical EMSE curves versus μ for LMMN at $\delta = \delta_o$.

Fig. 3.   Experimental and theoretical EMSE curves versus δ for LMMN at $\mu = \mu_o$.

and Gauss–Newton type by using the energy relation of [30]. We have also used the approach in [11], [12], and [27] to study the steady-state tracking and convergence performance of fractionally spaced blind adaptive schemes.

APPENDIX A
PARAMETER OPTIMIZATION FOR LMMN AND LMF

We explain here how the expressions that were derived in the body of the paper for the EMSE for LMF and LMMN enable us to investigate the existence of optimum design parameters $\{\delta_o, \mu_o\}$ that minimize the steady-state EMSE, as given by (40) and (41). This is done for two cases labeled A (fixed δ) and B.

A. Fixed δ and Optimal μ

If the norm mixing parameter δ is *a priori* chosen to fulfill some convergence properties, then there will always exist an optimum value of μ that minimizes ζ^{LMMN}, which is directly given from (40) by

$$\boxed{\mu_o^{\mathsf{LMMN}} = \sqrt{\mathrm{Tr}(\mathbf{Q})/a\,\mathrm{Tr}(\mathbf{R})}}. \qquad (46)$$
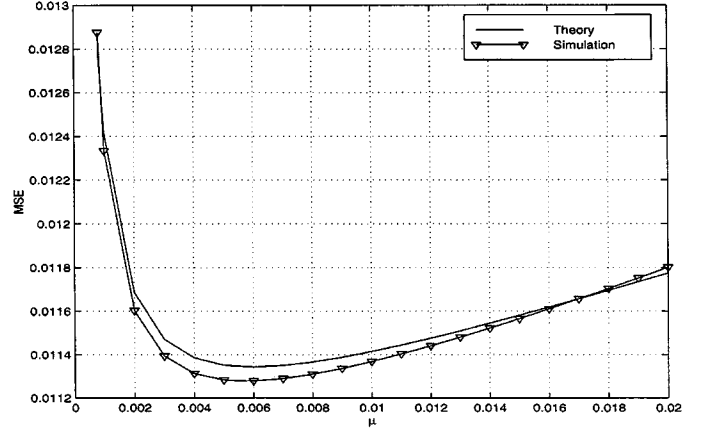

Fig. 4.   Experimental and theoretical MSE versus μ for LMMN in a nonstationary environment.

The corresponding minimum value of the steady-state EMSE is given by

$$\boxed{\zeta_{\min}^{\mathsf{LMMN}} = \frac{\sqrt{a\,\mathrm{Tr}(\mathbf{Q})\,\mathrm{Tr}(\mathbf{R})}}{b}} \qquad (\text{small } \mu). \qquad (47)$$

The LMF algorithm always has a constrained δ that is equal to zero. Therefore, the optimum step-size that minimizes its steady-state EMSE, which is given in (41), and the corresponding minimum steady-state EMSE, are respectively, given by

$$\boxed{\mu_o^{\mathsf{LMF}} = \sqrt{\mathrm{Tr}(\mathbf{Q})/\xi_6^v\,\mathrm{Tr}(\mathbf{R})}} \qquad (48)$$

and

$$\boxed{\zeta_{\min}^{\mathsf{LMF}} = \frac{\sqrt{\xi_6^v\,\mathrm{Tr}(\mathbf{Q})\,\mathrm{Tr}(\mathbf{R})}}{3\sigma_v^2}} \qquad (\text{small } \mu). \qquad (49)$$

We can see from the above expressions that μ_o decreases with $\mathrm{Tr}(\mathbf{R})$ and increases with the system nonstationarity variance $\mathrm{Tr}(\mathbf{Q})$. On the other hand, the minimum achievable EMSE of both algorithms increases with the square root of both $\mathrm{Tr}(\mathbf{R})$ and $\mathrm{Tr}(\mathbf{Q})$.

1) Simulation Results: Fig. 4 compares the simulation and theoretical results for the case $\mathbf{Q} = \sigma_q^2\mathbf{I}$ and $\mathbf{R} = \mathbf{I}$ with $\sigma_q = 5 \times 10^{-4}$ and $\delta = 0.8$. Moreover, the noise sequence is a mixture of Gaussian noise and uniform noise with variances $\sigma_n^2 = 10^{-2}$ and $\sigma_c^2 = 10^{-2}/12$, respectively. It is seen in the figure that the theoretical and experimental MSE are a good match. The figure also shows that the steady-state MSE possesses a minimum value of $\zeta = 0.0113$ at $\mu = 0.006$, which are in good agreement with the corresponding theoretical values obtained from (47) and (46) as $\zeta_{\min}^{\mathsf{LMMN}} = 0.011\,36$ and $\mu_o^{\mathsf{LMMN}} = 0.0061$, respectively.

Fig. 5 shows the experimental MSE and the theoretical MSE obtained from (40) versus the norm mixing parameter δ for Gaussian noise of variance $\sigma_v^2 = 10^{-2}, \sigma_q = 10^{-3}$, and $\mu = 0.001$. It is clear that the minimum value of the MSE occurs at $\delta = 1$ for Gaussian noise.
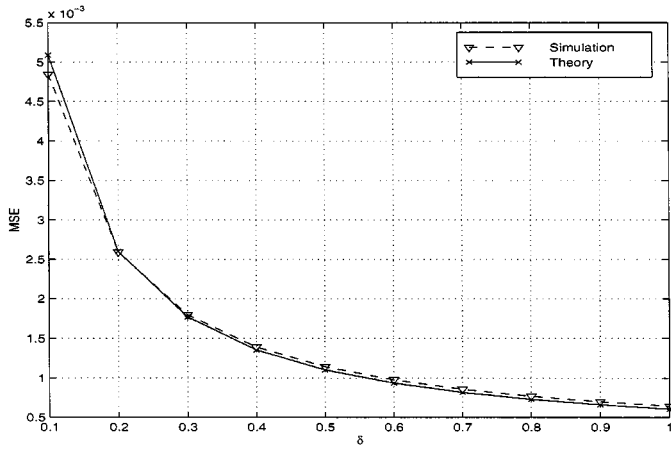
Fig. 5. Experimental and theoretical MSE versus μ for LMMN in a nonstationary environment with Gaussian noise.

2) Comparison with LMS*:* We can also compare the ability of the LMF and LMMN algorithms to track random variations in nonstationary environments with that of the LMS algorithm, which is known to have excellent tracking properties (see, e.g., [1], [3], and [19]). We use the ratio of the minimum achievable steady-state MSE of each of the algorithms to that of the LMS algorithm as a performance measure.

For the LMF algorithm, this ratio is given, from the results of Table IV, by

$$\frac{\zeta_{\min}^{\text{LMS}}}{\zeta_{\min}^{\text{LMF}}} = \frac{3\sigma_v^3}{\sqrt{\xi_6^v}}. \tag{50}$$

Here, we can see that the ratio depends only on the statistical properties of the measurement noise $v(i)$. For the case of the LMMN algorithm, the same ratio is given by

$$\frac{\zeta_{\min}^{\text{LMS}}}{\zeta_{\min}^{\text{LMMN}}} = \frac{\sigma_v b}{\sqrt{a}} \tag{51}$$

which is also dependent on the statistical properties of the noise, as well as on the norm mixing parameter δ. We specialize these results for the following noise distributions.

Gaussian Noise: In this case, $\xi_4^v = 3\sigma_v^4$, and $\xi_6^v = 15\sigma_v^6$. Then, we can verify from (50) that

$$\frac{\zeta_{\min}^{\text{LMS}}}{\zeta_{\min}^{\text{LMF}}} = \sqrt{\frac{3}{5}} \approx -1.1 \text{ dB}.$$

This indicates that the minimum achievable value of steady-state MSE of the LMS algorithm is less than that of the LMF algorithm by approximately 1.1 dB for all values of the noise variance σ_v^2! For the case of the LMMN algorithm, (51) yields

$$\frac{\zeta_{\min}^{\text{LMS}}}{\zeta_{\min}^{\text{LMMN}}} = \frac{\sigma_v b}{\sqrt{\delta^2 \sigma_v^2 + 6\delta\bar{\delta}\sigma_v^4 + 15\bar{\delta}^2\sigma_v^6}}. \tag{52}$$
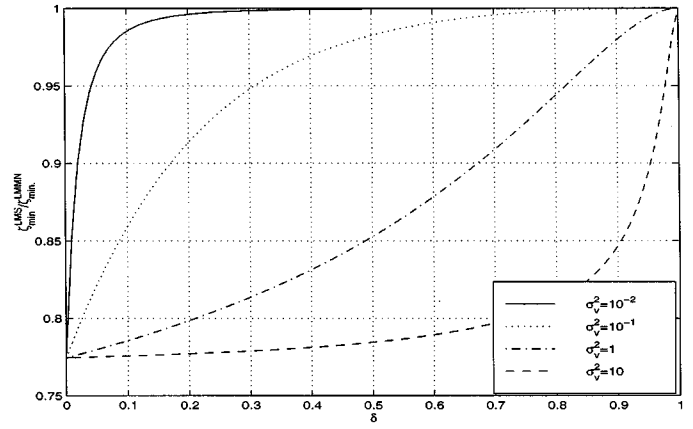

Fig. 6. Comparison of the tracking performance of LMS, LMF, and LMMN for Gaussian noise.
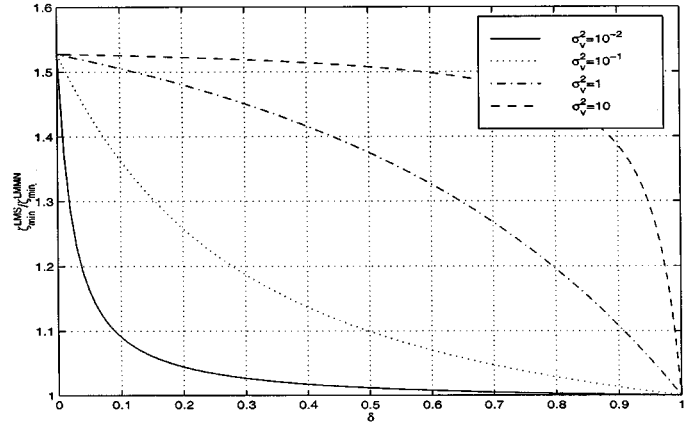

Fig. 7. Comparison of the tracking performance of LMS, LMF, and LMMN for uniform noise.

Fig. 6 shows a plot of this ratio versus the design parameter δ for various values of σ_v^2. The figure shows that this ratio is always less than unity for all values of δ and σ_v^2. These results reflect the superiority of the LMS algorithm over both the LMF and LMMN for tracking nonstationary systems in Gaussian noise environments.

Uniform Noise: For a uniformly distributed noise in the interval $[-\Delta, \Delta]$, we have $\sigma_v^2 = \Delta^2/3, \xi_4^v = \Delta^4/5$ and $\xi_6^v = \Delta^6/7$. Then, we can verify from (50) that

$$\frac{\zeta_{\min}^{\text{LMS}}}{\zeta_{\min}^{\text{LMF}}} = \sqrt{\frac{7}{3}} \approx 3.7 \text{ dB}.$$

This indicates that the minimum achievable value of steady-state MSE of the LMF algorithm is less than that of the LMS algorithm by approximately 3.7 dB for uniformly distributed noise. Fig. 7 shows a plot of the ratio of the minimum achievable EMSE of the LMS and LMMN algorithms versus the design parameter δ for various values of σ_v^2. The figure shows that this ratio is always larger than unity for all values of δ and σ_v^2. We can also see that $\delta = 0$ results in the best tracking performance, which reflects the superiority of the LMF algorithm in this case.

Mixed Gaussian and Uniform Noise: We now consider the case where the noise is a mix of Gaussian and uniform distributions (for example, a mix of Gaussian system noise and uniformly distributed roundoff errors). Fig. 8 shows the ratio of the minimum achievable EMSE of the LMS and LMMN algorithms versus δ for different values of the system noise variance σ_v^2, which is a combination of Gaussian and uniformly distributed noise with variance ratio 1:3. We can see that in this case, the LMMN algorithm will have the best tracking performance. The choice of the optimal norm mixing parameter δ is given in the final section of the paper.

B. Optimal δ and μ

We now derive an expression for the optimal values of $\{\delta, \mu\}$ jointly (recall that δ should lie in the interval $[0, 1]$). Differentiating (40) separately with respect to μ and δ and setting the derivatives equal to zero, we find that ζ^{LMMN} has a unique stationary point at the pair

$$\delta_o = \frac{\xi_6^v - 3\sigma_v^2 \xi_4^v}{(\xi_6^v - 3\sigma_v^2 \xi_4^v) - (\xi_4^v - 3\sigma_v^4)} \qquad (53)$$

and

$$\mu_o = \sqrt{\text{Tr}(\mathbf{Q})/a_o \text{Tr}(\mathbf{R})} \qquad (54)$$

where a_o is a in (23) with δ replaced by δ_o (similarly, we define b_o and c_o). The pair $\{\delta_o, \mu_o\}$ would correspond to a global minimum if, and only if, the Hessian matrix of ζ^{LMMN} is positive-definite at $\{\delta_o, \mu_o\}$. Some algebra will show that the Hessian matrix is given by

$$\mathbf{H}(\delta_o, \mu_o)$$
$$= \begin{bmatrix} \dfrac{\text{Tr}(\mathbf{Q})}{\mu_o^3 b_o} & \dfrac{(1 - 3\sigma_v^2)\, \text{Tr}(\mathbf{Q})}{\mu_o b_o} \\ \dfrac{(1 - 3\sigma_v^2)\, \text{Tr}(\mathbf{Q})}{\mu_o b_o} & \dfrac{\mu_o (\sigma_v^2 - 2\xi_4^v + \xi_6^v)\, \text{Tr}(\mathbf{R})}{b_o} \end{bmatrix}.$$

Now, since its $(1, 1)$ entry is positive, the Hessian matrix will be positive definite if, and only if, the Schur complement with respect to this entry is positive. This leads to the following conditions:

$$\sigma_v^2 + \xi_6^v > 2\xi_4^v, \qquad (55)$$

and

$$\text{Tr}(\mathbf{Q}) < \frac{(\sigma_v^2 - 2\xi_4^v + \xi_6^v)}{(1 - 3\sigma_v^2)^2} \text{Tr}(\mathbf{R}). \qquad (56)$$

Still, these conditions do not guarantee that the δ_o in (53) will lie in the interval $[0, 1]$. Using the above results, and the fact that ζ^{LMMN} has a unique stationary point, we arrive at the following conclusion.

1) If conditions (55) and (56) apply *and* if the resulting δ_o lies in $[0, 1]$, then the optimal parameters are given by (53) and (54).
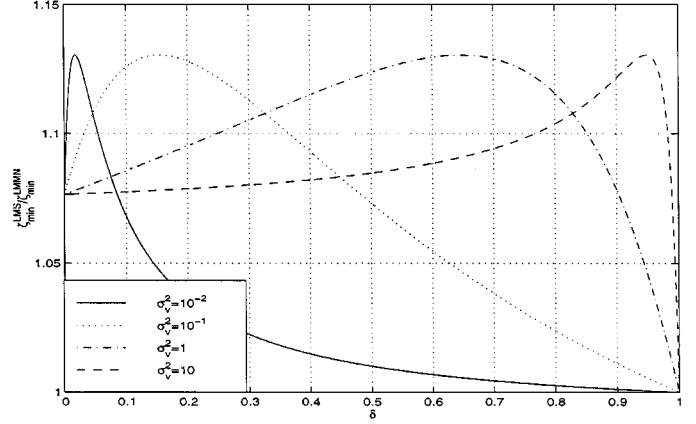

Fig. 8. Comparison of the tracking performance of LMS, LMF, and LMMN for a mixed Gaussian/uniform noise distribution.

2) Otherwise, the optimal δ is either $\delta_o = 0$ or $\delta_o = 1$ with the corresponding μ still given by (54).

The resulting minimum value of the EMSE will be

$$\zeta_{\text{min}}^{\text{LMMN}} = \frac{\sqrt{a_o \text{Tr}(\mathbf{Q}) \text{Tr}(\mathbf{R})}}{b_o} \qquad \text{(small } \mu\text{).} \qquad (57)$$

Gaussian Noise: For Gaussian system noise, $\xi_4^v = 3\sigma_v^4$, and $\xi_6^v = 15\sigma_v^6$. Then

$$\sigma_v^2 + \xi_6^v - 2\xi_4^v = \sigma_v^2 \left(1 - 6\sigma_v^2 + 15\sigma_v^4\right)$$
$$= 15\sigma_v^2 \left(\left(\sigma_v^2 - \frac{1}{5}\right)^2 + \frac{2}{75}\right) > 0$$

which implies that (55) is always true for the Gaussian noise case. Then, if the system degree of nonstationarity satisfies

$$\text{Tr}(\mathbf{Q}) < \frac{\sigma_v^2 \left(1 - 6\sigma_v^2 + 15\sigma_v^4\right) \text{Tr}(\mathbf{R})}{(1 - 3\sigma_v^2)^2} \qquad (58)$$

the optimum value of δ is given from (53) by $\delta_o = 1$, which corresponds to the LMS algorithm with an optimal step size given by

$$\mu_o = \sqrt{\frac{\text{Tr}(\mathbf{Q})}{\sigma_v^2 \text{Tr}(\mathbf{R})}}$$

and a corresponding minimum EMSE of

$$\zeta_{\text{min}} = \sigma_v \sqrt{\text{Tr}(\mathbf{Q}) \text{Tr}(\mathbf{R})}.$$

That is, for Gaussian system noise, if (58) holds, the LMS algorithm outperforms the LMF and LMMN algorithms, which is consistent with the results of the comparison in the previous section. Using a similar approach, we can show that the LMF tracking performance is superior in the case of uniform system noise (i.e., $\delta_o = 0$).

REFERENCES

[1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

[2] O. Macchi, *Adaptive Processing: The LMS Approach with Applications in Transmission*. New York: Wiley, 1995.

[3] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1996.

[4] P. S. R. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*. Boston, MA: Kluwer, 1997.

[5] B. Farhang-Boroujeny, *Adaptive Filters: Theory and Applications*. New York: Wiley, 1998.

[6] B. Widrow, J. McCool, M. G. Larimore, and C. R. Johnson, "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 46, pp. 1151–1162, Aug. 1976.

[7] N. J. Bershad, "Analysis of the normalized LMS algorithm with Gaussian inputs," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 793–806, 1986.

[8] A. H. Sayed and M. Rupp, "A time-domain feedback analysis of adaptive algorithms via the small gain theorem," *Proc. SPIE*, vol. 2563, pp. 458–469, July 1995.

[9] M. Rupp and A. H. Sayed, "A time-domain feedback analysis of filtered-error adaptive gradient algorithms," *IEEE Trans. Signal Processing*, vol. 44, pp. 1428–1439, June 1996.

[10] A. H. Sayed and M. Rupp, *DSP Handbook*. Boca Raton, FL: CRC, 1998, ch. 20.

[11] J. Mai and A. H. Sayed, "A feedback approach to the steady-state performance of fractionally-spaced blind adaptive equalizers," *IEEE Trans. Signal Processing*, vol. 48, pp. 80–91, Jan. 2000.

[12] M. Rupp and A. H. Sayed, "On the convergence of blind adaptive equalizers for constant-modulus signals," *IEEE Trans. Commun.*, vol. 48, pp. 795–803, May 2000.

[13] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*. New York: Springer-Verlag, 1990.

[14] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms*. Englewood Cliffs, NJ, 1995.

[15] S. K. Jones, R. K. Cavin, and W. M. Reed, "Analysis of error-gradient adaptive linear estimators for a class of stationary dependent processes," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 318–329, Mar. 1982.

[16] E. Pfann and R. Stewart, "LMS adaptive filtering with $\Sigma\Delta$ modulated input signals," *IEEE Signal Processing Lett.*, vol. 5, pp. 95–97, Apr. 1998.

[17] D. T. M. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Trans. Signal Processing*, vol. 41, pp. 2811–2825, Sept. 1993.

[18] V. Mathews and S. Cho, "Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 450–454, Apr. 1987.

[19] E. Eweda, "Comparison of RLS, LMS, and sign algorithms for tracking randomly time-varying channels," *IEEE Trans. Signal Processing*, vol. 42, pp. 2937–2944, Nov. 1994.

[20] R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IRE Trans. Inform. Theory*, vol. IT-4, pp. 69–72, June 1958.

[21] N. R. Yousef and A. H. Sayed, "Steady-state and tracking analyzes of the sign algorithm without the explicit use of the independence assumption," *IEEE Signal Processing Lett.*, vol. 7, pp. 307–309, Nov. 2000.

[22] J. A. Chambers, O. Tanrikulu, and A. G. Constantindes, "Least mean mixed-norm adaptive filtering," *Electron. Lett.*, vol. 30, no. 19, pp. 1574–1575, Sept. 1994.

[23] E. Walach and B. Widrow, "The least mean fourth (LMF) adaptive algorithm and its family," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 275–283, Mar. 1984.

[24] O. Tanrikulu and J. A. Chambers, "Convergence and steady-state properties of the least-mean mixed-norm (LMMN) adaptive algorithm," *Proc. Inst. Elect. Eng., Vis., Image Signal Process.*, vol. 143, no. 3, pp. 137–142, June 1996.

[25] I. Fijakow, C. E. Manlove, and C. R. Johnson, "Adaptive fractionally spaced blind CMA equalization: Excess MSE," *IEEE Trans. Signal Processing*, vol. 46, pp. 227–231, Jan. 1998.

[26] N. R. Yousef and A. H. Sayed, "A generalized tracking analysis of adaptive filtering algorithms in cyclically and randomly varying environments," in *Proc. ICASSP*, vol. I, Istanbul, Turkey, June 2000, pp. 440–443.

[27] ——, "A feedback analysis of the tracking performance of blind adaptive equalization algorithms," in *Proc. Conf. Decision Contr.*, vol. 1, Phoenix, AZ, Dec. 1999, pp. 174–179.

[28] S. Cho and V. Mathews, "Tracking analysis of sign algorithm in nonstationary environments," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 2046–2057, Dec. 1990.

[29] E. Eweda, "Optimum step size of the sign algorithm for nonstationary adaptive filtering," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1897–1901, Nov. 1990.

[30] M. Rupp and A. H. Sayed, "Robustness of Gauss–Newton recursive methods: A deterministic feedback analysis," *Signal Process.*, vol. 50, no. 3, pp. 165–188, May 1996.

Nabil R. Yousef (S'99) was born in Cairo, Egypt, in 1971. He received the B.S. and M.S. degrees in electrical engineering from Ain Shams University, Cairo, in 1994 and 1997, respectively. Since September 1997, he has been pursuing the Ph.D. degree in electrical engineering at the University of California, Los Angeles (UCLA).

His research interests include adaptive filtering, equalization, CDMA systems, and wireless location.

Mr. Yousef is the recipient of a 1999 Best Student Paper Award at an international meeting for work on adaptive filtering and of the 1999 NOKIA Fellowship Award.

Ali H. Sayed (SM'99–F'01) received the Ph.D. degree in electrical engineering in 1992 from Stanford University, Stanford, CA.

He is Associate Professor of electrical engineering at the University of California, Los Angeles (UCLA). He has over 135 journal and conference publications, is co-author of the research monograph *Indefinite Quadratic Estimation and Control* (Philadelphia, PA: SIAM, 1999) and of the graduate-level textbook *Linear Estimation* (Englewood Cliffs, NJ: Prentice-Hall, 2000). He is also co-editor of the volume *Fast Reliable Algorithms for Matrices with Structure* (Philadelphia, PA: SIAM, 1999). He is a member of the editorial boards of the *SIAM Journal on Matrix Analysis and Its Applications* and of the *International Journal of Adaptive Control and Signal Processing*, and has served as co-editor of special issues of the journal *Linear Algebra and Its Applications*. He has contributed several articles to engineering and mathematical encyclopedias and handbooks and has served on the program committees of several international meetings. He has also consulted to several companies in the areas of adaptive filtering, adaptive equalization, and echo cancellation. His research interests span several areas including adaptive and statistical signal processing, filtering and estimation theories, equalization techniques for communications, interplays between signal processing and control methodologies, and fast algorithms for large-scale problems.

Dr. Sayed is a recipient of the 1996 IEEE Donald G. Fink Award. He is Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING. To learn more about his work, visit the website of the UCLA Adaptive Systems Laboratory at http://www.ee.ucla.edu/asl.