Analysis of Stability and Performance of Adaptation Algorithms with Time-invariant Gains

Anders Ahlén, Lars Lindbom and Mikael Sternad*

February 1, 2003

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 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 zero mean and 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 can be neglected under certain conditions, and performance analysis then becomes straightforward. Performance analysis in the presence of a non-negligible time-varying feedback is performed for algorithms that use scalar measurements. Convergence in 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 FIR models with white inputs. This class includes Wiener LMS adaptation.**

Keywords: Tracking, adaptive signal processing, adaptive filtering, Least mean squares method.

EDICS: 2-ALGO

I. Introduction

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

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

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

$$\begin{aligned} \varepsilon_t &= y_t - \varphi_t^* \hat{h}_{t|t-1} & (2)\\ \hat{h}_{t+1|t} &= \hat{h}_{t|t-1} + \mu \varphi_t \varepsilon_t & (3)\end{aligned}$$

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), while 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

$$\begin{aligned} \varepsilon_t &= y_t - \varphi_t^* \hat{h}_{t|t-1} & (4)\\ \hat{h}_{t+k|t} &= \boldsymbol{\mathcal{M}}_k(q^{-1})(\varphi_t \varepsilon_t) = \sum_{i=0}^{\infty} \mathbf{M}_{k,i} \varphi_{t-i} \varepsilon_{t-i} \ , & (5)\end{aligned}$$

where $\boldsymbol{\mathcal{M}}_k(q^{-1})$ 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

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

while higher-order filters correspond to so-called multi-step algorithms [1], [2], [3]. The Momentum LMS algorithm [4] and the Lead-Lag LMS scheme of [5] are examples of multi-step algorithms with diagonal $\boldsymbol{\mathcal{M}}_1(q^{-1})$. In general, $\boldsymbol{\mathcal{M}}_k(q^{-1})$ need not be diagonal. Wiener methods that for arbitrary k optimize $\boldsymbol{\mathcal{M}}_k(q^{-1})$ 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

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

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

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

*Signals and Systems, Uppsala University, PO Box 528, SE-75120, Uppsala, Sweden. Phone +46 18 4713076, fax +46 18 555096, aa@signal.uu.se, ll@signal.uu.se, ms@signal.uu.se

The analysis of stability and performance of (4)(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 $\boldsymbol{\mathcal{M}}_1(q^{-1})$ of arbitrary structure, the expressions (1),(4)(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, 9, 10, 11, 12, 13]. Even for the LMS case, a strict analysis becomes very difficult.

Adaptation laws do inherently have a feedback structure, so 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 below. 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 below, 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),(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 $\boldsymbol{\mathcal{M}}_k(q^{-1})$ of arbitrary structure and complexity, for parameter variations with arbitrary power spectral density.

Independent regressors is a commonly used simplifying assumption [3, 20, 21, 22, 24, 25, 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 $\boldsymbol{\mathcal{M}}_k(q^{-1})$ 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 non-boldface 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 $z_t = \boldsymbol{\mathcal{R}}(z^{-1})u_t$ 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 zero mean and spectral density $\Phi_h(\omega)$, the additive noise v_t is stationary and zero mean, while 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 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. The Learning Filter

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

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

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 which is also useful for analyzing the tracking behavior of the algorithm (4),(5). The reformulation of the algorithm (4),(5) to a learning filter is central in this paper and it is outlined below. If

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

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

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

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

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

in which

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

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),

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

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

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

Define

$$\mathbf{Z}_t \triangleq \varphi_t \varphi_t^* - \mathbf{R} \tag{14}$$

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

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

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

The point of expressions (9), (10), (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

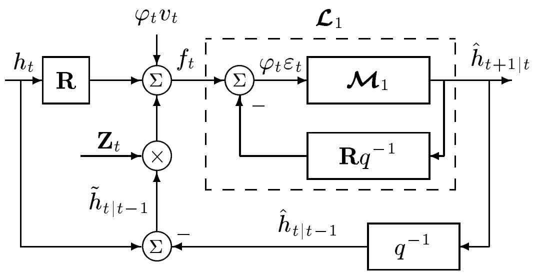

Figure 1: The adaptive algorithm (5) ($k = 1$) reformulated as a time-invariant filter operating on the signal f_t. The learning filter $\mathcal{L}_1(q^{-1})$ represents the inner time-invariant loop. The time-varying feedback loop around $\mathcal{L}_1(q^{-1})$, via the feedback noise $\mathbf{Z}_t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

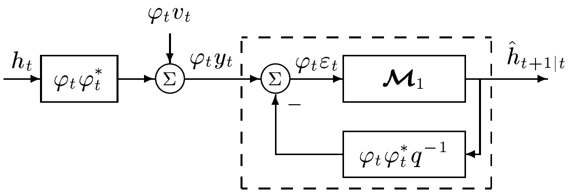

Figure 2: The algorithm (5) for $k = 1$, expressed as a time-varying filter operating on the signal $\varphi_t y_t$.

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.

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, described in subsection D below, 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

$$\mathcal{L}_1(q^{-1}) = (\mathbf{I} - (\mathbf{I} - \mu\mathbf{R})q^{-1})^{-1}\mu \quad , \tag{17}$$

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 Figure 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

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

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

C. The Tracking Error

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

$$\tilde{h}_{t+k|t} = (\mathbf{I} - q^{-k}\mathcal{L}_k\mathbf{R})h_{t+k} - \mathcal{L}_k(\varphi_t v_t) - \mathcal{L}_k(\mathbf{Z}_t\tilde{h}_{t|t-1}) \quad . \tag{19}$$

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

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

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

where

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

$$\mathbf{V}^k_{\varphi v,t} = \mathrm{E}\,\mathcal{L}_k(\varphi_t v_t)(\mathcal{L}_k(\varphi_t v_t))^* \tag{23}$$

$$\begin{aligned}\mathbf{V}^k_{Z\tilde{h},t} &= \mathrm{E}\,\mathcal{L}_k(\mathbf{Z}_t\tilde{h}_{t|t-1})(\mathcal{L}_k(\mathbf{Z}_t\tilde{h}_{t|t-1}))^* \qquad (24)\\ &= \sum_{i=0}^{t}\sum_{j=0}^{t}\mathbf{L}_{k,i}\mathrm{E}\,\{\mathbf{Z}_{t-i}\tilde{h}_{t-i|t-i-1}\tilde{h}^*_{t-j|t-j-1}\mathbf{Z}^*_{t-j}\}\mathbf{L}^*_{k,j}\end{aligned}$$

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

$$\begin{aligned}\mathbf{V}^k_{hZ\tilde{h},t} &= -\mathrm{E}\,(\mathbf{I} - q^{-k}\mathcal{L}_k\mathbf{R})h_{t+k}(\mathcal{L}_k(\mathbf{Z}_t\tilde{h}_{t|t-1}))^*\\ &- \mathrm{E}\,((\mathbf{I} - q^{-k}\mathcal{L}_k\mathbf{R})h_{t+k})^*\mathcal{L}_k(\mathbf{Z}_t\tilde{h}_{t|t-1}) \qquad (25)\\ \mathbf{V}^k_{\varphi vZ\tilde{h},t} &= \mathrm{E}\,\mathcal{L}_k(\varphi_t v_t)(\mathcal{L}_k(\mathbf{Z}_t\tilde{h}_{t|t-1}))^*\\ &+ \mathrm{E}\,(\mathcal{L}_k(\varphi_t v_t))^*\,\mathcal{L}_k(\mathbf{Z}_t\tilde{h}_{t|t-1}) \quad . \qquad (26)\end{aligned}$$

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), (21). In Section III, only the first two terms $\mathbf{V}^k_{h,t}$ and $\mathbf{V}^k_{\varphi v,t}$ are used while in Section IV, the term $\mathbf{V}^k_{Z\tilde{h},t}$ 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 3 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 below, 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

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

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, the innovation sequences of η_t and h_t, respectively with covariance matrices

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

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

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 □

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 □

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

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

and

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

respectively. A Wiener design of $\mathcal{L}_k(q^{-1})$ 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

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

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

$$\boldsymbol{\beta}\boldsymbol{\beta}_* = \lambda_e\boldsymbol{C}\,\bar{\mathbf{R}}_e\boldsymbol{C}_*NN_* + \lambda_\nu D\mathbf{R}^{-1}\boldsymbol{M}\bar{\mathbf{R}}_\nu\boldsymbol{M}_*\mathbf{R}^{-1}D_* \tag{32}$$

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

$$\lambda_e q^k \boldsymbol{C}\,\bar{\mathbf{R}}_e \boldsymbol{C}_* N_* = \boldsymbol{Q}_k \boldsymbol{\beta}_* + q\mathbf{I}D\boldsymbol{U}_{k*} \quad . \tag{33}$$

Proof: Obtained by Theorem 1 in [7], modified to an ARMA model structure for h_t with common scalar denominator polynomial. □

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

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

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

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

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

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

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

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

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 which neglects the feedback noise $\mathbf{Z}_t\tilde{h}_{t|t-1}$ and the outer time-varying loop in Fig. 1. It thus assumes $\eta_t = \varphi_t v_t$ and considers only the first two steady-state covariance matrices in (21).

Theorem 1: If the feedback noise $\mathbf{Z}_t\tilde{h}_{t|t-1}$ was zero, then the steady-state tracking error covariance matrix (20) would under Assumption 1 be given by

$$\mathbf{P}_{k,\mathrm{slow}} = \lim_{t\to\infty}(\;\mathbf{V}^k_{h,t} + \mathbf{V}^k_{\varphi v,t}\;) \;\stackrel{\Delta}{=}\; \mathbf{V}^k_h + \mathbf{V}^k_{\varphi v} \quad , \tag{37}$$

where

$$\begin{aligned}
\mathbf{V}^k_h \;&=\; \frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\mathbf{I} - e^{-j\omega k}\boldsymbol{\mathcal{L}}_k(\omega)\mathbf{R}\right)\Phi_h(\omega) \\
&\qquad \left(\mathbf{I} - e^{j\omega k}\boldsymbol{\mathcal{L}}_{k*}(\omega)\mathbf{R}\right)d\omega & (38)\\
\mathbf{V}^k_{\varphi v} \;&=\; \frac{1}{2\pi}\int_{-\pi}^{\pi} \boldsymbol{\mathcal{L}}_k(\omega)\Phi_\eta(\omega)\boldsymbol{\mathcal{L}}_{k*}(\omega)d\omega & (39)
\end{aligned}$$

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

Proof: Immediate from (22),(23) □

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 $\mathbf{V}^k_{\varphi v}$ 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

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

This is guaranteed by the Wiener design of [6],[7] □

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 below. 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),(30) differ only by the scaling factor λ_e. Introduce the following relative mean square approximation error

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

where

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

is an approximation of the tracking error (19) which neglects the feedback noise.

Lemma 2: Assume a parameter vector h_t described by the ARMA process (30), with $D(q^{-1})$ stable. A Wiener design of the learning filter is 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. The resulting estimation error is $\tilde{h}_{t+k|t}$, while the estimation error would be $\tilde{h}^a_{t+k|t}$ if the design assumption $\eta_t = \varphi_t v_t$ were true. 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 □

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 Example 1 below. 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) [30, 12, 19, 22]. In the book [19] by Macchi (see also [8],[12] for a similar definition), the DNS is for the purpose of LMS analysis characterized by the quantity

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

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) and the applicability of (43) will be investigated for the FIR system presented in Subsection 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 below. We compare to 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. Figure 3 displays these spectral densities for $\omega_o = 0.01$ in (36). Note the peak of the LMS lag error around ω_o and the contribution of high-frequency noise to the LMS error spectrum.

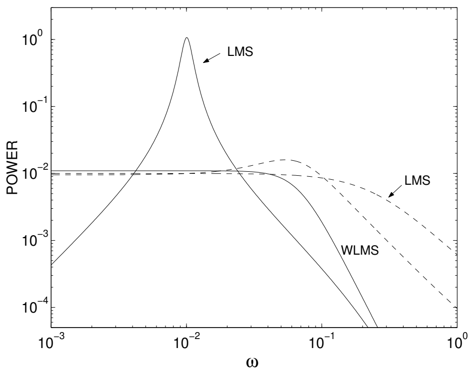

Figure 3: The spectral density of the lag error, or integrand of (38) $\|(\mathbf{I} - e^{-j\omega}\boldsymbol{\mathcal{L}}_1(\omega))(\sqrt{\lambda_e}/D(\omega))\|_2^2$ (solid) and the noise term, or integrand of (39) $0.01\|\boldsymbol{\mathcal{L}}_1(\omega)\|_2^2$ (dashed), for Wiener estimators (WLMS) and for LMS, for a FIR system with two taps varying as AR2-processes with $\omega_o = 0.01$ in (36).

Table 1: The asymptotic tracking error when second order FIR models (35),(36) with various ω_o are tracked by LMS and WLMS. Theoretical predictions $\mathrm{tr}\,(\mathbf{P}_{1,\mathrm{slow}})$ from (37) (bold) are compared to simulations of $\mathrm{tr}\,(\mathbf{P}_1)$ (italics). The largest neglected term $\mathrm{tr}\,(\mathbf{V}^1_{Z\tilde{h}})$ obtained from (24), is estimated by simulation.

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

We now vary ω_o in (36). Table 1 compares $\mathrm{tr}\,(\mathbf{P}_{1,\mathrm{slow}})$ from (37) (bold figures) with corresponding estimates of $\mathrm{tr}\,(\mathbf{P}_1)$ obtained by simulation over 100000 data (italic figures). Note the much lower tracking error of the Wiener design as compared to LMS.

The term $\mathrm{tr}\,(\mathbf{V}^1_{Z\tilde{h}}) = \lim_{t\to\infty} \mathrm{tr}\,(\mathbf{V}^1_{Z\tilde{h},t})$, 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) that neglects the feedback noise, and the measured performance. Let 10% represent a significant deviation of $\mathrm{tr}\,(\mathbf{P}_{1,\mathrm{slow}})$ relative to the true tracking error $\mathrm{tr}\,(\mathbf{P}_1)$. For LMS, the expression (37) then

[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.)

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, while 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 2 shows that the feedback noise grows in significance with an increasing M and can no longer be considered negligible for $\mathrm{M} \geq 4$.

Table 2: The asymptotic tracking error $\mathrm{tr}\,(\mathbf{P}_{1,\text{slow}})$ when FIR models (35),(36) with $\omega_o = 0.005$ of various orders M are tracked by WLMS. Theoretical predictions from (37) (bold) are compared to simulations (italics). Also shown is the largest neglected term $\mathrm{tr}\,(\mathbf{V}^1_{Z\tilde{h}})$.

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

IV. Time-varying Feedback

Our aim will now be to obtain closed-form expressions which 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, 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 □

The error $\tilde{h}_{t+k|t}$ given by (19) will then have zero mean, since e_t, 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

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

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

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

where

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

In (44), $\mathbf{V}^k_{h,t}$ and $\mathbf{V}^k_{\varphi v,t}$ 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

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

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

$$\mathbf{P}_{Z\tilde{h},t} = \mathrm{E}\,\{\mathbf{Z}_t\tilde{h}_{t|t-1}\tilde{h}^*_{t|t-1}\mathbf{Z}^*_t\} = \mathrm{E}\,\mathbf{Z}_t\mathbf{P}_{t|t-1}\mathbf{Z}^*_t$$
$$= \mathrm{E}\,\varphi_t\varphi^*_t\mathbf{P}_{t|t-1}\varphi_t\varphi^*_t - \mathbf{R}\mathbf{P}_{t|t-1}\mathbf{R} = \mathbf{R}\,\mathrm{tr}\,(\mathbf{R}\mathbf{P}_{t|t-1}) \;, \quad (47)$$

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

The main result of this section can now be presented.

Theorem 2: Consider a scalar linear regression (1) and let h_t be tracked by an estimator (10) with asymptotically stable $\boldsymbol{\mathcal{L}}_1$, that makes the integrand of (38) finite for all ω. 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

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

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

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

Proof: See Appendix C □

The expressions in Theorem 2 can be used for the adaptive algorithm (4)(5) with arbitrary $\boldsymbol{\mathcal{M}}_k(q^{-1})$, and it holds for arbitrary lags k. The stability condition involving (48) will limit the gain of $\boldsymbol{\mathcal{L}}_1(e^{j\omega})$. 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 a 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

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

where $\mathrm{L}_{k,i}$ 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

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

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:

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

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

Under Approximation 2, the cross-terms (25),(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 ≤ 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.

Approximation 1 and 2 do not hold exactly for M > 2, but they are good approximations also 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 $\mathrm{E}\,\mathbf{Z}_t \mathbf{Z}_\tau^*$ are now derived.

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

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

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

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

$$\kappa_u \triangleq \frac{\mathrm{E}\,|u_t|^4}{(\mathrm{E}\,|u_t|^2)^2} \quad . \tag{54}$$

- For real-valued regressors,

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

  and for $t \neq \tau$,

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

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

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

$$\mathcal{G}(z^{-1}) \;=\; \frac{1}{1 - m\sigma_u^4 \sum_{i=0}^{\infty} |\mathrm{L}_{1,i}|^2 z^{-i-1}} \tag{57}$$

is stable, where

$$m \;\triangleq\; \kappa_u + \mathrm{M} - 2 \quad . \tag{58}$$

---

[2] Coherent demodulated signals and fractional 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 fractional spaced method can here be regarded as providing two symbol spaced sequences, each separated by a half symbol period and each modelled 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.

The k-step estimation error is given by

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

where

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

in which

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

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

Here κ_u is defined as in (54), while $\sigma_u^2 = E|u_t|^2$ 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 $\boldsymbol{\mathcal{L}}_k(q^{-1})$, determined by (57), decreases as M increases. Also, 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, 22].

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

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

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

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, so an estimation by Monte Carlo simulation can be avoided. Performance and robustness for various mis-matched designs can also be investigated by Theorem 3, see [31, 32].

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 1. 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 $\sigma_u^2 = 1$, $\boldsymbol{\mathcal{L}}_1(q^{-1})$ and $\boldsymbol{\mathcal{M}}_1(q^{-1})$ can be expressed as

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

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

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

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

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 $\mathrm{tr}\,(\mathbf{P}_1)$ in (64).

Table 3 displays terms contributing to the minimal criterion value obtained from Theorem 3 (bold). Feedback noise-related terms are also estimated by simulation over 100000 data (italics). The two cross-terms (25) and (26) which are neglected by Theorem 3 are seen to indeed be small as compared to $\mathrm{tr}\,(\mathbf{V}_{Z\tilde{h}}^1)$. 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 2, is below 7%. The deviation between the theoretical and the measured performance peaks around M = 10.

Table 3: Contributions to the asymptotic one-step parameter prediction error $\mathrm{tr}\,(\mathbf{P}_1)$ 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 to simulations (italics).

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

Fig. 4 displays the tracking performance $\mathrm{tr}\,(\mathbf{P}_k)$ for different lags k, from Theorem 3 (o) and by simulations (∗). The step-size μ is tuned to minimize $\mathrm{tr}\,(\mathbf{P}_1)$ and the polynomials $Q_k(q^{-1})$ are obtained from Corollary 1 in [6]. (The performance for $k = 1$ can be found in Table 3.) 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 $\mathrm{tr}\,(\mathbf{P}_k)$ 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 3 and Fig 4 indicate a good agreement between the theory and measurements, with Theorem 3 giving a small overestimate of $\mathrm{tr}\,(\mathbf{P}_k)$ for $\mathrm{M} > 2$ in this example.

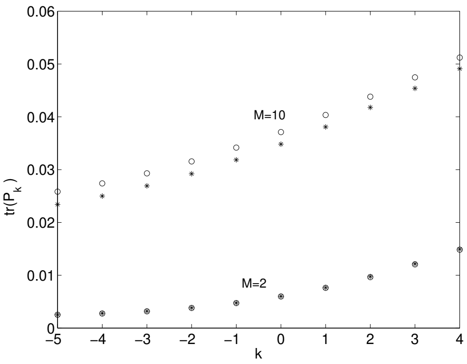

Figure 4: The tracking performance $\mathrm{tr}\,(\mathbf{P}_k)$ of WLMS for different lags k, from Theorem 3 (o) and by simulations ($*$), for FIR systems with $\mathrm{M} = 2$ and $\mathrm{M} = 10$ taps.

Example 3: Wiener LMS Stability. We increased the Wiener LMS gain μ in (68),(69), until instability occurred. In Figure 5, the FIR system (1),(35),(36) with $\omega_o = 0.1$ and $\mathrm{M} = 2$, 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 $\mathrm{tr}\,(\mathbf{P}_1)$, obtained from Theorem 3, is plotted as a function of μ, and compared to 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 $0 \leq \mu < 1$, 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 (A.13), will have singular values bounded by unity. The use of the small gain theorem (18) then gives the condition $\|e^{-j\omega}\boldsymbol{\mathcal{L}}_1(\omega)\|_\infty < 1$. 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 Figure 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.

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.

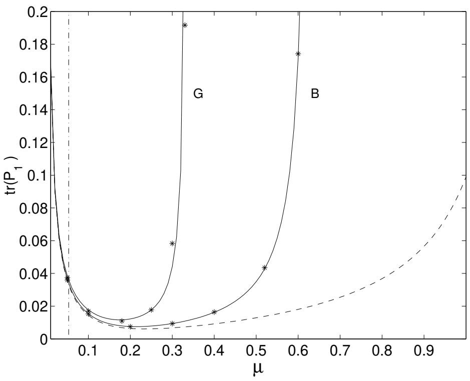

Figure 5: The performance $\mathrm{tr}\,(\mathbf{P}_1)$ as a function of μ in the WLMS algorithm (67) in Example 3, when tracking a second order FIR system (1),(35),(36) with $\omega_o = 0.1$. Results by theory from Theorem 3 (solid) and by simulation ($*$) for white real-valued binary (B) and Gaussian (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.

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 feedback noise $\mathbf{Z}_t\tilde{h}_{t|t-1}$ is small relative to $\varphi_t v_t$. 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, 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. Another useful generalization of the analysis tools (and the tools for Wiener design) would be to cover non-zero mean processes h_t, such as Ricean fading mobile radio channels.

A further generalization of Theorem 3 to FIR systems with colored regressors would be interesting but appears problematic, 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].

Acknowledgement. We thank the anonymous reviewers for very insightful comments, which have indeed helped us to improve the presentation considerably.

A 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 $\boldsymbol{\mathcal{L}}_k(q^{-1})$ 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 $|z| < 1$ while $\det \boldsymbol{\beta}_*(z)$ has all zeros in $|z| > 1$. Furthermore, from (32),

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

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 $D_*(z)$ 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 Appendix A of [7]. 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 $(z^k\boldsymbol{C}(z^{-1})\bar{\mathbf{R}}_e\boldsymbol{C}_*(z)N_*(z))^T$,

$$\underbrace{\left[\begin{array}{ccc|ccc} 0 & & 0 & \mathbf{I} & & 0 \\ \beta^{*T}_{n_\beta} & & & d_1\mathbf{I} & \ddots & \\ \vdots & \ddots & \beta^{*T}_{n_\beta} & \vdots & & \mathbf{I} \\ \beta^{*T}_0 & & \vdots & d_{n_D}\mathbf{I} & & d_1\mathbf{I} \\ & \ddots & \vdots & & \ddots & \vdots \\ 0 & & \beta^{*T}_0 & 0 & & d_{n_D}\mathbf{I} \end{array}\right]}_{\mathbf{A}_{\beta D}} \underbrace{\left[\begin{array}{c} \mathbf{Q}_0^T \\ \vdots \\ \mathbf{Q}_{n_Q}^T \\ \mathbf{U}_{n_U*} \\ \vdots \\ \mathbf{U}_{0*} \end{array}\right]}_{\mathbf{X}_{QU}}$$

$$= \lambda_e \mathbf{B}_{CN} \ . \tag{A.2}$$

Above, n_D, n_β, n_Q and n_U denote the degrees of the polynomial D and the polynomial matrices $\beta_*, \mathbf{Q}$ and $\boldsymbol{U}_{k*}$, respectively. Since the the polynomial matrix Diophantine equation (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 $\boldsymbol{\beta}^{-1}(z^{-1})$ 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 $\boldsymbol{\mathcal{L}}_1(e^{j\omega}) \to 0 \ \forall \ \omega$ 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 $\tilde{h}_{t|t-1}$ 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

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\,(\boldsymbol{\mathcal{L}}_k(\omega)\Phi_{Z\tilde{h}}(\omega)\boldsymbol{\mathcal{L}}_{k*}(\omega))d\omega \leq \lambda_e^2\alpha \ , \tag{A.4}$$

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 $\bar{\Phi}_\eta$, given by (30) and (29), respectively. Then,

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

$$+ \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\,(\boldsymbol{\mathcal{L}}_k\,\bar{\Phi}_\eta\,\boldsymbol{\mathcal{L}}_{k*})d\omega \ . \tag{A.5}$$

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

B Proof of Lemma 3.

The estimation error $\tilde{h}_{t|t-1}$ depends only on data up to time $t-1$, while $\mathbf{Z}_t$, 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, $\mathrm{E}\,(\mathbf{Z}_t\tilde{h}_{t|t-1}) = \mathrm{E}\,(\mathbf{Z}_t)\mathrm{E}\,(\tilde{h}_{t|t-1}) = 0$. Whiteness follows since

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

for $t > \tau$, since $\mathbf{Z}_t$ will be independent of all other factors and $\mathrm{E}\,(\mathbf{Z}_t) = 0$. 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 $\tilde{h}_{t|t-1}$,

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

Now, consider $\mathrm{E}\,\{\varphi_\tau v_\tau \tilde{h}^*_{t|t-1}\mathbf{Z}^*_t\}$, which is zero for $\tau \neq t$ 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, so

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

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

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

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

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

Pre-multiply by $\mathbf{R}$ and take the trace, to obtain

$$\begin{aligned}\mathrm{tr}\,(\mathbf{R}\mathbf{P}_{t+1|t}) \;&=\; \mathrm{tr}\,(\mathbf{R}\mathbf{V}^1_{h,t}) + \mathrm{tr}\,(\mathbf{R}\mathbf{V}^1_{\varphi v,t}) \\ &+ \mathrm{tr}\,(\mathbf{R}\sum_{i=0}^{t}\mathbf{L}_{1,i}\mathbf{R}\mathrm{tr}\,(\mathbf{R}\mathbf{P}_{t-i|t-i-1})\mathbf{L}^*_{1,i}) \quad . \qquad (A.6)\end{aligned}$$

Replacing $\mathbf{P}_{t-i|t-i-1}$ by $q^{-i-1}\mathbf{P}_{t+1|t}$ in (A.6) yields

$$\begin{aligned}(1 - \mathrm{tr}\,(\mathbf{R}\sum_{i=0}^{t}\mathbf{L}_{1,i}\mathbf{R}\mathbf{L}^*_{1,i}q^{-i-1}))\mathrm{tr}\,(\mathbf{R}\mathbf{P}_{t+1|t}) \\ = \mathrm{tr}\,(\mathbf{R}\mathbf{V}^1_{h,t}) + \mathrm{tr}\,(\mathbf{R}\mathbf{V}^1_{\varphi v,t}) \quad ,\end{aligned}$$

or, if we let $t \to \infty$,

$$\mathrm{tr}\,(\mathbf{R}\mathbf{P}_{t+1|t}) = \mathcal{K}(q^{-1})(\mathrm{tr}\,(\mathbf{R}\mathbf{V}^1_{h,t}) + \mathrm{tr}\,(\mathbf{R}\mathbf{V}^1_{\varphi v,t})) \quad , \qquad (A.7)$$

where the scalar transfer operator (48)

$$\mathcal{K}(q^{-1}) \triangleq \frac{1}{1 - \sum_{i=0}^{\infty}\mathrm{tr}\,(\mathbf{R}\mathbf{L}_{1,i}\mathbf{R}\mathbf{L}^*_{1,i})q^{-i-1}} \qquad (A.8)$$

determines stability. By using the static gain i.e. substituting $z = 1$ for q in (A.8) and using the limits $\mathbf{V}^1_h$ and $\mathbf{V}^1_{\varphi v}$ by (37) in (A.7), we obtain

$$\lim_{t\to\infty}\mathrm{tr}\,(\mathbf{R}\mathbf{P}_{t+1|t}) \triangleq \mathrm{tr}\,(\mathbf{R}\mathbf{P}_1) = \mathcal{K}(z^{-1})|_{z=1}\ \mathrm{tr}\,(\mathbf{R}\mathbf{P}_{1,\mathrm{slow}}) \qquad (A.9)$$

The expression (A.9) gives the *excess MSE* due to model misadjustment. The k-step covariance matrix is then given by (46), with

$$\lim_{t\to\infty}\mathbf{V}^k_{Z\tilde{h},t} \;\triangleq\; \mathbf{V}^k_{Z\tilde{h}} \;=\; \sum_{i=0}^{\infty}\mathbf{L}_{k,i}\mathbf{P}_{Z\tilde{h}}\mathbf{L}^*_{k,i} \quad . \qquad (A.10)$$

Here, $\mathbf{P}_{Z\tilde{h}} = \lim_{t\to\infty}\mathbf{P}_{Z\tilde{h},t}$. For stable $\boldsymbol{\mathcal{L}}_k(q^{-1})$ and bounded $\mathbf{P}_{Z\tilde{h}}$, $\mathbf{V}^k_{Z\tilde{h}}$ by (A.10) is bounded. By (47), (A.9),

$$\mathbf{P}_{Z\tilde{h}} \;=\; \mathbf{R}\lim_{t\to\infty}\mathrm{tr}\,(\mathbf{R}\mathbf{P}_{t|t-1}) = \mathbf{R}\,\mathcal{K}(z^{-1})|_{z=1}\ \mathrm{tr}\,(\mathbf{R}\mathbf{P}_{1,\mathrm{slow}}) \ . \qquad (A.11)$$

The use of (A.11) in (A.10) then gives (49).

D Second Order FIR Systems

Under the additional assumptions that u_t in (35) have *constant modulus*, that is, $|u_t|^2 = \sigma^2_u$, and that u_t are independent, the following result holds.

Lemma 4: Under Assumption 1, consider the FIR system (1), (35) with $\mathrm{M} = 2$ and a white noise v_t. Assume the input data u_t and u_τ $t \neq \tau$ to be independent with zero mean and constant modulus and to be either real-valued or circular complex-valued. Let the parameters h_t be estimated via a stable learning filter (10). Then, $\mathbf{Z}_t$ will be uncorrelated with $\tilde{h}_{t|t-1}$. The feedback noise $\mathbf{Z}_t\tilde{h}_{t|t-1}$ will be white with zero mean and uncorrelated with h_τ and $\varphi_\tau v_\tau$ for all τ. It will have covariance matrix

$$\mathbf{P}_{Z\tilde{h},t} = \sigma^4_u\begin{pmatrix} p^{22}_{t|t-1} & \bar{c}p^{12}_{t|t-1} \\ \bar{c}p^{12}_{t|t-1} & p^{11}_{t|t-1}\end{pmatrix} \qquad (A.12)$$

where $p^{ij}_{t|t-1}$ denotes the ij-th element of the prediction error covariance matrix $\mathbf{P}_{t|t-1}$, with $\bar{c} = 0$ for circular regressors and $\bar{c} = 1$ for real-valued regressors. □

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

$$\begin{pmatrix} u^*_t \\ u^*_{t-1}\end{pmatrix}\begin{pmatrix} u_t & u_{t-1}\end{pmatrix} - \sigma^2_u\mathbf{I} = \begin{pmatrix} 0 & u^*_t u_{t-1} \\ u^*_{t-1}u_t & 0\end{pmatrix} \ . \qquad (A.13)$$

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

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

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

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

$$\begin{aligned}&\mathrm{E}\,\mathbf{Z}_{t-i}\tilde{h}_{t-i|t-i-1}\tilde{h}^*_{t-j|t-j-1}\mathbf{Z}^*_{t-j} \\ &= \mathrm{E}\begin{pmatrix} z_{t-i}\tilde{h}^1_{t-i-1} \\ z^*_{t-i}\tilde{h}^0_{t-i-1}\end{pmatrix}\begin{pmatrix} z_{t-j}\tilde{h}^1_{t-j-1} \\ z^*_{t-j}\tilde{h}^0_{t-j-1}\end{pmatrix}^* \\ &= \mathrm{E}\begin{pmatrix} \tilde{h}^1_{t-i-1}\tilde{h}^{1*}_{t-j-1}z_{t-i}z^*_{t-j} & \tilde{h}^1_{t-i-1}\tilde{h}^{0*}_{t-j-1}z_{t-i}z_{t-j} \\ \tilde{h}^0_{t-i-1}\tilde{h}^{1*}_{t-j-1}z^*_{t-i}z^*_{t-j} & \tilde{h}^0_{t-i-1}\tilde{h}^{0*}_{t-j-1}z^*_{t-i}z_{t-j}\end{pmatrix} . \qquad (A.14)\end{aligned}$$

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. Also note that for u_t without constant modulus, each element of (A.14) 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 (A.14) 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

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

For circular complex data, factors $z_t z_t$ in (A.14) 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 (A.14) 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 (A.13) 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 $\tau < t$,

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

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

Remark 5: From the proof of Lemma 4, it is evident that the feedback noise will become correlated when the regressors do not have constant modulus □

Under the conditions stated in Lemma 4, $\mathbf{V}^k_{hZ\tilde{h},t}=0$ and $\mathbf{V}^k_{\varphi v Z\tilde{h},t}=0$ in (21). The terms $\mathbf{V}^k_{h,t}$ and $\mathbf{V}^k_{\varphi v,t}$ are asymptotically time-invariant under conditions discussed in Remark 4 in Section III. When $\mathbf{P}_{t|t-1}$ converges, the covariance matrix (21) is therefore given by (46), where the feedback noise contribution $\mathbf{V}^k_{Z\tilde{h}}$ is obtained by using $\lim_{t\to\infty}\mathbf{P}_{Z\tilde{h},t}$ from (A.12) in (A.10). Thus, $\mathbf{V}^k_{Z\tilde{h}}$ becomes a function of $\mathbf{P}_1=\lim_{t\to\infty}\mathbf{P}_{t+1|t}$, so $\mathbf{P}_1$ must be obtained iteratively from (46), (A.12) and (A.10). As initial value for the iteration, the approximation $\mathbf{V}^1_{Z\tilde{h}}=0$, representing slow variations, is used. Hereafter $\mathbf{P}_k$ can be obtained for arbitrary lags k.

E Proof of Lemma 5

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

$$\mathrm{E}\,\mathbf{Z}_t\mathbf{Z}^*_\tau = \mathrm{E}\,\varphi_t\varphi^*_t\varphi_\tau\varphi^*_\tau - \sigma_u^4\mathbf{I} \;. \tag{A.15}$$

Let the ij'th element of $\mathrm{E}\,\varphi_t\varphi^*_t\varphi_\tau\varphi^*_\tau$ be denoted by x^{ij}. Then, with $\varphi^*_t=(u_t\, u_{t-1}\,\ldots\, u_{t-\mathrm{M}+1})$, this element can be expressed as

$$x^{ij}=\sum_{\ell=1}^{\mathrm{M}}\mathrm{E}\,u^*_{t-i+1}u_{t-\ell+1}u^*_{\tau-\ell+1}u_{\tau-j+1} \tag{A.16}$$

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

$$\begin{aligned}\mathrm{E}\,u^*_i u_j u^*_k u_\ell &= \mathrm{E}\,u^*_i u_j \mathrm{E}\,u^*_k u_\ell \\ &+ \mathrm{E}\,u^*_i u^*_k \mathrm{E}\,u_j u_\ell + \mathrm{E}\,u^*_i u_\ell \mathrm{E}\,u_j u^*_k \\ &+ \frac{1}{\sigma_u^2}(\kappa_u-\kappa_u^G)\mathrm{E}\,u^*_i u_j \mathrm{E}\,u^*_k u_\ell \mathrm{E}\,u_j u^*_k \;,\end{aligned} \tag{A.17}$$

see e.g. page 549 of [36]. Here, $\kappa_u=\mathrm{E}\,|u_t|^4/(\sigma_u^2)^2$ is defined in (54) while κ_u^G is the kurtosis of Gaussian variables: $\kappa_u^G=2$ for circular complex variables while $\kappa_u^G=3$ 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 (A.17) then equals $\mathrm{E}\,|u_t|^4$.

Applying (A.17) on (A.16) for white circular data (for which $\kappa_u^G=2$) yields, with $\mathrm{E}\,u^*_i u_j=\sigma_u^2\delta_{i,j}$,

$$x^{ij}=\sigma_u^4\sum_{\ell=1}^{\mathrm{M}}(\delta_{i,\ell}\delta_{\ell,j}+\delta_{t-i,\tau-j}\delta_{t,\tau}+(\kappa_u-2)\delta_{i,\ell}\delta_{\ell,j}\delta_{t,\tau}) \;. \tag{A.18}$$

We notice that the third term within the parenthesis of (A.18) contributes to the sum only for $i=j=\ell$ and $\tau=t$, while the first contributes when $i=j=\ell$ and the second contributes for $\tau=t$ when $i=j$. Thus,

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

This observation and the use of (A.15), which subtracts $\sigma_u^4\mathbf{I}$, completes the first part of the proof. Applying the formula (A.17) on (A.16) to real-valued data ($\kappa_u^G=3$) yields

$$\begin{aligned}x^{ij}=\sigma_u^4\sum_{\ell=1}^{\mathrm{M}}(&\delta_{i,\ell}\delta_{\ell,j}+\delta_{t-i,\tau-\ell}\delta_{t-\ell,\tau-j}+\delta_{t-i,\tau-j}\delta_{t,\tau}\\ &+(\kappa_u-3)\delta_{i,\ell}\delta_{\ell,j}\delta_{t,\tau}) \;.\end{aligned} \tag{A.19}$$

Since we have real valued input data, also the second right-hand term of (A.17) contributes. The second term within the parenthesis in (A.19) contributes to the sum whenever $i-j=2(t-\tau)$, while the first, the third and the fourth terms contribute as for the case of circular data above. Thus,

$$x^{ii} = \begin{cases}\sigma_u^4(\kappa_u+\mathrm{M}-1) & \tau=t\\ \sigma_u^4 & \tau\neq t\end{cases} \tag{A.20}$$

$$\begin{matrix}x^{ij}\\ i\neq j\end{matrix} = \begin{cases}0 & \tau=t\\ 0 & \tau\neq t \text{ and } i,j=1,2\\ \sigma_u^4 & i-j=2(t-\tau),\\ & \tau\neq t,\; 2<i,j\leq \mathrm{M}\end{cases} \tag{A.21}$$

The use of (A.15) then completes the proof.

F Proof of Theorem 3

For colored v_t, Assumption 1 and white regressors imply

$$\mathrm{E}\,\mathrm{tr}\,(\varphi_\tau v_\tau v^*_t\varphi^*_t) = \mathrm{E}\,(\varphi^*_t\varphi_\tau)\mathrm{E}\,(v_\tau v^*_t) = \mathrm{M}\sigma_u^2\sigma_v^2\delta_{t,\tau} \;, \tag{A.22}$$

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

$$\operatorname{tr}(\mathrm{E}\,\eta_t \eta_\tau^*) = \operatorname{tr}(\mathrm{E}\,(\mathbf{Z}_t \tilde{h}_{t|t-1} + \varphi_t v_t)(\mathbf{Z}_\tau \tilde{h}_{\tau|\tau-1} + \varphi_\tau v_\tau)^*)$$

$$= \operatorname{tr}(\mathrm{E}\,\mathbf{Z}_\tau^* \mathbf{Z}_t \mathrm{E}\,\tilde{h}_{t|t-1} \tilde{h}_{\tau|\tau-1}^*) + \mathrm{E}\,\varphi_\tau^* \varphi_t \mathrm{E}\, v_t v_\tau^* \quad (A.23)$$

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), (A.22), and (A.23),

$$\begin{aligned} \operatorname{tr}(\mathrm{E}\,\eta_t \eta_\tau^*) &= \left(\sigma_u^4(\kappa_u + \mathrm{M} - 2)\operatorname{tr}(\mathbf{P}_{t|t-1}) + \mathrm{M}\sigma_u^2 \sigma_v^2\right) \delta_{t,\tau} \\ &= \sigma_u^4 \left(m \operatorname{tr}(\mathbf{P}_{t|t-1}) + \mathrm{M}\frac{\sigma_v^2}{\sigma_u^2} \right) \delta_{t,\tau} \quad , \quad (A.24) \end{aligned}$$

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

$$\operatorname{tr}(\mathbf{P}_{t+1|t}) = \operatorname{tr}(\mathbf{V}_{h,t}^1) + \operatorname{tr}(\mathbf{V}_{\varphi v,t}^1) + \operatorname{tr}(\mathbf{V}_{Z\tilde{h},t}^1) \quad (A.25)$$

where, from

$$\tilde{h}_{t+k|t} = (\mathbf{I} - q^{-k}\boldsymbol{\mathcal{L}}_k(q^{-1})\sigma_u^2) h_{t+k} - \boldsymbol{\mathcal{L}}_k(q^{-1})\eta_t \quad (A.26)$$

the asymptotic lag error is given by

$$\operatorname{tr}(\mathbf{V}_h^k) = \lim_{t\to\infty} \operatorname{tr}(\mathbf{V}_{h,t}^k) = \left\| (\mathbf{I} - q^{-k}\boldsymbol{\mathcal{L}}_k(q^{-1})\sigma_u^2) h_{t+k} \right\|_2^2 . \quad (A.27)$$

By using (63) and (A.22) in (23),

$$\operatorname{tr}(\mathbf{V}_{\varphi v}^k) = \lim_{t\to\infty} \operatorname{tr}(\mathbf{V}_{\varphi v,t}^k) = \mathrm{M}\sigma_v^2 \sigma_u^2 \Sigma_k \quad . \quad (A.28)$$

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

$$\operatorname{tr}(\mathbf{V}_{Z\tilde{h},t}^1) = m\sigma_u^4 \sum_{i=0}^{t} |\mathrm{L}_{1,i}|^2 q^{-i} \operatorname{tr}(\mathbf{P}_{t|t-1}) \quad . \quad (A.29)$$

The use of (A.29) in (A.25), gives, for $t \to \infty$, a scalar linear time-invariant recursion

$$\operatorname{tr}(\mathbf{P}_{t+1|t}) = \mathcal{G}(q^{-1})(\operatorname{tr}(\mathbf{V}_{h,t}^1) + \operatorname{tr}(\mathbf{V}_{\varphi v,t}^1)) \quad (A.30)$$

with $\mathcal{G}(q^{-1})$ given by (57). Under the stated assumptions, $\operatorname{tr}(\mathbf{P}_{t+1|t})$ will converge to a finite stationary limit $\operatorname{tr}(\mathbf{P}_1)$ if and only if the filter $\mathcal{G}(q^{-1})$ is stable. The steady-state error $\operatorname{tr}(\mathbf{P}_1) = \lim_{t\to\infty} \operatorname{tr}(\mathbf{P}_{t+1|t})$ 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 (A.27) and (A.28) for $k = 1$ in (A.30). This gives (64). The expression (62) is obtained by replacing $q^{-i}\operatorname{tr}(\mathbf{P}_{t|t-1})$ by $\operatorname{tr}(\mathbf{P}_1)$ in (A.29) and using (63). With (60) obtained from (A.27) and (61) from (A.28), we have the expression (59) for $\operatorname{tr}(\mathbf{P}_k)$.

References

[1] A. Benveniste, "Design of adaptive algorithms for tracking of time–varying systems," *International Journal of Adaptive Control and Signal Processing*, vol. 1, pp. 3-29, 1987.

[2] A. Benveniste, M. Métivier and P. Priouret, *Adaptive Algorithms and Stochastic Approximations.* Springer-Verlag, Berlin Heidelberg, 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 Transactions on Acoustics, Speech and Signal Processing*, vol. 38, pp. 2088–2098, 1990.

[5] H. S. Dabis and T. J. Moir, "Least mean squares as a control system," *International Journal of Adaptive Control and Signal Processing*, 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. on Communication.*, vol 49, pp. 2207-2217, 2001.

[7] M. Sternad, L Lindbom, and A. Ahlén. "Wiener design of adaptation algorithms with time-invariant gains," *IEEE Trans. on Signal Processing*, vol. 50, pp. 1895-1907, 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 Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, pp. 439-446, 1981.

[10] L. Guo and L. Ljung, "Stability of general tracking algorithms," *IEEE Transactions on Automatic Control*, vol. 40, pp. 1376-1387, 1995.

[11] L. Guo and L. Ljung, "Performance analysis of general tracking algorithms," *IEEE Transactions on Automatic Control*, vol. 40, pp. 1388-1402, 1995.

[12] O. Macchi, "Optimization of adaptive identification for time-varying filters," *IEEE Transactions on Automatic Control*, vol. 31, pp. 283-287, 1986.

[13] G.V. Moustakides, "Exponential convergence of products of random matrices: application to adaptive algorithms," *International Journal of Adaptive Control and Signal Processing*, vol. 12, pp. 579-597, 1998.

[14] I.D. Landau, "A feedback system approach to adaptive filtering," *IEEE Transactions on Information Theory*, vol. 30, pp. 251-262, 1984.

[15] C. P. Kwong, "Further results on the control-theoretic analysis of the LMS algorithm," *IEEE Transactions on Signal Processing*, vol. 41, pp. 943–946, 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 Transactions on Signal Processing*, vol. 44, pp. 1428-1439, 1996.

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

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

[20] A. Feuer and E. Weinstein, "Convergence analysis of the LMS filters with uncorrelated Gaussian data," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, pp. 222-229, 1985.

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

[22] W. A. Gardner, "Nonstationary learning characteristics of the LMS algorithm," *IEEE Transactions on Circuits and Systems*, vol. 34, pp. 1199-1207, 1987.

[23] S. Haykin, *Adaptive Filter Theory.* 2d. ed., Prentice Hall, Englewood Cliffs, NJ, 1991.

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

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

[26] B. Widrow and E. Walach, "On the statistical efficiency of the LMS algorithm with nonstationary inputs," *IEEE Transactions on Information Theory*, vol. 30, pp. 211-221, 1984.

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

[28] M. Vidyasagar, *Nonlinear Systems Analysis.* Second ed. Prentice-Hall International, London 1993.

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

[30] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proceedings of IEEE*, vol. 64, pp. 1151-1162, 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. on Communication*, vol. 50, pp. 156-167, 2002.

[32] M.Sternad, L.Lindbom and A.Ahlén, "Robust Wiener design of adaptation laws with constant gains," *IFAC Workshop on Adaptation and Learning in Control and Signal Processing*, Como, Italy, August 2001. www.signal.uu.se/Publications/abstracts/c0113.html

[33] L. Lindbom, *A Wiener filtering approach to the design of tracking algorithms.* PhD Thesis, Dept. of Technology, Uppsala University, Sweden, 1995. www.signal.uu.se/Publications/abstracts/a951.html

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

[35] V. Kŭcera, *Analysis and Design of Discrete Linear Control Systems.* Academia, Prague, 1991.

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