Analysis of Stability and Performance of Adaptation Algorithms With Time-Invariant Gains

Anders Ahlén, *Senior Member, IEEE*, Lars Lindbom, and Mikael Sternad, *Senior Member, IEEE*

Abstract—Adaptation laws that track parameters of linear regression models are investigated. The considered class of algorithms apply linear time-invariant filtering on the instantaneous gradient vector and includes least mean squares (LMS) as its simplest member. The asymptotic stability and steady-state tracking performance for prediction and smoothing estimators is analyzed for parameter variations described by stochastic processes with time-invariant statistics. The analysis is based on a novel technique that decomposes the inherent feedback of adaptation algorithms into one time-invariant loop and one time-varying loop. The impact of the time-varying feedback on the tracking error covariance can be neglected under certain conditions, and the performance analysis then becomes straight-forward. Performance analysis in the presence of a non-negligible time-varying feedback is performed for algorithms that use scalar measurements. Convergence in mean square error (MSE) and the MSE tracking performance is investigated, assuming independent consecutive regression vectors. Closed-form expressions for the tracking MSE are thereafter derived without this independence assumption for a subclass of algorithms applied to finite impulse response (FIR) models with white inputs. This class includes Wiener LMS adaptation.

Index Terms—Adaptive filtering, adaptive signal processing, least mean squares method, tracking.

I. INTRODUCTION

A COMMON adaptation task is to estimate the parameter vector h_t in a linear regression

$$y_t = \varphi_t^* h_t + v_t \tag{1}$$

from measurements of the discrete-time signal y_t and a possibly complex-valued regression matrix φ_t^* that is known at time t. We may obtain an estimate $\hat{h}_{t+1|t}$ via the LMS algorithm

$$\varepsilon_t = y_t - \varphi_t^* \hat{h}_{t|t-1} \tag{2}$$

$$\hat{h}_{t+1|t} = \hat{h}_{t|t-1} + \mu \varphi_t \varepsilon_t \tag{3}$$

where $\varphi_t \varepsilon_t$ is the instantaneous negative gradient of $|\varepsilon_t|^2/2$ at the estimate $\hat{h}_{t|t-1}$. When tracking time-varying parameters, a too-small gain μ leads to a large average estimation error (lag error), whereas a too-high gain results in high sensitivity to the noise v_t in (1).

It may sometimes be hard to obtain adequate tracking performance with LMS, mainly because it implicitly assumes random

walk parameter variations: an assumption that is frequently invalid. This paper considers a larger class of adaptation laws with constant gains, which allow prior information on the dynamics of h_t to be utilized to introduce an appropriate amount of coupling and inertia in the estimates. We may also obtain filtering $(k=0)$, prediction $(k>0)$, or fixed-lag smoothing $(k<0)$ estimates $\hat{h}_{t+k|t}$ for arbitrary horizons k. This class of algorithms can be expressed as

$$\varepsilon_t = y_t - \varphi_t^* \hat{h}_{t|t-1} \tag{4}$$

$$\hat{h}_{t+k|t} = \mathcal{M}_k(q^{-1})(\varphi_t \varepsilon_t) = \sum_{i=0}^{\infty} \mathbf{M}_{k,i} \varphi_{t-i} \varepsilon_{t-i} \tag{5}$$

where $\mathcal{M}_k(q^{-1})$ is a matrix of transfer operators in the backward shift operator $q^{-1} x_t = x_{t-1}$. Thus, the estimate $\hat{h}_{t+k|t}$ is formed by linear time-invariant filtering of the instantaneous gradient. For the special case of LMS (3), in which $k = 1$, the filter is

$$\mathcal{M}_1(q^{-1}) = \frac{\mu}{1-q^{-1}} \mathbf{I} \tag{6}$$

while higher order filters correspond to so-called multistep algorithms [1]–[3]. The momentum LMS algorithm [4] and the lead-lag LMS scheme of [5] are examples of multistep algorithms with diagonal $\mathcal{M}_1(q^{-1})$. In general, $\mathcal{M}_k(q^{-1})$ need not be diagonal. Wiener methods that for arbitrary k optimize $\mathcal{M}_k(q^{-1})$ in MSE are presented in [6] and [7]. Such Wiener designed estimators with constant gains may attain steady-state performance close to that of the optimal Kalman estimator at much lower complexity [7]. A subclass that is suitable for tracking mobile radio channels is represented by the Wiener LMS structure discussed in [6], for which

$$\mathcal{M}_k(q^{-1}) = \frac{Q_k(q^{-1})}{D(q^{-1})} \mathbf{R}^{-1} \tag{7}$$

where $Q_k(q^{-1})$ and $D(q^{-1})$ are polynomials, and

$$\mathbf{R} = \mathrm{E}\varphi_t \varphi_t^*. \tag{8}$$

The analysis of stability and performance of (4) and (5) is the topic of the present paper. This task is, in general, far from trivial. By using a state-space realization of a one-step prediction filter $\mathcal{M}_1(q^{-1})$ of arbitrary structure, the expressions (1), (4), and (5) could be iterated to obtain explicit, but very involved, expressions for the one-step prediction error. Similar expressions form the basis of many works; see, e.g., [8]–[13]. Even for the LMS case, a strict analysis becomes very difficult.

Adaptation laws do inherently have a feedback structure; therefore, analysis of the feedback, primarily to ascertain

stability, has been an important theme in the literature. For example, in [14], the usefulness of exploiting the underlying feedback structure of adaptive algorithms is addressed thoroughly, and stability conditions are derived for both the deterministic and the stochastic case. A feedback analysis somewhat related to our approach can be found in [15] for LMS assuming independent regressors. The work presented in [16] and [17] exploits the inherent feedback structure of such algorithms even further by the use of the small gain theorem, and [16] also highlights connections to results in H_∞ theory. Although the small gain theorem is a powerful tool, it may, for certain feedback structures, give conservative stability bounds, as will be evident from Example 3. On the other hand, the energy conservation approach of [17], relying on this theorem, requires no assumptions on the regressor distributions. See also [18].

As outlined in Section II-B, the inherent feedback loop may be partitioned into an inner loop with linear time invariant dynamics and an outer loop with time-varying gain. This decomposition is of use not only for design [6], [7] but also for analysis. We here utilize it as a new tool for understanding the feedback path (4).

The MSE tracking performance of adaptation laws described by (4) and (5) can easily be obtained when the time-varying outer feedback loop is insignificant. This will be the case either when the parameter variations are slow or when the variance of the noise v_t in (1) is large. Expressions for the spectra of lag errors and noise induced errors are derived under this assumption in Section III. We connect the property of a negligible time-varying loop with the concepts of the degree of nonstationarity and "slow time variations," as defined by Macchi in [19]. However, it is emphasized that the impact of the time-varying loop on the tracking performance depends on the algorithm. The impact will be smaller on algorithms with superior tracking performance.

For situations where the time-varying feedback is not negligible, Section IV-A proposes an analysis based on the assumption that consecutive regressor vectors $\varphi_t^*, \varphi_\tau^*, t \neq \tau$ are independent, with Gaussian regressors and scalar y_t. Stability in MSE is then guaranteed by the asymptotic stability of a linear scalar transfer function, and the tracking MSE is derived. The results are novel in that they hold for $\mathcal{M}_k(q^{-1})$ of arbitrary structure and complexity for parameter variations with arbitrary power spectral density.

Independent regressors is a commonly used simplifying assumption [3], [20]–[22], [24]–[26]. While this assumption is quite restrictive and does not apply to the modeling of dynamic systems, useful approximative results can be obtained in situations where it is not true.

Section IV-B then discusses FIR systems with uncorrelated (white) zero mean regressors. Important examples are fading baseband channels in wireless transmission. The transmitted data are then regressors (elements of φ_t^*), and they are, in general, uncorrelated due to coding and interleaving. A novel analysis is presented for this case, without assuming the time-varying loop to be vanishing and without assuming independent φ_t^*. Performance results are derived for algorithms with

diagonal $\mathcal{M}_k(q^{-1})$ under approximations that hold exactly for FIR systems with two parameters and zero mean inputs with constant modulus. Stability and convergence in MSE can here be ascertained by checking the asymptotic stability of a transfer function.

II. PRELIMINARIES

Notation: The identity matrix is denoted $\mathbf{I}$. A superscript asterisk represents conjugation and transposition. Polynomial matrix fractions [27] are used to represent transfer functions/rational matrices. For polynomial matrices $\boldsymbol{P}(q^{-1})$ and rational matrices $\boldsymbol{\mathcal{R}}(q^{-1})$, conjugate matrices $\boldsymbol{P}_*(q)$ or $\boldsymbol{\mathcal{R}}_*(q)$ are obtained by conjugating coefficients, transposing and substituting the forward shift operator q for the backward shift operator q^{-1}. The arguments q or q^{-1} are sometimes omitted. Scalar polynomials $P(q^{-1})$ are represented by nonboldface capitals.

Square polynomial matrices $\boldsymbol{P}(q^{-1})$ will be called stable if all zeros of $\det(\boldsymbol{P}(z^{-1}))$ are located in $|z| < 1$ and marginally stable if these zeros are located in $|z| \leq 1$. A transfer operator $\boldsymbol{\mathcal{R}}(z^{-1}) = \boldsymbol{R}^{-1}(z^{-1})\boldsymbol{S}(z^{-1})$ is denoted stable if $\boldsymbol{R}(z^{-1})$ is a stable polynomial matrix. The linear dynamic system $z_t = \boldsymbol{\mathcal{R}}(q^{-1})u_t$ is then asymptotically stable and L_p-stable for all $p \in [1, \infty]$ [28].

"White" denotes sequences that are uncorrelated but not necessarily independent.

A. Basic Assumptions

The following assumptions on the signals in (1) are used throughout the paper.

Assumption 1: The parameter vector h_t has spectral density $\Phi_h(\omega)$ and the additive noise v_t is stationary and zero mean, whereas φ_t^*, with known dimension, is stationary with zero mean and finite second- and higher order moments. The regressor correlation matrix $\mathbf{R}$ is nonsingular and time-invariant. Moreover, h_t, v_t, and φ_t^* are mutually independent processes. □

Remark 1: The parameter vector h_t does not necessarily have zero mean. The assumption that φ_t^* is independent of h_t and of v_t excludes models that use (filtered) measurements y_{t-i} for $i > 0$ as regressors, such as AR or ARX regression models. □

B. Learning Filter

The algorithm (4) and (5) can be expressed as a causal and time-invariant filter, which is denoted the *learning filter* $\mathcal{L}_k(q^{-1})$, that operates on a signal vector

$$f_t \triangleq \varphi_t \varepsilon_t + \mathbf{R}\hat{h}_{t|t-1}. \tag{9}$$

The introduction of the learning filter enables the use of Wiener theory for the design of $\mathcal{M}_k(q^{-1})$ in (5); see [7]. This leads to a decomposition of the inherent feedback structure that is also useful for analyzing the tracking behavior of the algorithm (4) and (5). The reformulation of the algorithm (4) and (5) to a learning filter is central in this paper, and it is outlined below. If

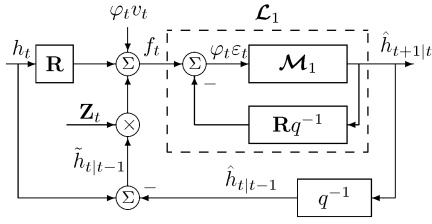$$\hat{h}_{t|t-1} = q^{-1}\mathcal{M}_1(q^{-1})(\varphi_t \varepsilon_t)$$

Fig. 1. Adaptive algorithm (5) ($k = 1$) reformulated as a time-invariant filter operating on the signal f_t. The learning filter $\mathcal{L}_1(q^{-1})$ represents the inner time-invariant loop. The time-varying feedback loop around $\mathcal{L}_1(q^{-1})$, via the feedback noise $\mathbf{Z}_t \tilde{h}_{t|t-1}$, may cause instability even if the learning filter is stable.

from (5) with $k = 1$ is used in (9), then we obtain

$$\varphi_t \varepsilon_t = \left(\mathbf{I} + \mathbf{R} q^{-1} \mathcal{M}_1 \left(q^{-1} \right) \right)^{-1} f_t.$$

The estimator (5) can thus, for arbitrary k, be expressed as

$$\hat{h}_{t+k|t} = \mathcal{L}_k(q^{-1}) f_t = \sum_{i=0}^{\infty} \mathbf{L}_{k,i} f_{t-i} \qquad (10)$$

in which

$$\mathcal{L}_k(q^{-1}) = \mathcal{M}_k(q^{-1}) \left(\mathbf{I} + \mathbf{R} q^{-1} \mathcal{M}_1(q^{-1}) \right)^{-1}. \qquad (11)$$

We now investigate the signal f_t. Denote the k-step estimation error $\tilde{h}_{t|t-k} = h_t - \hat{h}_{t|t-k}$. By (4) and (1)

$$\varphi_t \varepsilon_t = \varphi_t \varphi_t^* \tilde{h}_{t|t-1} + \varphi_t v_t. \qquad (12)$$

Adding and subtracting $\mathbf{R} \tilde{h}_{t|t-1}$ in (12) gives

$$\varphi_t \varepsilon_t = \mathbf{R} h_t - \mathbf{R} \hat{h}_{t|t-1} + (\varphi_t \varphi_t^* - \mathbf{R}) \tilde{h}_{t|t-1} + \varphi_t v_t. \qquad (13)$$

Define

$$\mathbf{Z}_t \triangleq \varphi_t \varphi_t^* - \mathbf{R} \qquad (14)$$

$$\eta_t \triangleq \mathbf{Z}_t \tilde{h}_{t|t-1} + \varphi_t v_t \qquad (15)$$

where we denote $\mathbf{Z}_t$ and η_t as the *autocorrelation matrix noise* and the *gradient noise*, respectively. By inserting (13) into (9), f_t may now be expressed as

$$f_t = \mathbf{R} h_t + \mathbf{Z}_t \tilde{h}_{t|t-1} + \varphi_t v_t = \mathbf{R} h_t + \eta_t. \qquad (16)$$

The point of expressions (9)–(11) and (16) is to decompose the feedback loop into two parts; see Fig. 1: an inner time-invariant loop via $\mathbf{R}$, which is absorbed into the definition of $\mathcal{L}_1(q^{-1})$ by (11), and an outer time-varying feedback represented by $\mathbf{Z}_t \tilde{h}_{t|t-1}$, which we call the *feedback noise*.

The inherent feedback of adaptation algorithms can, of course, be expressed in various ways. For example, use of the gradient expression (12) in (5) results in a single loop with a time-varying feedback gain matrix, as illustrated by Fig. 2. The filter operating on $\varphi_t y_t$ to produce $\hat{h}_{t+1|t}$ will be time-varying (except for cases with scalar φ_t with constant modulus). It can be noted that an approximation of $\varphi_t \varphi_t^*$ by its average $\mathbf{R}$ in Fig. 2, i.e., using the direct averaging method described by Kushner [29], would correspond to setting $\mathbf{Z}_t \tilde{h}_{t|t-1} = 0$ in Fig. 1.
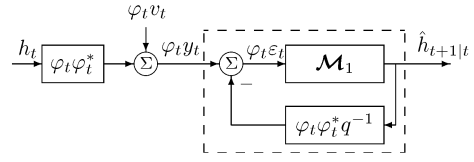

Fig. 2. Algorithm (5) for $k = 1$, expressed as a time-varying filter operating on the signal $\varphi_t y_t$.

The one-step prediction learning filter $\mathcal{L}_1(q^{-1})$ is part of the loop for all k, and its properties are crucial for the algorithm stability. A necessary, but not sufficient, condition for asymptotic stability is *internal stability of the learning filter $\mathcal{L}_1(q^{-1})$* (representing the inner loop in Fig. 1). Wiener design, which is described in Section II-D, guarantees asymptotic stability of $\mathcal{L}_k(q^{-1}) \, \forall \, k$. For algorithms obtained by other means, stability of the learning filter will have to be verified separately.

The stability requirement on $\mathcal{L}_1(q^{-1})$ has an interesting interpretation for LMS, where

$$\mathcal{L}_1(q^{-1}) = \left(\mathbf{I} - (\mathbf{I} - \mu \mathbf{R}) q^{-1} \right)^{-1} \mu \qquad (17)$$

which is obtained by inserting (6) into (11). Asymptotic stability of (17) coincides with the classical LMS condition for stability in the mean [24], $0 < \mu < 2/\lambda_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{R}$.

In general, the outer feedback $\mathbf{Z}_t \tilde{h}_{t|t-1}$ will also have to be taken into account. A sufficient but conservative deterministic condition for stability is then provided by the small gain theorem; see, e.g., [28]. Consider the mapping from $\tilde{h}_{t|t-1}$ to $\tilde{h}_{t+1|t}$ in Fig. 1 without external inputs, i.e., for $h_t = 0$, $\varphi_t v_t = 0$. If $\mathcal{L}_1(q^{-1})$ is causal and asymptotically stable, L_p-stability will be assured if

$$\| \mathcal{L}_1(q^{-1}) \left(\mathbf{Z}_t \tilde{h}_{t|t-1} \right) \|_p \leq \gamma \| \tilde{h}_{t|t-1} \|_p; \quad \gamma < 1. \qquad (18)$$

Condition (18) is of general use, but less conservative conditions for special cases will be derived below.

C. Tracking Error

By (10) and (16), the tracking error can be expressed as

$$\tilde{h}_{t+k|t} = \left(\mathbf{I} - q^{-k} \mathcal{L}_k \mathbf{R} \right) h_{t+k} - \mathcal{L}_k(\varphi_t v_t) - \mathcal{L}_k \left(\mathbf{Z}_t \tilde{h}_{t|t} \right). \qquad (19)$$

The three error contributions are the *lag error* $(\mathbf{I} - q^{-k} \mathcal{L}_k(q^{-1}) \mathbf{R}) h_{t+k}$, a noise term $\mathcal{L}_k(q^{-1})(\varphi_t v_t)$, and a feedback noise term $\mathcal{L}_k(q^{-1})(\mathbf{Z}_t \tilde{h}_{t|t-1})$ caused by old parameter tracking errors.

By Assumption 1, h_t and $\varphi_t v_t$ are independent. The steady-state tracking error covariance matrix (after the initial transient) can thus be expressed as

$$\mathbf{P}_k \triangleq \lim_{t \to \infty} \mathbf{P}_{t+k|t} = \lim_{t \to \infty} \mathrm{E} \tilde{h}_{t+k|t} \tilde{h}_{t+k|t}^* \qquad (20)$$

$$= \lim_{t \to \infty} \left(\mathbf{V}_{h,t}^k + \mathbf{V}_{\varphi v,t}^k + \mathbf{V}_{Z\tilde{h},t}^k + \mathbf{V}_{hZ\tilde{h},t}^k + \mathbf{V}_{\varphi v Z\tilde{h},t}^k \right) \qquad (21)$$

where

$$\mathbf{V}_{h,t}^k = \mathrm{E} \left(\mathbf{I} - q^{-k}\boldsymbol{\mathcal{L}}_k\mathbf{R}\right) h_{t+k}$$
$$\times \left(\left(\mathbf{I} - q^{-k}\boldsymbol{\mathcal{L}}_k\mathbf{R}\right) h_{t+k}\right)^* \qquad (22)$$

$$\mathbf{V}_{\varphi v,t}^k = \mathrm{E}\boldsymbol{\mathcal{L}}_k(\varphi_t v_t) \left(\boldsymbol{\mathcal{L}}_k(\varphi_t v_t)\right)^* \qquad (23)$$

$$\mathbf{V}_{Z\tilde{h},t}^k = \mathrm{E}\boldsymbol{\mathcal{L}}_k \left(\mathbf{Z}_t\tilde{h}_{t|t-1}\right) \left(\boldsymbol{\mathcal{L}}_k \left(\mathbf{Z}_t\tilde{h}_{t|t-1}\right)\right)^*$$
$$= \sum_{i=0}^{t} \sum_{j=0}^{t} \mathbf{L}_{k,i}$$
$$\times \mathrm{E}\left\{\mathbf{Z}_{t-i}\tilde{h}_{t-i|t-i-1}\tilde{h}_{t-j|t-j-1}^*\mathbf{Z}_{t-j}^*\right\} \mathbf{L}_{k,j}^* \quad (24)$$

with $\mathbf{L}_{k,i}$ denoting the impulse response matrices of $\boldsymbol{\mathcal{L}}_k(q^{-1})$; see (10). The last terms are

$$\mathbf{V}_{hZ\tilde{h},t}^k = -\mathrm{E}\left(\mathbf{I} - q^{-k}\boldsymbol{\mathcal{L}}_k\mathbf{R}\right) h_{t+k}$$
$$\times \left(\boldsymbol{\mathcal{L}}_k \left(\mathbf{Z}_t\tilde{h}_{t|t-1}\right)\right)^*$$
$$- \mathrm{E}\left(\left(\mathbf{I} - q^{-k}\boldsymbol{\mathcal{L}}_k\mathbf{R}\right) h_{t+k}\right)^*$$
$$\times \boldsymbol{\mathcal{L}}_k \left(\mathbf{Z}_t\tilde{h}_{t|t-1}\right) \qquad (25)$$

$$\mathbf{V}_{\varphi v Z\tilde{h},t}^k = \mathrm{E}\boldsymbol{\mathcal{L}}_k(\varphi_t v_t) \left(\boldsymbol{\mathcal{L}}_k \left(\mathbf{Z}_t\tilde{h}_{t|t-1}\right)\right)^*$$
$$+ \mathrm{E}\left(\boldsymbol{\mathcal{L}}_k(\varphi_t v_t)\right)^* \boldsymbol{\mathcal{L}}_k \left(\mathbf{Z}_t\tilde{h}_{t|t-1}\right). \qquad (26)$$

We analyze the convergence properties in MSE of (19) and provide conditions under which the feedback noise does not cause instability. Furthermore, expressions are derived for the error covariance matrix (20) and (21). In Section III, only the first two terms $\mathbf{V}_{h,t}^k$ and $\mathbf{V}_{\varphi v,t}^k$ are used, whereas in Section IV, the term $\mathbf{V}_{Z\tilde{h},t}^k$ is added and estimated by taking feedback effects into account. The two last terms in (21) will be significantly smaller than the terms (22)–(24); see Table III in Section IV for an illustration. These contributions are either zero or are neglected in the present analysis.

D. Wiener Design

When the term η_t in (16) is regarded as a zero mean stationary additive noise, then stable Wiener-based learning filters can be designed to minimize the steady-state tracking error covariance matrix. The Wiener design presented in [7] will be used in Section III, and it is for that purpose summarized here with slightly modified notation.

Assumption 2: Let the signals representing the gradient noise η_t, and the parameter vector h_t, be described by ARMA processes with common denominator polynomials

$$\eta_t = \frac{1}{N(q^{-1})}\boldsymbol{M}(q^{-1})\nu_t; \quad h_t = \frac{1}{D(q^{-1})}\boldsymbol{C}(q^{-1})e_t \quad (27)$$

where $N(q^{-1})$ is required to be stable, whereas $D(q^{-1})$ may be marginally stable. Furthermore, let $\boldsymbol{M}(q^{-1})$ and $\boldsymbol{C}(q^{-1})$ be stably invertible with full rank on $|z| = 1$. Here, ν_t and e_t are zero mean white noises, which are the innovation sequences of η_t and h_t, respectively, with covariance matrices

$$\mathrm{E}\nu_t\nu_t^* \triangleq \mathbf{R}_\nu = \lambda_\nu\bar{\mathbf{R}}_\nu; \quad \mathrm{E}e_te_t^* \triangleq \mathbf{R}_e = \lambda_e\bar{\mathbf{R}}_e \quad (28)$$

where $\bar{\mathbf{R}}_e$ is nonsingular. The white sequence ν_t is assumed uncorrelated with h_{t-i} and with $\hat{h}_{t-i|t-i-1}, i \geq 0$. $\square$

Remark 2: Uncorrelatedness between ν_t and $\hat{h}_{t-i|t-i-1}, i \geq 0$ is a key assumption for Wiener design. It will in general not hold exactly $\square$

Remark 3: In (28), both λ_ν and λ_e represent scalar scaling factors. In the modeling of the signals η_t and h_t, these scalings are normally set to unity. The statistical properties of η_t and h_t are then solely determined by (27) and the covariance matrices $\bar{\mathbf{R}}_\nu$ and $\bar{\mathbf{R}}_e$, respectively. In order to investigate slowly varying parameters, we will regard λ_ν and λ_e as scalings that affect all signal components of η_t and h_t equally, while leaving the correlation properties of η_t and h_t unaffected. $\square$

Under Assumption 2, the spectral densities of η_t and h_t are readily found to be

$$\Phi_\eta(\omega) = \lambda_\nu\bar{\Phi}_\eta(\omega) = \lambda_\nu\frac{\boldsymbol{M}(\omega)\bar{\mathbf{R}}_\nu\boldsymbol{M}_*(\omega)}{N(\omega)N_*(\omega)} \qquad (29)$$

and

$$\Phi_h(\omega) = \lambda_e\bar{\Phi}_h(\omega) = \lambda_e\frac{\boldsymbol{C}(\omega)\bar{\mathbf{R}}_e\boldsymbol{C}_*(\omega)}{D(\omega)D_*(\omega)} \qquad (30)$$

respectively. A Wiener design of $\boldsymbol{\mathcal{L}}_k(q^{-1})$ is now obtained as follows.

Lemma 1: Under Assumption 2 and assuming a nonsingular $\mathbf{R}$, the MSE-optimal learning filter (10) is asymptotically stable and is given by

$$\boldsymbol{\mathcal{L}}_k^{opt}(q^{-1}) = \boldsymbol{Q}_k(q^{-1})\boldsymbol{\beta}(q^{-1})^{-1}N(q^{-1})\mathbf{R}^{-1}. \qquad (31)$$

Here, $\boldsymbol{\beta}(q^{-1})$ is the stably and causally invertible solution to the left polynomial matrix spectral factorization

$$\boldsymbol{\beta}\boldsymbol{\beta}_* = \lambda_e C\bar{\mathbf{R}}_e C_* N N_* + \lambda_\nu D\mathbf{R}^{-1}\mathbf{M}\bar{\mathbf{R}}_\nu\mathbf{M}_*\mathbf{R}^{-1}D_* \quad (32)$$

whereas $\boldsymbol{Q}_k(q^{-1})$, together with a polynomial matrix $\boldsymbol{U}_{k*}(q)$, is the unique solution to the Diophantine equation

$$\lambda_e q^k C\bar{\mathbf{R}}_e C_* N_* = \boldsymbol{Q}_k\boldsymbol{\beta}_* + q\mathbf{I}DU_{k*}. \qquad (33)$$

Proof: The proof is obtained by Theorem 1 in [7] and modified to an ARMA model structure for h_t with common scalar denominator polynomial. $\square$

Integrating design models, i.e., models (27) with $1 - q^{-1}$ as factors in $D(q^{-1})$, can be used to obtain unbiased estimators of vectors h_t that are not really random walks but have nonzero means. An example is Rician fading mobile radio channels. The general condition for unbiased estimation of nonzero mean vectors h_t is $\boldsymbol{\mathcal{L}}_k(z^{-1})\big|_{z=1} = \mathbf{R}^{-1}$. Wiener design based on integrating design models assures this property. The LMS algorithm (17) will always satisfy this condition since it is based on a random walk model; see [7, ex. 1].

The Wiener LMS or WLMS algorithm (4), (5), and (7) assumes η_t to be white and all components of h_t to have the same

dynamics, i.e., $\boldsymbol{C}(q^{-1}) = C(q^{-1})\mathbf{I}$. It has a learning filter given by

$$\boldsymbol{\mathcal{L}}_k(q^{-1}) = \frac{Q_k(q^{-1})}{\beta(q^{-1})}\mathbf{R}^{-1}. \tag{34}$$

As outlined in [6], scalar versions of (32) and (33) provide optimized polynomials $Q_k(q^{-1})$ and $\beta(q^{-1})$.

E. Example for AR (2) Variations

For the purpose of illustrating results presented in Section III and in Section IV-B, we consider tracking of the coefficients of a scalar M-tap time-varying FIR system (1), in which

$$\varphi_t^* = (u_t \; u_{t-1} \cdots u_{t-M+1}). \tag{35}$$

The regressors u_i are white with zero mean and unit variance, yielding $\mathbf{R} = \mathbf{I}$. The parameter dynamics are governed by a second-order oscillatory vector AR process (27) with

$$D(q^{-1}) = 1 + d_1 q^{-1} + d_2 q^{-2} \text{ and } \boldsymbol{C}(q^{-1}) = \mathbf{I}.$$

We select $d_1 = -2p\cos(\omega_o)$ and $d_2 = p^2$, or

$$h_t - 2p\cos(\omega_o)h_{t-1} + p^2 h_{t-2} = e_t \tag{36}$$

and use $p = 0.999$ and $\mathbf{R}_e = \lambda_e \mathbf{I}$. The spectrum of h_t then follows readily from (30) as

$$\Phi_h(\omega) = \frac{\lambda_e}{|D(\omega)|^2}\mathbf{I}$$

with $D(\omega) = 1 - 2p\cos(\omega_o)e^{-j\omega} + p^2 e^{-j2\omega}$. In the illustrations, the number of regressors M and the frequency ω_o will be varied. The variance of the noise v_t is set to 0.01, and λ_e is adjusted to give an output SNR of 20 dB, with h_t having Euclidean norm $|h_t|^2 = 1$ for all values of M and ω_o.

III. NEGLECTING THE TIME-VARYING FEEDBACK

We begin with a simplified analysis that neglects the feedback noise $\mathbf{Z}_t \tilde{h}_{t|t-1}$ and the outer time-varying loop in Fig. 1. It thus considers only the first two steady-state covariance matrices in (21).

Theorem 1: Let the contribution of the feedback noise $\mathbf{Z}_t \tilde{h}_{t|t-1}$ to the error covariance matrix (20) be neglected. If $\boldsymbol{\mathcal{L}}_k(q^{-1})$ is stable, then (20) will under Assumption 1 be given by

$$\lim_{t \to \infty}(\mathbf{V}_{h,t}^k + \mathbf{V}_{\varphi v,t}^k) \triangleq \mathbf{V}_h^k + \mathbf{V}_{\varphi v}^k \triangleq \mathbf{P}_{k,\text{slow}} \tag{37}$$

where

$$\mathbf{V}_h^k = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\mathbf{I} - e^{-j\omega k}\boldsymbol{\mathcal{L}}_k(\omega)\mathbf{R}\right)\Phi_h(\omega)$$
$$\times \left(\mathbf{I} - e^{j\omega k}\boldsymbol{\mathcal{L}}_{k*}(\omega)\mathbf{R}\right)d\omega \tag{38}$$

$$\mathbf{V}_{\varphi v}^k = \frac{1}{2\pi}\int_{-\pi}^{\pi} \boldsymbol{\mathcal{L}}_k(\omega)\Phi_\eta(\omega)\boldsymbol{\mathcal{L}}_{k*}(\omega)d\omega \tag{39}$$

and where $\Phi_\eta(\omega)$ is the spectral density of $\varphi_t v_t$.

Proof: The proof is immediate from (22) and (23). $\square$

The integrands of (38) and (39) describe the spectral densities of the lag error and the noise-induced error, respectively. They describe situations that include arbitrary spectral densities $\Phi_h(\omega)$ of the parameters as well as colored measurement noise v_t and correlated regressor matrices φ_t.

Remark 4: By Assumption 1, the noise term $\mathbf{V}_{\varphi v}^k$ is bounded if $\boldsymbol{\mathcal{L}}_k(q^{-1})$ is stable. The parameter estimation errors $\tilde{h}_{t+k|t}$ must have bounded variance even for drifting parameters, e.g., due to $D(z^{-1})$ in (27) having zeros on $|z| = 1$. This is guaranteed if and only if $\boldsymbol{\mathcal{L}}_k(q^{-1})$ is stable and marginally stable factors of $D(z^{-1})$ in (27) are canceled by all elements of the lag error matrix

$$\mathbf{I} - q^{-k}\boldsymbol{\mathcal{L}}_k(q^{-1})\mathbf{R}. \tag{40}$$

This is guaranteed by the Wiener design of [6] and [7]. $\square$

A natural question is now if and when the feedback noise can be neglected. This will depend on the structure and the tuning of the adaptation algorithm. A general guideline for Wiener-designed constant gain algorithms is given by Lemma 2. It considers the situation when a sequence of Wiener designs is performed based on the (not necessarily good) approximation $\eta_t = \varphi_t v_t$ for a sequence of problems in which the noise- and parameter spectral densities (29) and (30) differ only by the scaling factor λ_e. Introduce the following relative mean square approximation error:

$$\frac{\|\tilde{h}_{t+k|t} - \tilde{h}_{t+k|t}^a\|_2^2}{\|\tilde{h}_{t+k|t}^a\|_2^2} = \frac{\| - \boldsymbol{\mathcal{L}}_k(q^{-1})(\mathbf{Z}_t \tilde{h}_{t|t-1})\|_2^2}{\|\tilde{h}_{t+k|t}^a\|_2^2} \tag{41}$$

where

$$\tilde{h}_{t+k|t}^a = \left(\mathbf{I} - q^{-k}\boldsymbol{\mathcal{L}}_k(q^{-1})\mathbf{R}\right)h_{t+k} - \boldsymbol{\mathcal{L}}_k(q^{-1})(\varphi_t v_t) \tag{42}$$

is an approximation of the tracking error (19) that neglects the feedback noise. While the resulting estimation error is $\tilde{h}_{t+k|t}$, it would be $\tilde{h}_{t+k|t}^a$ if the design assumption $\eta_t = \varphi_t v_t$ was true.

Lemma 2: Assume a zero mean parameter vector h_t described by (30), with $D(z^{-1})$ stable. Let a Wiener design of the learning filter be performed, based on the assumption that $\eta_t = \varphi_t v_t$, with the spectrum of $\varphi_t v_t$ described by (29), with $\boldsymbol{M}(z^{-1})$ stable and nonsingular on $|z| = 1$ and $\bar{\mathbf{R}}_\nu$ nonsingular. Under Assumptions 1, and with bounded regressors φ_t, the relative approximation error (41) will then tend to zero as $\lambda_e/\lambda_\nu \to 0$.

Proof: See Appendix A $\square$

Thus, for optimally adjusted Wiener designed adaptation laws, the feedback noise becomes negligible when $\lambda_e/\lambda_\nu \to 0$. This can be interpreted as a situation with either slowly varying parameters h_t or with a high variance of v_t. In other words, we have a small parameter-drift-to-noise ratio. We may then substitute $\varphi_t \varphi_t^*$ by its average $\mathbf{R}$ in Fig. 2, and the resulting performance analysis via Theorem 1 will provide a small relative error (41). For misadjusted algorithms or algorithms with inappropriate structure, it may very well be the case that the estimation error is so large that the feedback noise in (15) cannot be neglected in situations where it would be negligible for a well-adjusted Wiener design. This is illustrated in Ex-
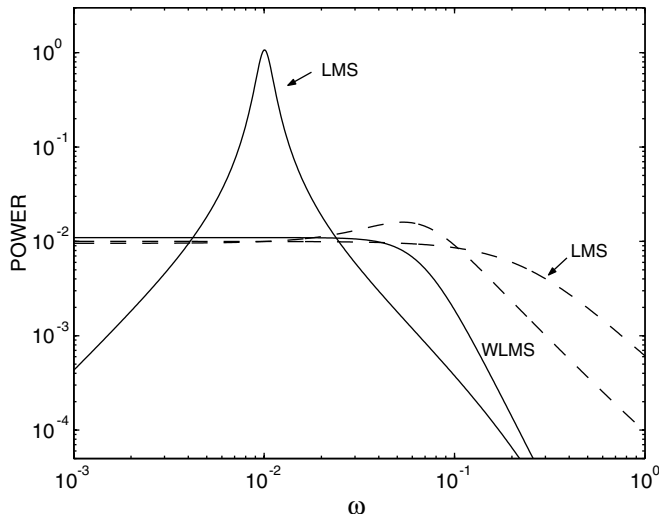
Fig. 3. Spectral density of the lag error or integrand of (38) $\|(\mathbf{I} - e^{-j\omega}\mathcal{L}_1(\omega))(\sqrt{\lambda_e}/D(\omega))\|_2^2$ (solid) and the noise term or integrand of (39) $0.01\|\mathcal{L}_1(\omega)\|_2^2$ (dashed), for Wiener estimators (WLMS), and for LMS, for a FIR system with two taps varying as AR2-processes with $\omega_o = 0.01$ in (36).

ample 1. The degree of approximation will thus be algorithm dependent.

A small parameter-drift-to-noise ratio is a situation sometimes referred to as slow variations, measured by various indicators of "the degree of nonstationarity" (DNS) [12], [19], [22], [30]. In the book [19] by Macchi (see also [8] and [12] for a similar definition), the DNS is for the purpose of LMS analysis characterized by the quantity

$$\sqrt{\frac{\mathrm{E}\|\varphi_t^*(h_t - h_{t-1})\|_2^2}{\mathrm{E}|v_t|^2}}. \tag{43}$$

Parameter variations are considered slow if this quantity is always small. For vanishing parameter-drift-to-noise ratios $\lambda_e/\lambda_\nu \to 0$, the variations will be slow according to (43).[1]

A low DNS by (43) is thus related, but not identical, to the algorithm dependent property that the feedback noise can be neglected in the performance analysis. This relation is illustrated below.

Example 1: The validity of expression (37) as an approximation of $\mathbf{P}_k$ and the applicability of (43) will be investigated for the FIR system presented in Section II-E and for regressors $u_t \in \{1, -1, i, -i\}$. We first consider a system with $M = 2$ taps. One-step prediction estimates $\hat{h}_{t+1|t}$ are obtained by a WLMS law (34), tuned to the dynamics (36). The design equations are given by (67)–(69) in Example 2. We compare this with LMS (17), with step-size μ tuned to optimize the simulated performance.

In these designs with $\mathbf{R} = \mathbf{I}$ and $\Phi_\eta = \mathbf{R}\mathrm{E}|v_t|^2 = 0.01\mathbf{I}$, the integrands of (38) and (39) will become diagonal 2×2 matrices, with equal spectral densities along the diagonals. Fig. 3 displays these spectral densities for $\omega_o = 0.01$ in (36). Note the peak

TABLE I
ASYMPTOTIC TRACKING ERROR WHEN SECOND-ORDER FIR MODELS (35), (36) WITH VARIOUS ω_o ARE TRACKED BY LMS AND WLMS. THEORETICAL PREDICTIONS tr($\mathbf{P}_{1,\mathrm{slow}}$) FROM (37) (BOLD) ARE COMPARED WITH SIMULATIONS OF tr($\mathbf{P}_1$) (ITALICS). THE LARGEST NEGLECTED TERM tr($\mathbf{V}_{Z\tilde{h}}^1$) OBTAINED FROM (24) IS ESTIMATED BY SIMULATION

	eqn	$\omega_o = .005$	0.01	0.02	0.10
DNS:	(43)	.0510	.1005	.2002	.9996
LMS:					
tr($\mathbf{P}_{1,\mathrm{slow}}$)	(37)	**.0027**	**.0045**	**.0075**	**.0360**
tr($\mathbf{P}_1$)	(20)	*.0030*	*.0052*	*.0099*	*.0650*
tr($\mathbf{V}_{Z\tilde{h}}^1$)	(24)	*.0003*	*.0007*	*.0020*	*.0278*
WLMS:					
tr($\mathbf{P}_{1,\mathrm{slow}}$)	(37)	**.0013**	**.0019**	**.0028**	**.0061**
tr($\mathbf{P}_1$)	(20)	*.0014*	*.0021*	*.0031*	*.0076*
tr($\mathbf{V}_{Z\tilde{h}}^1$)	(24)	*.0001*	*.0002*	*.0003*	*.0015*

TABLE II
ASYMPTOTIC TRACKING ERROR tr($\mathbf{P}_{1,\mathrm{slow}}$) WHEN FIR MODELS (35) AND (36) WITH $\omega_o = 0.005$ OF VARIOUS ORDERS M ARE TRACKED BY WLMS. THEORETICAL PREDICTIONS FROM (37) (BOLD) ARE COMPARED WITH SIMULATIONS (ITALICS). THE LARGEST NEGLECTED TERM tr($\mathbf{V}_{Z\tilde{h}}^1$) IS ALSO SHOWN

(WLMS)	eqn	M = 2	M = 4	M = 10	M = 20
tr($\mathbf{P}_{1,\mathrm{slow}}$)	(37)	**.0013**	**.0022**	**.0044**	**.0078**
tr($\mathbf{P}_1$)	(20)	*.0014*	*.0025*	*.0057*	*.0116*
tr($\mathbf{V}_{Z\tilde{h}}^1$)	(24)	*.0001*	*.0003*	*.0015*	*.0039*

of the LMS lag error around ω_o and the contribution of high-frequency noise to the LMS error spectrum.

We now vary ω_o in (36). Table I compares tr($\mathbf{P}_{1,\mathrm{slow}}$) from (37) (bold figures) with corresponding estimates of tr($\mathbf{P}_1$) obtained by simulation over 100 000 data (italic figures). Note the much lower tracking error of the Wiener design as compared with LMS.

The term tr($\mathbf{V}_{Z\tilde{h}}^1$) = $\lim_{t\to\infty}$ tr($\mathbf{V}_{Z\tilde{h},t}^1$), which is the largest term due to the feedback noise in (21) is also measured. This term essentially explains the difference between the expression (37), which neglects the feedback noise, and the measured performance. Let 10% represent a significant deviation of tr($\mathbf{P}_{1,\mathrm{slow}}$) relative to the true tracking error tr($\mathbf{P}_1$). For LMS, (37) then predicts the performance reasonably well for ω_o up to 0.005. For the Wiener design, with better tracking ability and thus less feedback noise, the performance is well predicted up to $\omega_o = 0.02$. A tuning based on neglecting feedback effects will here provide adequate performance in a wider set of circumstances. These differences are not captured by the DNS (43).

We now fix ω_0 to 0.0050 and increase the number of estimated parameters M, whereas λ_e is adjusted so that the output SNR remains at 20 dB, with a DNS by (43) remaining fixed at the small value 0.0510. However, the result for one-step prediction with WLMS in Table II shows that the feedback noise grows in significance with an increasing M and can no longer be considered negligible for $M \geq 4$.

---

[1]More precisely, (43) will become small when $\lambda_e/\lambda_\nu \to 0$ if the system generating h_t in (27) is stable, except for at most one integrator in $1/D(q^{-1})$. (Integrators are, however, not covered by Lemma 2.)

IV. TIME-VARYING FEEDBACK

Our aim will now be to obtain closed-form expressions that include the impact of the outer time-varying feedback loop and derive conditions for convergence in MSE. The analysis is restricted to scalar measurements y_t.

A. Independent Gaussian Regression Vectors

We first use an assumption common in LMS analysis (see, e.g., [24] and [25]) and apply it to the more general algorithm structure (5).

Assumption 3: Consecutive regression vectors $\varphi_t^*, \varphi_\tau^*$ are independent, with circular Gaussian elements, and $\mathbf{R}$ is a Toeplitz matrix with r_{j-i} as the element (i, j). ☐

We then obtain the following properties.

Lemma 3: By Assumptions 1 and 3, $\mathbf{Z}_t$ and $\tilde{h}_{t|t-1}$ are independent. Then, the feedback noise $\mathbf{Z}_t\tilde{h}_{t|t-1}$ is white with zero mean, and it is uncorrelated with h_τ and $\varphi_\tau v_\tau$ for all τ. ☐

Proof: See Appendix B ☐

Let the parameter vector h_t be tracked by an estimator (10) with asymptotically stable $\mathcal{L}_k(q^{-1})$, resulting in stationary lag errors $(\mathbf{I} - q^{-k}\mathcal{L}_k\mathbf{R})h_{t+k}$ with zero mean. Then, the error $\tilde{h}_{t+k|t}$ given by (19) will have zero mean since φ_t and v_t have zero mean by Assumption 1, and $\mathbf{Z}_t\tilde{h}_{t|t-1}$ has zero mean by Lemma 3. Furthermore, the two last terms in (21) are zero. Thus, the asymptotic covariance matrix, if it exists, is

$$\mathbf{P}_k = \lim_{t\to\infty} \mathbf{P}_{t+k|t} = \lim_{t\to\infty} \left(\mathbf{V}_{h,t}^k + \mathbf{V}_{\varphi v,t}^k + \mathbf{V}_{Z\tilde{h},t}^k \right) \quad (44)$$

and since $\mathbf{Z}_t\tilde{h}_{t|t-1}$ is uncorrelated with zero mean

$$\mathbf{V}_{Z\tilde{h},t}^k = \sum_{i=0}^t \mathbf{L}_{k,i}\mathbf{P}_{Z\tilde{h},t-i}\mathbf{L}_{k,i}^* \quad (45)$$

where

$$\mathbf{P}_{Z\tilde{h},t-i} = \mathrm{E}\left\{ \mathbf{Z}_{t-i}\tilde{h}_{t-i|t-i-1}\tilde{h}_{t-i|t-i-1}^*\mathbf{Z}_{t-i}^* \right\}.$$

In (44), $\mathbf{V}_{h,t}^k$ and $\mathbf{V}_{\varphi v,t}^k$ are here assumed bounded and asymptotically time-invariant. The conditions for this were discussed in Remark 4 in Section III. Using (37), we may then express (44) as

$$\mathbf{P}_k = \mathbf{P}_{k,\text{slow}} + \lim_{t\to\infty} \mathbf{V}_{Z\tilde{h},t}^k. \quad (46)$$

By Assumptions 1 and 3, $\mathbf{Z}_t$ and $\tilde{h}_{t|t-1}$ are independent; therefore

$$\begin{aligned} \mathbf{P}_{Z\tilde{h},t} &= \mathrm{E}\left\{ \mathbf{Z}_t\tilde{h}_{t|t-1}\tilde{h}_{t|t-1}^*\mathbf{Z}_t^* \right\} = \mathrm{E}\mathbf{Z}_t\mathbf{P}_{t|t-1}\mathbf{Z}_t^* \\ &= \mathrm{E}\varphi_t\varphi_t^*\mathbf{P}_{t|t-1}\varphi_t\varphi_t^* - \mathbf{R}\mathbf{P}_{t|t-1}\mathbf{R} \\ &= \mathbf{R}\mathrm{tr}(\mathbf{R}\mathbf{P}_{t|t-1}) \end{aligned} \quad (47)$$

where the second last equality follows from direct use of (14) and the last from [23, Sec. 9.6].

The main result of this section can now be presented.

Theorem 2: Consider a scalar linear regression (1), and let h_t be estimated by (9) and (10) with asymptotically stable $\mathcal{L}_k(q^{-1})$, resulting in stationary lag errors $(\mathbf{I} - q^{-k}\mathcal{L}_k\mathbf{R})h_{t+k}$

with zero mean and finite second-order moments. Under Assumptions 1 and 3, the tracking error $\tilde{h}_{t+k|t}$ and the feedback noise $\mathbf{Z}_t\tilde{h}_{t|t-1}$ will then have bounded covariance matrices and be asymptotically stationary if and only if the linear time-invariant scalar system

$$\mathcal{K}(z^{-1}) \triangleq \frac{1}{1 - \sum_{i=0}^\infty \mathrm{tr}\left(\mathbf{R}\mathbf{L}_{1,i}\mathbf{R}\mathbf{L}_{1,i}^* \right) z^{-i-1}} \quad (48)$$

is asymptotically stable, where $\{\mathbf{L}_{1,i}\}$ are impulse response coefficient matrices of the one-step prediction learning filter $\mathcal{L}_1(q^{-1})$. The steady-state tracking error covariance matrix is given by (46), with

$$\lim_{t\to\infty} \mathbf{V}_{Z\tilde{h},t}^k = \mathcal{K}(z^{-1})|_{z=1}\mathrm{tr}(\mathbf{R}\mathbf{P}_{1,\text{slow}}) \sum_{i=0}^\infty \mathbf{L}_{k,i}\mathbf{R}\mathbf{L}_{k,i}^*. \quad (49)$$

Proof: See Appendix C. ☐

The expressions in Theorem 2 can be used for the adaptive algorithm (4) and (5) with arbitrary $\mathcal{M}_k(q^{-1})$, and it holds for arbitrary lags k. The stability condition involving (48) will limit the gain of $\mathcal{L}_1(e^{j\omega})$. Stability of the scalar transfer function $\mathcal{K}(z^{-1})$ can be verified, for example, via the Nyquist criterion.

B. FIR Systems With White Inputs

Consider now (1) being an FIR system with scalar y_t and regressor vector (35) of length M. We will assume the input data u_t to be uncorrelated (white) so that $\mathrm{E}u_t u_\tau^* = 0$ for $t \neq \tau$ and to have zero mean, with variance $\mathrm{E}|u_t|^2 = \sigma_u^2$. Hence, $\mathbf{R} = \sigma_u^2\mathbf{I}$. The learning filter will in this subsection be constrained to be diagonal

$$\mathcal{L}_k(q^{-1}) = \sum_{i=0}^\infty L_{k,i}\mathbf{I}q^{-i} \quad (50)$$

where $L_{k,i}$ are scalars. Such learning filters appear, for example, in the LMS algorithm in (17) and the WLMS algorithm in (34) with white regressors u_t.

We here also specialize to

$$\mathrm{tr}(\mathbf{P}_k) = \lim_{t\to\infty} \mathrm{E}\|\tilde{h}_{t+k|t}\|_2^2 \quad (51)$$

which is the sum of mean square FIR-tap estimation errors, as the criterion to be evaluated.

The feedback noise will be correlated for $M > 2$. However, its trace remains uncorrelated under certain circumstances. This simplified an analysis of the properties of $\mathrm{tr}(\mathbf{P}_k)$ considerably. Expressions will be derived for $\mathrm{tr}(\mathbf{P}_k)$ under the following approximations.

Approximation 1:

$$\mathrm{tr}\left\{ \mathrm{E}\mathbf{Z}_\tau^*\mathbf{Z}_t\tilde{h}_{t|t-1}\tilde{h}_{\tau|\tau-1}^* \right\} = \mathrm{tr}\left\{ \mathrm{E}\mathbf{Z}_\tau^*\mathbf{Z}_t\mathrm{E}\tilde{h}_{t|t-1}\tilde{h}_{\tau|\tau-1}^* \right\}. \quad (52)$$

Approximation 2: The feedback noise $\mathbf{Z}_t\tilde{h}_{t|t-1}$ is uncorrelated with $\varphi_\tau v_\tau$ and h_τ, $\forall \tau$.

Under Approximation 2, the cross terms (25) and (26) in (21) are neglected. Approximations 1 and 2 make it possible to base the analytical expressions on fourth-order moment properties.

As shown by Lemma 4 in Appendix D, they hold exactly for FIR systems of order $M \leq 2$ when the regressors u_t have constant modulus. Those conditions coincide with the channel properties in the North American Digital Cellular system IS-136 [31].[2] They also hold when tracking each RAKE finger separately in, e.g., the WCDMA system, with 4-QAM modulated data.

Approximations 1 and 2 do not hold exactly for $M > 2$, but they are also good approximations for FIR systems of high order with white zero mean regressors.

An assumption of independent regression vectors would *imply* (52), but such an assumption is not *equivalent* to (52); it would place unnecessarily strong restrictions on the statistics. Independence will furthermore not apply in some cases when Approximations 1–2 hold exactly, such as under the conditions of Lemma 4 in Appendix D.

Explicit expressions for $E\mathbf{Z}_t\mathbf{Z}_\tau^*$ are now derived.

Lemma 5: Consider the FIR system (1) and (35) of order M, having as input stationary white data u_t with zero mean and variance σ_u^2. Then, we have the following.

- For circular complex valued regressors, the covariance function of $\mathbf{Z}_t$ is given by

$$E\mathbf{Z}_t\mathbf{Z}_\tau^* = \sigma_u^4(\kappa_u + M - 2)\mathbf{I}\delta_{t,\tau} \qquad (53)$$

where $\delta_{t,\tau}$ is the Kronecker delta, and κ_u is the Pearson kurtosis of u_t, which is defined for zero mean variables u_t as[3]

$$\kappa_u \triangleq \frac{E|u_t|^4}{(E|u_t|^2)^2}. \qquad (54)$$

- For real-valued regressors

$$E\mathbf{Z}_t\mathbf{Z}_t^* = \sigma_u^4(\kappa_u + M - 2)\mathbf{I} \qquad (55)$$

and for $t \neq \tau$

$$(E\mathbf{Z}_t\mathbf{Z}_\tau^*)_{ij} = 0 \quad (i,j \leq 2)$$
$$(E\mathbf{Z}_t\mathbf{Z}_\tau^*)_{ij} = \sigma_u^4 \ i - j = 2(t - \tau), \quad 2 < i,j \leq M \qquad (56)$$

where $(\cdot)_{ij}$ denotes the ijth element of $E\mathbf{Z}_t\mathbf{Z}_\tau^*$.

Proof: See Appendix E □

Thus, $\mathbf{Z}_t$ is a white sequence if u_t is white and circular complex. This is, however, not true for real-valued regressors unless $M \leq 2$, as illustrated by (56). When $\mathbf{Z}_t$ is white, the trace of the feedback noise correlation will be "white" under Approximation 1, i.e., (52) will be zero for $t \neq \tau$.

The main result of the section can now be presented.

Theorem 3: Under Assumption 1, consider the FIR system (1) and (35) with white input data u_t that are either circular

complex for arbitrary FIR degree M or real-valued, for $M \leq 2$. Let the parameter vector h_t be estimated by (9) and (10) with a stable $\mathcal{L}_k(q^{-1})$ having the structure (50), resulting in stationary lag errors $(\mathbf{I} - q^{-k}\mathcal{L}_k\mathbf{R})h_{t+k}$ with zero mean and finite second-order moments. Under Approximations 1 and 2, a finite steady-state mean square parameter error (51) then exists if and only if

$$\mathcal{G}(z^{-1}) = \frac{1}{1 - m\sigma_u^4 \sum_{i=0}^{\infty} |L_{1,i}|^2 z^{-i-1}} \qquad (57)$$

is stable, where

$$m \triangleq \kappa_u + M - 2. \qquad (58)$$

The k-step estimation error is given by

$$\text{tr}(\mathbf{P}_k) = \text{tr}(\mathbf{V}_h^k) + \text{tr}(\mathbf{V}_{\varphi v}^k) + \text{tr}(\mathbf{V}_{Z\tilde{h}}^k) \qquad (59)$$

where

$$\text{tr}(\mathbf{V}_h^k) = \left\| (\mathbf{I} - q^{-k}\mathcal{L}_k(q^{-1})\sigma_u^2)h_{t+k} \right\|_2^2 \qquad (60)$$
$$\text{tr}(\mathbf{V}_{\varphi v}^k) = M\sigma_v^2\sigma_u^2\Sigma_k \qquad (61)$$
$$\text{tr}(\mathbf{V}_{Z\tilde{h}}^k) = m\sigma_u^4\text{tr}(\mathbf{P}_1)\Sigma_k \qquad (62)$$

in which

$$\Sigma_k \triangleq \sum_{i=0}^{\infty} |L_{k,i}|^2 \qquad (63)$$

$$\text{tr}(\mathbf{P}_1) = \frac{\text{tr}(\mathbf{V}_h^1) + \text{tr}(\mathbf{V}_{\varphi v}^1)}{1 - m\sigma_u^4\Sigma_1}. \qquad (64)$$

Here, κ_u is defined as in (54), whereas $\sigma_u^2 = E|u_t|^2$ and $\sigma_v^2 = E|v_t|^2$.

Proof: See Appendix F □

Note the presence of the FIR order M in (58). As a consequence, the allowed gain of $\mathcal{L}_k(q^{-1})$, which is determined by (57), decreases as M increases. In addition, note the dependence of m on the kurtosis κ_u. The impact of M and of the distribution of the regressors on adaptation performance has been noted by Gardner in [21] and [22].

In the particular case of considering WLMS tracking with the learning filter (34), the terms (60) and (63) can be expressed as

$$\text{tr}(\mathbf{V}_h^k) = \left\| \frac{\beta(q^{-1}) - q^{-k}Q_k(q^{-1})}{\beta(q^{-1})}h_{t+k} \right\|_2^2 \qquad (65)$$

$$\Sigma_k = \frac{1}{2\pi\jmath} \oint_{|z|=1} \left| \frac{1}{\sigma_u^2} \frac{Q_k(z^{-1})}{\beta(z^{-1})} \right|^2 \frac{dz}{z}. \qquad (66)$$

Wiener design of constant-gain adaptation laws will in general have to be performed iteratively, since it depends on the variance and color of the gradient noise (29), which, by (15), will depend on the properties of the estimation error [7]. Theorem 3 can be utilized to simplify the iterative Wiener design of WLMS algorithms. The contribution of the feedback noise to the gradient noise variance can be computed analytically; therefore, an estimation by Monte Carlo simulation can be avoided. Performance and robustness for various mismatched designs can also be investigated by Theorem 3; see [31] and [32].

---

[2]Coherently demodulated signals and fractionally spaced equalization, with two samples per symbol period, will provide good detection performance, although the optimal sampling phase may vary with time over a TDMA slot. The fractionally spaced method can here be regarded as providing two symbol spaced sequences, each separated by a half symbol period and each modeled by a two-tap channel. Two trackers, operating on each data sequence, can be used to track the taps.

[3]For constant modulus data, $\kappa_u = 1$ for circular complex Gaussian data $\kappa_u = 2$, and for real-valued Gaussian regressors, $\kappa_u = 3$. For real-valued stochastic variables, $\kappa_u - 3$ is also known as the Fisher kurtosis, or normalized kurtosis.

C. Illustration of Theorem 3

Example 2—Wiener LMS Performance: The parameters of the FIR system discussed in Section II-E, with regressors $u_t \in \{-1, 1, -i, i\}$, are estimated. We here select $\omega_o = 0.1$, which corresponds to a case where the feedback noise is significant, see Table I. The degree of nonstationarity as measured by (43) is 0.9996. The one-step prediction learning filter of the WLMS law (34) for $k = 1$ is tuned to the dynamics of (36). Since $\sigma_u^2 = 1$, $\mathcal{L}_1(q^{-1})$, and $\mathcal{M}_1(q^{-1})$ can be expressed as

$$\mathcal{L}_1(q^{-1}) = \frac{Q_1(q^{-1})}{\beta(q^{-1})}\mathbf{I}, \quad \mathcal{M}_1(q^{-1}) = \frac{Q_1(q^{-1})}{D(q^{-1})}\mathbf{I}. \quad (67)$$

Here, $Q_1(q^{-1})$ and $\beta(q^{-1}) = D(q^{-1}) + q^{-1}Q_1(q^{-1})$ (cf. [6, (34), (48), and (C.4)]) are given by

$$Q_1(q^{-1}) = \mu\left(\frac{-d_1}{1 + d_2(1 - \mu)} - d_2 q^{-1}\right) \quad (68)$$

$$\beta(q^{-1}) = 1 + \frac{d_1(1 + d_2)(1 - \mu)}{1 + d_2(1 - \mu)}q^{-1} + d_2(1 - \mu)q^{-2} \quad (69)$$

where d_1 and d_2 are obtained for second-order AR statistics (36) as $d_1 = -2p\cos\omega_o, d_2 = p^2$. A scalar step-size parameter μ, which should be $0 \le \mu < 1$ [6], is tuned to minimize $\text{tr}(\mathbf{P}_1)$ in (64).

Table III displays terms contributing to the minimal criterion value obtained from Theorem 3 (bold). Feedback noise-related terms are also estimated by simulation over 100 000 data (italics). The two cross-terms (25) and (26), which are neglected by Theorem 3 are seen to indeed be small as compared with $\text{tr}(\mathbf{V}_{Z\tilde{h}}^1)$. The validity of Approximation 1 is also investigated. The relative difference between right- and left-hand sides in (52), called the error in (52) in Table II, is below 7%. The deviation between the theoretical and the measured performance peaks around $M = 10$.

Fig. 4 displays the tracking performance $\text{tr}(\mathbf{P}_k)$ for different lags k from Theorem 3 ($\circ$) and by simulations ($*$). The step-size μ is tuned to minimize $\text{tr}(\mathbf{P}_1)$, and the polynomials $Q_k(q^{-1})$ are obtained from [6, Cor. 1]. (The performance for $k = 1$ can be found in Table III.) It is evident that the simulated results are in agreement with the theory of Theorem 3 for $M = 2$ and that (59) is a good approximation of $\text{tr}(\mathbf{P}_k)$, even for higher order FIR systems like $M = 10$. It is also evident from Fig. 4 that fixed-lag smoothing (k negative) substantially improves the tracking performance and should therefore be used in applications where the use of a delay is acceptable.

The bottom line is that Table III and Fig. 4 indicate a good agreement between the theory and measurements, with Theorem 3 giving a small overestimate of $\text{tr}(\mathbf{P}_k)$ for $M > 2$ in this example.

Example 3—Wiener LMS Stability: We increased the Wiener LMS gain μ in (68) and (69), until instability occurred. In Fig. 5, the FIR system (1), (35), and (36) with $\omega_o = 0.1$ and $M = 2$ is considered, and the adaptation gain is varied. Two types of real-valued inputs are used: Binary data (B) and Gaussian signals (G) with variance 1. The tracking MSE $\text{tr}(\mathbf{P}_1)$ obtained

TABLE III
CONTRIBUTIONS TO THE ASYMPTOTIC ONE-STEP PARAMETER PREDICTION ERROR $\text{tr}(\mathbf{P}_1)$ WHEN FIR MODELS WITH $\omega_o = 0.1$ OF ORDER M ARE TRACKED BY WLMS IN EXAMPLE 2. THEORETICAL PREDICTIONS FROM RELEVANT EXPRESSIONS IN THEOREM 3 (BOLD) ARE COMPARED WITH SIMULATIONS (ITALICS)

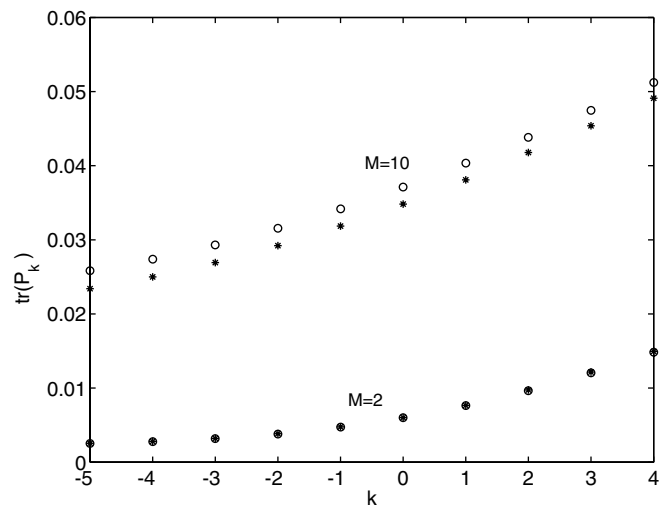| Term:	eqn	M = 2	M = 4	M = 10	M = 20
$\text{tr}(\mathbf{P}_1)$	(59)	.0076	.0153	.0403	.0785
Simulation:		*.0076*	*.0146*	*.0379*	*.0767*
$\text{tr}(\mathbf{V}_h^1)$	(60)	.0021	.0061	.0184	.0394
$\text{tr}(\mathbf{V}_{\varphi v}^1)$	(61)	.0040	.0043	.0047	.0046
$\text{tr}(\mathbf{V}_{Z\tilde{h}}^1)$	(62)	.0015	.0049	.0172	.0345
Simulation:		*.0015*	*.0042*	*.0153*	*.0305*
$-\text{tr}(\mathbf{V}_{hZ\tilde{h}}^1)$	(25)	*0*	*.0002*	*.0011*	*.0011*
$\text{tr}(\mathbf{V}_{\varphi v Z\tilde{h}}^1)$	(26)	*0*	*.0001*	*.0002*	*.0002*
Error in	(52):	*0*	*2.7%*	*6.3%*	*3.7%*

Fig. 4. Tracking performance $\text{tr}(\mathbf{P}_k)$ of WLMS for different lags k from Theorem 3 ($\circ$) and by simulations ($*$) for FIR systems with $M = 2$ and $M = 10$ taps.

from Theorem 3 is plotted as a function of μ and compared with measured values from simulation ($*$). The dashed curve is the theoretical result obtained from (37) by neglecting the feedback noise. It results in large errors and predicts stability for all admissible $0 \le \mu < 1$, instead of the correct stability limits $\mu_{\max} = 0.336$ (G) and $\mu_{\max} = 0.632$ (B). Evidently, the results obtained by simulation are in agreement with the theory of Theorem 3, but a simplified analysis based on Theorem 1 (neglecting the feedback noise) would be unsatisfactory in this example.

In the binary case (B), $\mathbf{Z}_t$, which is given by (D.1), will have singular values bounded by unity. The use of the small gain theorem (18) then gives the condition $\|e^{-j\omega}\mathcal{L}_1(\omega)\|_\infty < 1$. For (67)–(69), the largest magnitude of the spectral density of Q_1/β should thus be < 1. This is fulfilled for $\mu < 0.0529$. Hence, the small gain theorem applied to the outer loop of Fig. 1 is a too-conservative tool in this example. For Gaussian regressors, the magnitude of the elements of $\mathbf{Z}_t$ will not be bounded, and (18) cannot be used directly.
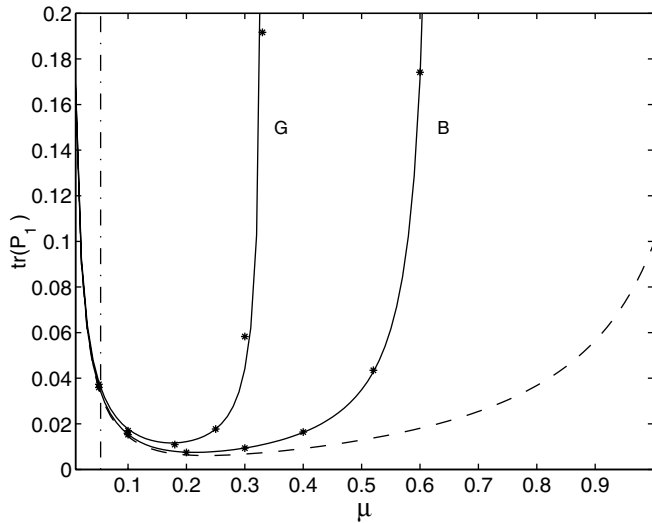
Fig. 5. Performance $\mathrm{tr}(\mathbf{P}_1)$ as a function of μ in the WLMS algorithm (67) in Example 3 when tracking a second-order FIR system (1), (35), and (36) with $\omega_o = 0.1$. Results by theory from Theorem 3 (solid) and by simulation ($*$) for white real-valued binary (B) and Gaussian (G) regressors. The dashed curve neglects the feedback noise. The vertical line at $\mu = 0.05$ is the stability bound obtained from the small gain theorem (18) for binary regressors.

The plots clearly show that input data with larger kurtosis (Gaussian data) result in larger tracking errors and in performance curves that are more sensitive to the choice of gain μ. However, one should note from (58) that the relative impact of the regressor distribution (via κ_u) diminishes with an increasing FIR order M.

V. CONCLUDING REMARKS

We have proposed a novel formalism for analysis of adaptive algorithms that perform tracking of linear regression models. The formalism was applied on general linear algorithms with constant gains.

The considered decomposition of the feedback loop can also be used as powerful tools for analyzing the tracking performance for slow time-variations of RLS algorithms with slow adaptation and of Kalman predictors. See [33] for such results.

Theorem 1 can be used for vector-valued measurements y_t and parameter variations described by general spectral densities, but it provides reasonable approximations only when the variance of the feedback noise $\mathbf{Z}_t \tilde{h}_{t|t-1}$ is small relative to the variance of $\varphi_t v_t$. The assumption of independent regressor vectors used in Theorem 2 is restrictive, but it is approximately fulfilled for the antenna array application discussed by Horowitz in [25].

An interesting insight resulting from the present work is that tracking of FIR systems with white regressors can be analyzed under approximations that are milder than an assumption of independent regression vectors, which would evidently be wrong in such situations. This analysis, which is presented in Section IV-B, was here limited to algorithms with diagonal learning filters. Theorem 3 can be generalized to the criterion (20) and to general algorithms (11), but the feedback noise will then not be uncorrelated.

A further generalization of Theorem 3 to FIR systems with colored regressors would be interesting but appears problem-

atic since whiteness of the regressors u_t is a key assumption in Lemma 5. An exact tracking analysis for fast variations with colored regressors might require considerably more complicated tools, perhaps along the lines explored by Douglas and Pan in [34] or by using the energy conserving approach of [18].

APPENDIX A
PROOF OF LEMMA 2

We will investigate how the relative approximation error (41) behaves as $\lambda_e/\lambda_\nu \to 0$. In the sequel, we set λ_ν to unity and let λ_e tend to zero. The design of $\mathcal{L}_k(q^{-1})$ is based on (31)–(33). Under the stated assumptions, the right-hand side of the spectral factorization (32) is nonsingular on the unit circle. Therefore, a causally and stably invertible polynomial matrix $\boldsymbol{\beta}(z^{-1})$ exists. Thus, $\det \boldsymbol{\beta}(z^{-1})$ has all zeros in $|z| < 1$, whereas $\det \boldsymbol{\beta}_*(z)$ has all zeros in $|z| > 1$. Furthermore, from (32)

$$\lim_{\lambda_e \to 0} \boldsymbol{\beta} = D\mathbf{R}^{-1}\boldsymbol{M}\bar{\mathbf{R}}_\nu^{1/2}; \quad \lim_{\lambda_e \to 0} \boldsymbol{\beta}_* = \bar{\mathbf{R}}_\nu^{1/2}\boldsymbol{M}_*\mathbf{R}^{-1}D_* \tag{A.1}$$

which are nonsingular on $|z| = 1$ since $\boldsymbol{M}(e^{i\omega})$ is assumed nonsingular and $\bar{\mathbf{R}}_\nu$ is nonsingular. Note that $D(z^{-1})$ is here assumed to have all zeros in $|z| < 1$ so that $D_*(z)$ will have all zeros outside $|z| = 1$.

Equation (33) is a unilateral Diophantine equation for a scalar $D(z^{-1})$. It will have a unique solution since $D(z^{-1})$ with all zeros in $|z| < 1$ and $\det \boldsymbol{\beta}_*(z)$ with zeros in $|z| > 1$ are coprime; see [7, App. A]. By equating for different powers of z and z^{-1} in (33) and transposing, a linear system of equations (A.2) is obtained, where $\mathbf{B}_{CN}$ contains the corresponding coefficient matrices of $(z^k \boldsymbol{C}(z^{-1})\bar{\mathbf{R}}_e \boldsymbol{C}_*(z)N_*(z))^T$

$$\underbrace{\begin{bmatrix} 0 & 0 & | & \mathbf{I} & 0 \\ \beta_{n_\beta}^{*T} & & | & d_1\mathbf{I} & \ddots \\ \vdots & \ddots & \beta_{n_\beta}^{*T} & | & \vdots & & \mathbf{I} \\ \beta_0^{*T} & & \vdots & | & d_{n_D}\mathbf{I} & & d_1\mathbf{I} \\ & \ddots & \vdots & | & & \ddots & \vdots \\ 0 & & \beta_0^{*T} & | & 0 & & d_{n_D}\mathbf{I} \end{bmatrix}}_{\mathbf{A}_{\beta D}} \underbrace{\begin{bmatrix} \mathbf{Q}_0^T \\ \vdots \\ \mathbf{Q}_{n_Q}^T \\ \mathbf{U}_{n_U *} \\ \vdots \\ \mathbf{U}_{0 *} \end{bmatrix}}_{\mathbf{X}_{QU}}$$
$$= \lambda_e \mathbf{B}_{CN}. \tag{A.2}$$

Previously, n_D, n_β, n_Q and n_U have denoted the degrees of the polynomial D and the polynomial matrices $\beta_*, \boldsymbol{Q}_k$ and $\boldsymbol{U}_{k*}$, respectively. Since the polynomial matrix Diophantine (33) is uniquely solvable for all λ_e, the corresponding block-Sylvester coefficient matrix $\mathbf{A}_{\beta D}$ is nonsingular for all λ_e [35]. By (A.1), it converges to a constant nonsingular matrix (involving the coefficients of $\lim_{\lambda_e \to 0} \beta_*(z)$) as $\lambda_e \to 0$. Denote this limiting matrix $\bar{\mathbf{A}}_{\beta D}$. Thus, the solution pair $(\boldsymbol{Q}_k(z^{-1}), \boldsymbol{U}_{k*}(z))$ tends to zero proportional to λ_e since

$$\mathbf{X}_{QU} = \lambda_e \mathbf{A}_{\beta D}^{-1} \mathbf{B}_{CN} \tag{A.3}$$

where $\mathbf{A}_{\beta D}^{-1} \to \bar{\mathbf{A}}_{\beta D}^{-1}$ as $\lambda_e \to 0$. By (31) and (A.1), the impulse response coefficients of the optimal learning filter will also tend

to zero proportional to λ_e when λ_e is sufficiently small so that $\beta^{-1}(z^{-1})$ is close to its limiting expression.

The matrix $\mathbf{Z}_t$ will have bounded elements since φ_t is assumed to have bounded elements. Therefore, the small gain theorem applies. Since by the above reasoning $\mathcal{L}_1(e^{j\omega}) \to 0 \ \forall \ \omega$ when $\lambda_e \to 0$, we can conclude that the feedback loop via $\mathbf{Z}_t \tilde{h}_{t|t-1}$ will be stable for Wiener-optimized learning filters designed for sufficiently small parameter drift-to-noise ratios.

With a stable filter generating h_t in (27) and a stable feedback loop, both h_t and $\hat{h}_{t|t-1}$ will be stationary processes with finite variance so that $\tilde{h}_{t|t-1}$ is stationary with finite variance.

Thus, since $\mathbf{Z}_t \tilde{h}_{t|t-1}$ is guaranteed to have finite power when λ_e is sufficiently small and since the impulse response coefficients of the optimized learning filter will, by (A.3), tend to zero proportional to λ_e, there will exist a constant $\alpha < \infty$ such that the numerator of (41) is bounded by

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr}(\mathcal{L}_k(\omega)\Phi_{Z\tilde{h}}(\omega)\mathcal{L}_{k*}(\omega))d\omega \le \lambda_e^2 \alpha \qquad \text{(A.4)}$$

where the spectral density of the feedback noise is denoted by $\Phi_{Z\tilde{h}}$.

The spectral densities of h_t and $\eta = \varphi_t v_t$ (with $\lambda_\nu = 1$) used in the design are $\lambda_e \bar{\Phi}_h$ and $\bar{\Phi}_\eta$, which are given by (30) and (29), respectively. Then

$$\|\tilde{h}_{t+k|t}^a\|_2^2 = \frac{\lambda_e}{2\pi} \int_{-\pi}^{\pi} \text{tr}((\mathbf{I} - e^{-j\omega k}\mathcal{L}_k \mathbf{R})\bar{\Phi}_h$$

$$\times (\mathbf{I} - e^{-j\omega k}\mathcal{L}_k \mathbf{R})_*)d\omega + \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr}(\mathcal{L}_k \bar{\Phi}_\eta \mathcal{L}_{k*})d\omega. \qquad \text{(A.5)}$$

Therefore, the relative approximation error vanishes as λ_e tends to zero due to the factor λ_e^2 in (A.4).

APPENDIX B
PROOF OF LEMMA 3

The estimation error $\tilde{h}_{t|t-1}$ depends only on data up to time $t - 1$, whereas $\mathbf{Z}_t$, due to the independence assumption on φ_t, will not depend on data older than t. Thus, $\mathbf{Z}_t$ and $\tilde{h}_{t|t-1}$ are independent. As a consequence, $\text{E}(\mathbf{Z}_t \tilde{h}_{t|t-1}) = \text{E}(\mathbf{Z}_t)\text{E}(\tilde{h}_{t|t-1}) = 0$. Whiteness follows since

$$\text{E}(\mathbf{Z}_t \tilde{h}_{t|t-1} \tilde{h}_{\tau|\tau-1}^* \mathbf{Z}_\tau^*) = \text{E}(\mathbf{Z}_t)\text{E}(\tilde{h}_{t|t-1} \tilde{h}_{\tau|\tau-1}^* \mathbf{Z}_\tau^*) = 0$$

for $t > \tau$ since $\mathbf{Z}_t$ will be independent of all other factors, and $\text{E}(\mathbf{Z}_t) = 0$. For $t < \tau$, $\mathbf{Z}_\tau^*$ will be independent of all other factors. Since $\mathbf{Z}_t$ depends neither on h_τ (by Assumption 1) nor on $\tilde{h}_{t|t-1}$

$$\text{E}h_\tau(\mathbf{Z}_t \tilde{h}_{t|t-1})^* = \text{E}(h_\tau \tilde{h}_{t|t-1}^*)\text{E}(\mathbf{Z}_t^*) = 0.$$

Now, consider $\text{E}\{\varphi_\tau v_\tau \tilde{h}_{t|t-1}^* \mathbf{Z}_t^*\}$, which is zero for $\tau \ne t$ since $\mathbf{Z}_t^*$ is then independent of all other factors. For $\tau = t$, v_t and $\tilde{h}_{t|t-1}^*$ are independent of the regressors; therefore

$$\text{E}\{\varphi_t v_t \tilde{h}_{t|t-1}^* \mathbf{Z}_t^*\} = \text{E}\{\varphi_t \text{E}(v_t \tilde{h}_{t|t-1}^*)\mathbf{Z}_t^*\}.$$

The outer expectation over first- and third-order moments of the Gaussian regressors is zero. Thus

$$\text{E}\{\varphi_t v_t \tilde{h}_{t|t-1}^* \mathbf{Z}_t^*\} = 0 = \text{E}\{\varphi_t v_t\}\text{E}\{\tilde{h}_{t|t-1}^* \mathbf{Z}_t^*\}.$$

APPENDIX C
PROOF OF THEOREM 2

Consider (44) for $k = 1$, without letting $t \to \infty$, as a recursion in t. This recursion must converge (convergence in MSE). The use of (47) in (45) gives

$$\mathbf{P}_{t+1|t} = \mathbf{V}_{h,t}^1 + \mathbf{V}_{\varphi v,t}^1 + \sum_{i=0}^{t} \mathbf{L}_{1,i}\mathbf{R}\text{tr}(\mathbf{R}\mathbf{P}_{t-i|t-i-1})\mathbf{L}_{1,i}^*.$$

Premultiply by $\mathbf{R}$ and take the trace to obtain

$$\text{tr}(\mathbf{R}\mathbf{P}_{t+1|t}) = \text{tr}(\mathbf{R}\mathbf{V}_{h,t}^1) + \text{tr}(\mathbf{R}\mathbf{V}_{\varphi v,t}^1)$$

$$+ \text{tr}(\mathbf{R}\sum_{i=0}^{t} \mathbf{L}_{1,i}\mathbf{R}\text{tr}(\mathbf{R}\mathbf{P}_{t-i|t-i-1})\mathbf{L}_{1,i}^*). \qquad \text{(C.1)}$$

Replacing $\mathbf{P}_{t-i|t-i-1}$ by $q^{-i-1}\mathbf{P}_{t+1|t}$ in (C.1) yields

$$\left(1 - \text{tr}\left(\mathbf{R}\sum_{i=0}^{t} \mathbf{L}_{1,i}\mathbf{R}\mathbf{L}_{1,i}^* q^{-i-1}\right)\right)\text{tr}(\mathbf{R}\mathbf{P}_{t+1|t})$$

$$= \text{tr}(\mathbf{R}\mathbf{V}_{h,t}^1) + \text{tr}(\mathbf{R}\mathbf{V}_{\varphi v,t}^1)$$

or

$$\text{tr}(\mathbf{R}\mathbf{P}_{t+1|t}) = \mathcal{K}(q^{-1})(\text{tr}(\mathbf{R}\mathbf{V}_{h,t}^1) + \text{tr}(\mathbf{R}\mathbf{V}_{\varphi v,t}^1)) \qquad \text{(C.2)}$$

where the transfer operator (48) is in the limit $t \to \infty$

$$\mathcal{K}(q^{-1}) \triangleq \frac{1}{1 - \sum_{i=0}^{\infty} \text{tr}(\mathbf{R}\mathbf{L}_{1,i}\mathbf{R}\mathbf{L}_{1,i}^*)q^{-i-1}} \qquad \text{(C.3)}$$

and it determines stability. By using the static gain, i.e., substituting $z = 1$ for q in (C.3) and using the limits $\mathbf{V}_h^1$ and $\mathbf{V}_{\varphi v}^1$ by (37) in (C.2), we obtain

$$\lim_{t\to\infty} \text{tr}(\mathbf{R}\mathbf{P}_{t+1|t}) \triangleq \text{tr}(\mathbf{R}\mathbf{P}_1) = \mathcal{K}(z^{-1})|_{z=1}\text{tr}(\mathbf{R}\mathbf{P}_{1,\text{slow}}). \qquad \text{(C.4)}$$

The expression (C.4) gives the *excess MSE* due to model misadjustment. The k-step covariance matrix is then given by (46), with

$$\lim_{t\to\infty} \mathbf{V}_{Z\tilde{h},t}^k \triangleq \mathbf{V}_{Z\tilde{h}}^k = \sum_{i=0}^{\infty} \mathbf{L}_{k,i}\mathbf{P}_{Z\tilde{h}}\mathbf{L}_{k,i}^*. \qquad \text{(C.5)}$$

Here, $\mathbf{P}_{Z\tilde{h}} = \lim_{t\to\infty} \mathbf{P}_{Z\tilde{h},t}$. For stable $\mathcal{L}_k(q^{-1})$ and bounded $\mathbf{P}_{Z\tilde{h}}$, $\mathbf{V}_{Z\tilde{h}}^k$ by (C.5) is bounded. By (47) and (C.4)

$$\mathbf{P}_{Z\tilde{h}} = \mathbf{R}\lim_{t\to\infty} \text{tr}(\mathbf{R}\mathbf{P}_{t|t-1}) = \mathbf{R}\mathcal{K}(z^{-1})|_{z=1}\text{tr}(\mathbf{R}\mathbf{P}_{1,\text{slow}}). \qquad \text{(C.6)}$$

The use of (C.6) in (C.5) then gives (49).

APPENDIX D
PROOF OF LEMMA 4

For $M = 2$ and independent inputs with constant modulus, $|u_t|^2 = \sigma_u^2$, (8) and (14) give $\mathbf{Z}_t$ as

$$\begin{pmatrix} u_t^* \\ u_{t-1}^* \end{pmatrix} (u_t \quad u_{t-1}) - \sigma_u^2 \mathbf{I} = \begin{pmatrix} 0 & u_t^* u_{t-1} \\ u_{t-1}^* u_t & 0 \end{pmatrix}. \quad \text{(D.1)}$$

Denote the elements of $\tilde{h}_{t|t-1}$ by $\tilde{h}_{t-1}^\ell, \ell = 0, 1$. The mean value of $\mathbf{Z}_t \tilde{h}_{t|t-1}$ becomes

$$\mathrm{E} \begin{pmatrix} u_t^* u_{t-1} \tilde{h}_{t-1}^1 \\ u_t u_{t-1}^* \tilde{h}_{t-1}^0 \end{pmatrix} = 0$$

since u_t, with zero mean, will be independent of u_{t-1} and of $\tilde{h}_{t-1}^\ell$.

Let $z_t \triangleq u_t^* u_{t-1}$. The correlation function of the feedback noise now follows as

$$\begin{aligned}
&\mathrm{E} \mathbf{Z}_{t-i} \tilde{h}_{t-i|t-i-1} \tilde{h}_{t-j|t-j-1}^* \mathbf{Z}_{t-j}^* \\
&= \mathrm{E} \begin{pmatrix} z_{t-i} \tilde{h}_{t-i-1}^1 \\ z_{t-i}^* \tilde{h}_{t-i-1}^0 \end{pmatrix} \begin{pmatrix} z_{t-j} \tilde{h}_{t-j-1}^1 \\ z_{t-j}^* \tilde{h}_{t-j-1}^0 \end{pmatrix}^* \\
&= \mathrm{E} \begin{pmatrix} \tilde{h}_{t-i-1}^1 \tilde{h}_{t-j-1}^{1*} z_{t-i} z_{t-j}^* & \tilde{h}_{t-i-1}^1 \tilde{h}_{t-j-1}^{0*} z_{t-i} z_{t-j} \\ \tilde{h}_{t-i-1}^0 \tilde{h}_{t-j-1}^{1*} z_{t-i}^* z_{t-j}^* & \tilde{h}_{t-i-1}^0 \tilde{h}_{t-j-1}^{0*} z_{t-i}^* z_{t-j} \end{pmatrix}.
\end{aligned} \quad \text{(D.2)}$$

If $i < j$, then u_{t-i}^* in $z_{t-i} = u_{t-i}^* u_{t-i-1}$ (or u_{t-i} in z_{t-i}^*) will be independent of all other factors of each product, and if $i > j$, then u_{t-j}^* is independent of all other factors. Thus, the feedback noise is white. In addition, note that for u_t without constant modulus, each element of (D.2) would have four terms, all of which would not be zero when $i \neq j$. The feedback noise would then not be white.

For $i = j$, the z-related factors in (D.2) are either $(|u_t|^2)^2 = \sigma_u^4$, due to the constant modulus, or zero. In the case of real-valued input data and models, the covariance matrix of the feedback noise becomes

$$\sigma_u^4 \mathrm{E} \begin{pmatrix} \tilde{h}_{t-1}^1 \tilde{h}_{t-1}^1 & \tilde{h}_{t-1}^1 \tilde{h}_{t-1}^0 \\ \tilde{h}_{t-1}^0 \tilde{h}_{t-1}^1 & \tilde{h}_{t-1}^0 \tilde{h}_{t-1}^0 \end{pmatrix} = \sigma_u^4 \begin{pmatrix} p_{t|t-1}^{22} & p_{t|t-1}^{12} \\ p_{t|t-1}^{12} & p_{t|t-1}^{11} \end{pmatrix}.$$

For circular complex data, factors $z_t z_t$ in (D.2) for $i = j$ contain $u_t^* u_t^*$, which is independent of all other factors. Since $u_t^* u_t^* = u_t u_t = 0$ in the circular case, off-diagonal elements of (D.2) are zero.

That $\mathbf{Z}_t \tilde{h}_{t|t-1}$ is uncorrelated with h_τ follows since u_t in $\mathbf{Z}_t$ in (D.1) is independent of h_τ as well as of $\tilde{h}_{t|t-1}$. Uncorrelatedness between $\mathbf{Z}_t \tilde{h}_{t|t-1}$ and $\varphi_\tau v_\tau$ is true for $\tau \geq t$ due to whiteness and zero mean of v_t and its independence of all other terms. For $\tau < t$

$$\mathrm{E}\{\varphi_\tau v_\tau \tilde{h}_{t|t-1}^* \mathbf{Z}_t^*\} = \mathrm{E}\{\varphi_\tau v_\tau (u_t u_{t-1}^* \tilde{h}_{t-1}^{1*} u_t^* u_{t-1} \tilde{h}_{t-1}^{0*})\}$$

vanishes since u_t is zero mean and independent of all other terms.

APPENDIX E
PROOF OF LEMMA 5

For FIR systems having stationary and white inputs with zero mean and variance σ_u^2

$$\mathrm{E} \mathbf{Z}_t \mathbf{Z}_\tau^* = \mathrm{E} \varphi_t \varphi_t^* \varphi_\tau \varphi_\tau^* - \sigma_u^4 \mathbf{I}. \quad \text{(E.1)}$$

Let the ijth element of $\mathrm{E} \varphi_t \varphi_t^* \varphi_\tau \varphi_\tau^*$ be denoted by x^{ij}. Then, with $\varphi_t^* = (u_t u_{t-1} \ldots u_{t-M+1})$, this element can be expressed as

$$x^{ij} = \sum_{\ell=1}^M \mathrm{E} u_{t-i+1}^* u_{t-\ell+1} u_{\tau-\ell+1}^* u_{\tau-j+1} \quad \text{(E.2)}$$

for $i, j = 1, \ldots, M$. For equally distributed zero mean variables u_t, fourth-order moments are expressed by

$$\begin{aligned}
\mathrm{E} u_i^* u_j u_k^* u_\ell &= \mathrm{E} u_i^* u_j \mathrm{E} u_k^* u_\ell + \mathrm{E} u_i^* u_k^* \mathrm{E} u_j u_\ell + \mathrm{E} u_i^* u_\ell \\
&\quad \times \mathrm{E} u_j u_k^* + \frac{1}{\sigma_u^2} (\kappa_u - \kappa_u^G) \mathrm{E} u_i^* u_j \mathrm{E} u_k^* u_\ell \mathrm{E} u_j u_k^* \quad \text{(E.3)}
\end{aligned}$$

see, e.g., [36, p. 549]. Here, $\kappa_u = \mathrm{E}|u_t|^4/(\sigma_u^2)^2$ is defined in (54), whereas κ_u^G is the kurtosis of Gaussian variables. $\kappa_u^G = 2$ is defined for circular complex variables, whereas $\kappa_u^G = 3$ is defined for real variables. The second right-hand term is zero for white circular variables, where $\mathrm{E} u_i u_j = 0 \, \forall \, i, j$. The last term vanishes for Gaussian data where $\kappa_u = \kappa_u^G$. It contributes only when $i = j = k = \ell$ and is constructed so that the right-hand side of (E.3) then equals $\mathrm{E}|u_t|^4$.

Applying (E.3) on (E.2) for white circular data (for which $\kappa_u^G = 2$) yields, with $\mathrm{E} u_i^* u_j = \sigma_u^2 \delta_{i,j}$

$$x^{ij} = \sigma_u^4 \sum_{\ell=1}^M (\delta_{i,\ell} \delta_{\ell,j} + \delta_{t-i,\tau-j} \delta_{t,\tau} + (\kappa_u - 2) \delta_{i,\ell} \delta_{\ell,j} \delta_{t,\tau}). \quad \text{(E.4)}$$

We notice that the third term within the parenthesis of (E.4) contributes to the sum only for $i = j = \ell$ and $\tau = t$, whereas the first contributes when $i = j = \ell$ and the second contributes for $\tau = t$ when $i = j$. Thus

$$x^{ij} = 0 \quad i \neq j; \quad x^{ii} = \begin{cases} \sigma_u^4 (\kappa_u + M - 1), & \tau = t \\ \sigma_u^4, & \tau \neq t. \end{cases}$$

This observation and the use of (E.1), which subtracts $\sigma_u^4 \mathbf{I}$, completes the first part of the proof. Applying the formula (E.3) on (E.2) to real-valued data ($\kappa_u^G = 3$) yields

$$\begin{aligned}
x^{ij} = \sigma_u^4 \sum_{\ell=1}^M (\delta_{i,\ell} \delta_{\ell,j} + \delta_{t-i,\tau-\ell} \delta_{t-\ell,\tau-j} + \delta_{t-i,\tau-j} \delta_{t,\tau} \\
+ (\kappa_u - 3) \delta_{i,\ell} \delta_{\ell,j} \delta_{t,\tau}). \quad \text{(E.5)}
\end{aligned}$$

Since we have real-valued input data, the second right-hand term of (E.3) also contributes. The second term within the parenthesis in (E.5) contributes to the sum whenever $i - j = 2(t - \tau)$,

whereas the first, the third, and the fourth terms contribute as for the previous case of circular data. Thus

$$x^{ii} = \begin{cases} \sigma_u^4(\kappa_u + M - 1), & \tau = t \\ \sigma_u^4, & \tau \neq t \end{cases} \qquad (\text{E.6})$$

$$\begin{aligned} x^{ij} \\ i \neq j \end{aligned} = \begin{cases} 0, & \tau = t \\ 0, & \tau \neq t \text{ and } i,j = 1,2 \\ \sigma_u^4, & i - j = 2(t - \tau) \\ & \tau \neq t, \quad 2 < i,j \leq M. \end{cases} \qquad (\text{E.7})$$

The use of (E.1) then completes the proof.

APPENDIX F
PROOF OF THEOREM 3

For colored v_t, Assumption 1 and white regressors imply

$$\text{Etr}(\varphi_\tau v_\tau v_t^* \varphi_t^*) = \text{E}(\varphi_t^* \varphi_\tau)\text{E}(v_\tau v_t^*) = M\sigma_u^2 \sigma_v^2 \delta_{t,\tau} \quad (\text{F.1})$$

since $\text{E}(\varphi_t^* \varphi_\tau) = \sum_{i=0}^{M-1} \text{E}u_{t-i}^* u_{\tau-i} = 0$ for $t \neq \tau$ and equals $M\sigma_u^2$ for $t = \tau$. Therefore

$$\begin{aligned} \text{tr}(\text{E}\eta_t \eta_\tau^*) &= \text{tr}(\text{E}(\mathbf{Z}_t \tilde{h}_{t|t-1} + \varphi_t v_t) \\ &\quad \times (\mathbf{Z}_\tau \tilde{h}_{\tau|\tau-1} + \varphi_\tau v_\tau)^*) \\ &= \text{tr}(\text{E}\mathbf{Z}_\tau^* \mathbf{Z}_t \text{E}\tilde{h}_{t|t-1} \tilde{h}_{\tau|\tau-1}^*) \\ &\quad + \text{E}\varphi_\tau^* \varphi_t \text{E}v_t v_\tau^* \end{aligned} \qquad (\text{F.2})$$

where Approximations 1 and 2 were used in the second equality. Under the stated assumptions, $\mathbf{Z}_t$ is white due to Lemma 5. Using (53), (55), (56), (F.1), and (F.2)

$$\begin{aligned} \text{tr}(\text{E}\eta_t \eta_\tau^*) &= \left(\sigma_u^4(\kappa_u + M - 2)\text{tr}(\mathbf{P}_{t|t-1}) + M\sigma_u^2 \sigma_v^2\right)\delta_{t,\tau} \\ &= \sigma_u^4 \left(m\,\text{tr}(\mathbf{P}_{t|t-1}) + M\frac{\sigma_v^2}{\sigma_u^2}\right)\delta_{t,\tau} \end{aligned} \qquad (\text{F.3})$$

where we introduced $m \triangleq \kappa_u + M - 2$ in the last equality. Since the two last terms in (21) are neglected under Approximation 2, we obtain

$$\text{tr}(\mathbf{P}_{t+1|t}) = \text{tr}(\mathbf{V}_{h,t}^1) + \text{tr}(\mathbf{V}_{\varphi v,t}^1) + \text{tr}(\mathbf{V}_{Z\tilde{h},t}^1) \qquad (\text{F.4})$$

where, from

$$\tilde{h}_{t+k|t} = (\mathbf{I} - q^{-k}\boldsymbol{\mathcal{L}}_k(q^{-1})\sigma_u^2)h_{t+k} - \boldsymbol{\mathcal{L}}_k(q^{-1})\eta_t \qquad (\text{F.5})$$

the asymptotic lag error is given by

$$\text{tr}(\mathbf{V}_h^k) = \lim_{t\to\infty}\text{tr}(\mathbf{V}_{h,t}^k) = \left\|(\mathbf{I} - q^{-k}\mathbf{L}_k(q^{-1})\sigma_u^2)h_{t+k}\right\|_2^2. \qquad (\text{F.6})$$

By using (63) and (F.1) in (23)

$$\text{tr}(\mathbf{V}_{\varphi v}^k) = \lim_{t\to\infty}\text{tr}(\mathbf{V}_{\varphi v,t}^k) = M\sigma_v^2 \sigma_u^2 \Sigma_k. \qquad (\text{F.7})$$

Note that (50) is assumed to have scalar impulse response coefficients $\text{L}_{k,i}$. By using (F.3) and the whiteness of the feedback noise in (24)

$$\text{tr}(\mathbf{V}_{Z\tilde{h},t}^1) = m\sigma_u^4 \sum_{i=0}^{t} |\text{L}_{1,i}|^2 q^{-i}\text{tr}(\mathbf{P}_{t|t-1}). \qquad (\text{F.8})$$

The use of (F.8) in (F.4) gives a scalar linear time-invariant recursion

$$\text{tr}(\mathbf{P}_{t+1|t}) = \mathcal{G}(q^{-1})(\text{tr}(\mathbf{V}_{h,t}^1) + \text{tr}(\mathbf{V}_{\varphi v,t}^1)) \qquad (\text{F.9})$$

where $\mathcal{G}(q^{-1})$ is given by (57) in the limit $t \to \infty$. Under the stated assumptions, $\text{tr}(\mathbf{P}_{t+1|t})$ will converge to a finite stationary limit $\text{tr}(\mathbf{P}_1)$ if and only if the filter $\mathcal{G}(q^{-1})$ is stable. The steady-state error $\text{tr}(\mathbf{P}_1) = \lim_{t\to\infty}\text{tr}(\mathbf{P}_{t+1|t})$ is obtained by using the steady-state gain $\mathcal{G}(1)$ (substituting $z = 1$ for q) of $\mathcal{G}(q^{-1})$ and using the steady-state values (F.6) and (F.7) for $k = 1$ in (F.9). This gives (64). Expression (62) is obtained by replacing $q^{-i}\text{tr}(\mathbf{P}_{t|t-1})$ by $\text{tr}(\mathbf{P}_1)$ in (F.8) and using (63). With (60) obtained from (F.6) and (61) from (F.7), we have (59) for $\text{tr}(\mathbf{P}_k)$.

ACKNOWLEDGMENT

REFERENCES

[1] A. Benveniste, "Design of adaptive algorithms for tracking of time-varying systems," *Int. J. Adaptive Contr. Signal Process.*, vol. 1, pp. 3–29, 1987.
[2] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations.* Berlin, Heidelberg, Germany: Springer-Verlag, 1990.
[3] L. Ljung and S. Gunnarsson, "Adaptation and tracking in system identification—A survey," *Automatica*, vol. 26, pp. 7–21, 1990.
[4] S. Roy and J. J. Shynk, "Analysis of the momentum LMS algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 2088–2098, 1990.
[5] H. S. Dabis and T. J. Moir, "Least mean squares as a control system," *Int. J. Adaptive Contr. Signal Process.*, vol. 54, pp. 321–335, 1991.
[6] L. Lindbom, M. Sternad, and A. Ahlén, "Tracking of time-varying mobile radio channels. Part I: The wiener LMS algorithm," *IEEE Trans. Commun.*, vol. 49, pp. 2207–2217, Dec. 2001.
[7] M. Sternad, L. Lindbom, and A. Ahlén, "Wiener design of adaptation algorithms with time-invariant gains," *IEEE Trans. Signal Processing*, vol. 50, pp. 1895–1907, Aug. 2002.
[8] E. Eweda and O. Macchi, "Tracking error bounds of adaptive nonstationary filtering," *Automatica*, vol. 21, pp. 293–302, 1985.
[9] D. C. Farden, "Tracking properties of adaptive signal processing algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 439–446, June 1981.
[10] L. Guo and L. Ljung, "Exponential stability of general tracking algorithms," *IEEE Trans. Automat. Contr.*, vol. 40, pp. 1376–1387, Aug. 1995.
[11] ——, "Performance analysis of general tracking algorithms," *IEEE Trans. Automat. Contr.*, vol. 40, pp. 1388–1402, Aug. 1995.
[12] O. Macchi, "Optimization of adaptive identification for time-varying filters," *IEEE Trans. Automat. Contr.*, vol. AC-31, pp. 283–287, Mar. 1986.
[13] G. V. Moustakides, "Exponential convergence of products of random matrices: Application to adaptive algorithms," *Int. J. Adaptive Contr. Signal Process.*, vol. 12, pp. 579–597, 1998.
[14] I. D. Landau, "A feedback system approach to adaptive filtering," *IEEE Trans. Inform. Theory*, vol. 30, pp. 251–262, Mar. 1984.

[15] C. P. Kwong, "Further results on the control-theoretic analysis of the LMS algorithm," *IEEE Trans. Signal Processing*, vol. 41, pp. 943–946, Feb. 1993.

[16] A. H. Sayed and M. Rupp, "An ℓ_2-stable feedback structure for nonlinear adaptive filtering and identification," *Automatica*, vol. 33, pp. 13–30, 1997.

[17] 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.

[18] A. H. Sayed and T. Y. Al-Naffouri, "Mean-square analysis of normalized leaky adaptive filters," in *Proc. ICASSP*, vol. 6, Salt Lake City, UT, May 2001, pp. 3873–3876.

[19] O. Macchi, *Adaptive Processing: The Least Mean Squares Approach With Applications in Transmission*. Chichester, U.K.: Wiley, 1995.

[20] A. Feuer and E. Weinstein, "Convergence analysis of the LMS filters with uncorrelated gaussian data," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 222–229, Feb. 1985.

[21] W. A. Gardner, "Learning characteristics of stochastic-gradient- descent algorithms: A general study, analysis, and critique," *Signal Process.*, vol. 6, pp. 113–133, 1984.

[22] ——, "Nonstationary learning characteristics of the LMS algorithm," *IEEE Trans. Circuits Syst.*, vol. CS-34, pp. 1199–1207, Oct. 1987.

[23] S. Haykin, *Adaptive Filter Theory*, second ed. Englewood Cliffs, NJ: Prentice Hall, 1991.

[24] ——, *Adaptive Filter Theory*, third ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[25] L. L. Horowitz and K. D. Senne, "Performance advantage of complex LMS for controlling narrow-band adaptive arrays," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 29, pp. 722–735, June 1981.

[26] B. Widrow and E. Walach, "On the statistical efficiency of the LMS algorithm with nonstationary inputs," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 211–221, Mar. 1984.

[27] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1980.

[28] M. Vidyasagar, *Nonlinear Systems Analysis*, second ed. London, U.K.: Prentice-Hall Int., 1993.

[29] H. J. Kushner, *Approximation and Weak Convergence Methods for Random Processes with Application to Stochastic System Theory*. Cambridge, MA: MIT Press, 1984.

[30] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 64, pp. 1151–1162, Aug. 1976.

[31] L. Lindbom, A. Ahlén, M. Sternad, and M. Falkenström, "Tracking of time-varying mobile radio channels. Part II: A case study," *IEEE Trans. Commun.*, vol. 50, pp. 156–167, Jan. 2002.

[32] M. Sternad, L. Lindbom, and A. Ahlén. Robust wiener design of adaptation laws with constant gains. presented at IFAC Workshop on Adaptation and Learning in Control and Signal Processing. [Online]. Available: www.signal.uu.se/Publications/abstracts/c0113.html

[33] L. Lindbom, "A Wiener filtering approach to the design of tracking algorithms," Ph.D. dissertation, Dept. Technol., Uppsala Univ., Uppsala, Sweden, 1995.

[34] S. C. Douglas and W. Pan, "Exact expectation analysis of the LMS adaptive filter," *IEEE Trans. Signal Processing*, vol. 43, pp. 2863–2871, Dec. 1995.

[35] V. Kucera, *Analysis and Design of Discrete Linear Control Systems*. Prague, Czech Republic: Academia, 1991.

[36] T. Söderström and P. Stoica, *System Identification*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

Andres Ahlén (S'80–M'84–SM'90) was born in Kalmar, Sweden. He received the Ph.D. degree in automatic control in 1986 from Uppsala University, Uppsala, Sweden.

He was an Assistant Professor with the Systems and Control Group, Uppsala University, from 1984 to 1989 and as an Associate Professor in automatic control from 1989 to 1992. During 1991, he was a visiting research fellow with the Department of Electrical and Computer Engineering, The University of Newcastle, Callaghan, Australia. In 1992, he was appointed Associate Professor of signal processing at Uppsala University. Since 1996, he has been a Full Professor and has held the chair in Signal Processing at Uppsala University. Since 1996, he has also been the head of the Signals and Systems Group at Uppsala University. His research interests, which include signal processing, communications, and control are currently focused on signal processing for wireless communications.

Dr Ahlén is an editor of the IEEE TRANSACTIONS ON COMMUNICATIONS.

Lars Lindbom was born in Västervik, Sweden. He received the M.S. degree in engineering physics and the Ph.D. degree in signal processing from Uppsala University (UU), Uppsala, Sweden, in 1989 and 1998, respectively.

From 1995 to 2002, he was with Ericsson Infotech, Karlstad, Sweden, where he held a senior specialist position in adaptive filtering for mobile radio systems. Since the end of 2002, he has been with the Signals and Systems Group at UU. His main research interests include adaptive filtering, equalization, and system identification with applications to wireless communications.

Mikael Sternad (S'83–M'88–SM'90) received the M.S. degree in engineering physics in 1981 and the Ph.D. degree in automatic control in 1987, both from the Institute of Technology at Uppsala University (UU), Uppsala, Sweden.

He has been a professor in automatic control at UU since 2001. His main research interest is signal processing applied to mobile radio communication problems, such as long-range channel prediction, which is used for fast link adaptation and scheduling of packet data flows in wireless mobile systems. He is also involved in acoustic signal processing, in particular, compensation of loudspeaker dynamics.