Tracking of Time-varying Mobile Radio Channels
Part I: The Wiener LMS Algorithm

Lars Lindbom, Mikael Sternad and Anders Ahlén*

July 2, 2001

Abstract- **Adaptation algorithms with constant gains are designed for tracking smoothly time-varying parameters of linear regression models, in particular channel models occurring in mobile radio communications. In a companion paper, an application to channel tracking in the IS-136 TDMA system is discussed. The proposed algorithms are based on two key concepts. First, the design is transformed into a Wiener filtering problem. Second, the parameters are modeled as correlated ARIMA processes with known dynamics. This leads to a new framework for systematic and optimal design of simple adaptation laws based on prior information. The algorithms can be realized as Wiener filters, called Learning Filters, or as "LMS/Newton" updates complemented by filters that provide predictions or smoothing estimates.**

The simplest algorithm, named Wiener LMS, is presented here. All parameters are here assumed governed by the same dynamics and the covariance matrix of the regressors is assumed known. The computational complexity is of the same order of magnitude as that of LMS for regressors which are either white or have autoregressive statistics. The tracking performance is however substantially improved.

Keywords: adaptive estimation, channel modeling, Least mean squares method.

I. Introduction

We here propose a novel way of extending and optimizing the structure of LMS-like adaptation laws. These results were originally motivated by the difficult problem of accurately tracking rapidly time-varying channel parameters in the IS-136 TDMA cellular system. A design related to the proposed algorithm [27] has successfully been used on IS-136 channels [21, 35], and a case study on this particular application can be found in Part II [31].

Motivated also by other applications such as multi-carrier systems, multi-antenna systems [35], and multiuser detectors [12, 42], a framework has been developed for designing low-complexity adaptation laws. These algorithms track coefficients of linear regression models under assumptions which are realistic in communications applications. Our aim is to improve upon the sometimes inadequate tracking performance offered by standard LMS and RLS algorithms, see for example [45] and [10].[1] A key insight is that prior information about the nature of time-variations has to be utilized if an adaptation law is to significantly outperform RLS and LMS. Do the parameters change erratically or smoothly, is their behavior oscillatory or do they rather drift as linear trends? Such information should be used in an efficient way.

Models of the dynamics of time-varying parameters have become known as *hypermodels*. While used earlier in time series modelling [15, 22], their use in the design of adaptive algorithms for regression models have been discussed by Benveniste and co-workers [4, 5]. This work, as well as many others, was focused on cases with slowly time-varying dynamics only. For such cases, the powerful tools of weak convergence theory and various methods of averaging can be used for both analysis and design [25, 32]. However, these tools are not applicable to fast time-variations. Our aim here is to systematically integrate prior information into the design of algorithms which can handle also fast parameter variations.

The time-varying Kalman filter constitutes the MSE-optimal linear algorithm for estimating regression parameters based on linear models of their behavior, [3, 34] and it can be used for channel tracking [8, 17, 40]. Unfortunately, its computational complexity may often preclude its use. In this paper, we propose an alternative approach, which avoids the on-line Riccati update required in ordinary Kalman adaptation laws, and potential numerical problems of fast Kalman implementations.

We shall focus on the class of adaptation laws with constant gains, of much lower complexity. Their design can be formulated as a Wiener problem by a transformation of variables, explained in Section III.A. Minimization of the mean square tracking error then becomes rather straightforward, using, for example, a polynomial approach to Wiener filter design [1, 2]. A family of adap-

*Lars Lindbom is with Ericsson Infotech, PO Box 1038, SE-65115 Karlstad, Sweden. Lars.Lindbom@ks.ericsson.se. Mikael Sternad and Anders Ahlén are with Signals and Systems, Uppsala University, PO Box 528, SE-75120, Uppsala, Sweden. Phone +46 4713076, fax +46 18 555096, ms@signal.uu.se, aa@signal.uu.se

[1] For fast time-variations, an RLS algorithm must use a short data window, so the estimated covariance matrix, or Hessian, will be inaccurate. In the problems considered in this paper, the regressors are stationary. If their statistics is unknown, it is then both possible and natural to estimate their covariance matrix separately and accurately, using a long data window.

tation laws has been obtained [28, 29], with several practically useful features:

- The complexity of the algorithm is directly coupled to the choice of hypermodel. The user will in this way be able to make systematic tradeoffs between complexity and model approximations, using insights offered by classical Wiener theory.

- Estimators providing predictions or smoothing estimates can be obtained by optimizing the adaptation law for the desired estimation horizon.

Channel parameter predictions are of use in decision-directed adaptive detectors [6] and in fast transmitter power control. Smoothing improves the tracking performance and should be used whenever the corresponding time delay is acceptable.

Within this class of algorithms our paper focuses on the variant of lowest complexity: The Wiener LMS algorithm (WLMS). It has close to LMS complexity for regressors with white or autoregressive statistics, but offers a significant performance improvement as compared to LMS. The traditional LMS-algorithm can be regarded as a special case, obtained for white regressors when assuming the time-varying parameters to be random walks.

Remarks on the notation. For any polynomial

$$P(q^{-1}) = p_0 + p_1 q^{-1} + \ldots + p_{n_P} q^{-n_P} \tag{1}$$

in the backward shift operator q^{-1} ($q^{-1}x_t = x_{t-1}$), a conjugate polynomial is then defined as $P_*(q) \stackrel{\Delta}{=} p_0^* + p_1^* q + \ldots + p_{n_P}^* q^{n_P}$ where q is the forward shift operator ($qx_t = x_{t+1}$) and p^* denotes the complex conjugate of p. The arguments q^{-1} or q are often omitted where there is no risk of misunderstanding. When $p_0 = 1$, the polynomial is called *monic*. Time-varying coefficients of filters or polynomials (1) are denoted as p_t^j, or $p_{j,t}$. Rational matrices, or transfer function matrices, $\boldsymbol{\mathcal{R}}(q^{-1})$ are defined stable if all elements have all poles in $|z| < 1$, and are marginally stable with poles in $|z| \leq 1$.

II. The Tracking Problem

A. Linear Regressions and Hypermodels

A sequence of measurement vectors $\{y_t\}$ of dimension $n_y|1$ is assumed available at the discrete time instants $t = 0, 1, 2, ...$ and to be generated by a linear regression

$$y_t = \varphi_t^* h_t + v_t \quad , \tag{2}$$

where v_t respresents noise and all terms may be complex-valued. The known regression matrix sequence $\{\varphi_t^*\}$, of dimension $n_y|n_h$, is defined as the Hermitian conjugate of a corresponding $n_h|n_y$ matrix φ_t. It is known up to time t and is assumed stationary with zero mean and nonsingular covariance matrix

$$\mathbf{R} \stackrel{\Delta}{=} \mathrm{E}\, \varphi_t \varphi_t^* \quad . \tag{3}$$

Our aim is to estimate the time-varying parameter column vector $h_t = [h_{0,t} \ldots h_{n_h-1,t}]^T$. We cannot, without further assumptions, determine the sequence of parameters h_t uniquely from a sequence of measurements $y_1 = \varphi_1^* h_1 + v_1$, $y_2 = \varphi_2^* h_2 + v_2$, ... even in the noise-free case, if $n_y < n_h$. We would have unknowns $h_1, h_2 \ldots$ with more elements than the available measurements $y_1, y_2 \ldots$. To avoid this dilemma, assumptions must be introduced on the relationship between parameters h_t and h_τ for $\tau \neq t$.

Over a time interval of interest, the elements of h_t could be assumed to evolve as linear combinations of deterministic basis functions, see e.g. [7, 26, 36, 37]. Polynomial basis functions are a possible choice, but they often result in noise-sensitive estimates [36, 37]. In the telecommunication area, the available information about the nature of time-variations is in general statistical as in, for example, different types of models for fading channels. We will therefore use a stochastic approach: The parameter vector h_t will be modelled by a stochastic process, with known statistical properties.

A large variety of parameter dynamics can be described by linear time-invariant stochastic hypermodels

$$h_t = \boldsymbol{\mathcal{H}}(q^{-1}) e_t \quad , \tag{4}$$

where $\boldsymbol{\mathcal{H}}$ is a stable or marginally stable transfer function matrix of dimension $n_h|n_h$ and e_t is a white noise vector. The model represents either prior information about time variations or design assumptions.

Special cases of (4) with simpler structure are often adequate. A useful specialization is when $\boldsymbol{\mathcal{H}}(q^{-1})$ is diagonal, with equal elements along the diagonal:

$$h_t = \boldsymbol{\mathcal{H}}(q^{-1}) e_t = \frac{C(q^{-1})}{D(q^{-1})} \mathbf{I}\; e_t \quad . \tag{5}$$

Our WLMS design will be based on this vector-ARIMA model. When (5) is further specialized to

$$h_t = \frac{1}{D(q^{-1})} \mathbf{I}\; e_t = \frac{1}{1 + d_1 q^{-1} + d_2 q^{-2}} \mathbf{I}\; e_t \tag{6}$$

with d_1 and d_2 being real-valued, we have introduced a model for which explicit solutions to the design equations are available, see Section IV.

If the channel parameters h_t have nonzero mean, it is appropriate to let $D(z^{-1})$ have a zero in $z = 1$. Unbiased estimates of h_t are then obtained, see (42) below. One special case is the random walk model ($d_1 = -1, d_2 = 0$) on which many adaptation schemes are indirectly based [18],[32],[40]. Another special case,

$$h_t = h_{t-1} + \frac{1}{1 - d_2 q^{-1}} \mathbf{I}\; e_t \tag{7}$$

is obtained by setting $d_1 = -(1 + d_2)$ in (6). This model is commonly referred to as a *filtered random walk* and with $d_2 = 1$ in (7), we obtain *integrated* random walk hypermodelling. Integrated random walks, used e.g. in [22] model the elements of h_t as noisy linear trends.

If no prior information is available about the dynamics of h_t, then (7), with d_2 chosen in the interval $[0.90 - 0.999]$, will often be a reasonable and robust first choice of hypermodel. It allows for a certain degree of smoothness in the parameter evolution, which is not the case with a random walk model ($d_2 = 0$). If some knowledge is available, the "design" parameter d_2 can be used to match the smoothness of h_t.

In the estimation of communication channels, two complications are encountered: Models of time variability and fading will not be exactly known. Furthermore, estimated transmitted symbols, which act as regressors as illustrated in the example below, may include decision errors and are thus not known completely. These issues are discussed further in Part II [31]. We there also discuss the adjustment of (5) to fading statistics described by Jakes model, in particular robust adjustment in situations when the Doppler frequency of mobile users is not exactly known.

Example 1: A fading multi-input-multi-output mobile radio channel [29]. Consider a TDMA-based mobile cellular communication system with several transmitters on the same frequency in the same cell and in the same time slot. The situation could either represent channel reuse within cells by several mobile users [42] or one transmitter which uses multiple antennas to increase the data rate [12]. Receivers with multiple diversity branches detect the users simultaneously. With two transmitters and two received signals obtained at different positions and/or differing polarizations, a suitable model is

$$\begin{pmatrix} y_t^1 \\ y_t^2 \end{pmatrix} = \begin{pmatrix} B_t^{11}(q^{-1}) & B_t^{12}(q^{-1}) \\ B_t^{21}(q^{-1}) & B_t^{22}(q^{-1}) \end{pmatrix} \begin{pmatrix} u_t^1 \\ u_t^2 \end{pmatrix} + \begin{pmatrix} v_t^1 \\ v_t^2 \end{pmatrix}, \tag{8}$$

where y_t^i is the baseband signal at antenna i. Here

$$B_t^{ij}(q^{-1}) = b_{0,t}^{ij} + b_{1,t}^{ij} q^{-1} + \ldots + b_{M-1,t}^{ij} q^{-M+1}$$

represents an M-tap multipath fading channel from user j to antenna i. The signals u_t^j are the (possibly pulse-shaped) symbols transmitted from user j and have known and time-invariant statistics. The transmitted symbols are either known, during the training phase, or estimated by a multiuser detector which, in turn, is based on the channel estimates. With

$$\varphi_t^{1*} = (u_t^1 \ldots u_{t-M+1}^1) \quad , \quad \varphi_t^{2*} = (u_t^2 \ldots u_{t-M+1}^2) \quad ,$$

and by collecting the unknown channel parameters in vectors $b_t^{ij} = (b_{0,t}^{ij} \quad b_{1,t}^{ij} \ldots b_{M-1,t}^{ij})^T$, a linear regression model (2) is obtained by stacking all vectors b_t^{ij}, as

$$\begin{aligned} \begin{pmatrix} y_t^1 \\ y_t^2 \end{pmatrix} &= \begin{pmatrix} \varphi_t^{1*} & \varphi_t^{2*} & 0 & 0 \\ 0 & 0 & \varphi_t^{1*} & \varphi_t^{2*} \end{pmatrix} \begin{pmatrix} b_t^{11} \\ b_t^{12} \\ b_t^{21} \\ b_t^{22} \end{pmatrix} + \begin{pmatrix} v_t^1 \\ v_t^2 \end{pmatrix} \\ &= \varphi_t^* h_t + v_t \end{aligned} \tag{9}$$

□

B. The WLMS Adaptation Structure

Assuming the system to be described by (2)-(4), parameter tracking becomes a signal estimation problem, with h_t in (4) being sought. Define the tracking error vector

$$\tilde{h}_{t+k} \triangleq h_{t+k} - \hat{h}_{t+k|t}$$

where $\hat{h}_{t+k|t}$ is an estimate of h_{t+k} at time t representing filtering ($k = 0$), prediction ($k > 0$) or fixed lag smoothing ($k < 0$). We will measure tracking performance by

$$\operatorname{tr} \mathbf{P}_k \triangleq \lim_{t\to\infty} \operatorname{tr} \mathrm{E}\, \tilde{h}_{t+k} \tilde{h}_{t+k}^* = \lim_{t\to\infty} \sum_{i=0}^{n_h - 1} \mathrm{E}\, \left| h_{i,t+k} - \hat{h}_{i,t+k|t} \right|^2 \tag{10}$$

where the expectation is taken with respect to e_t in (4) and v_t in (2) after the initial transients.

The class of adaptation algorithms, within which we here chose to minimize (10), corresponds to introducing two modifications in the LMS algorithm

$$\varepsilon_t = y_t - \varphi_t^* \hat{h}_{t|t-1} \tag{11}$$
$$\hat{h}_t = \hat{h}_{t|t-1} + \mu \varphi_t \varepsilon_t \tag{12}$$
$$\hat{h}_{t+1|t} = \hat{h}_t \quad , \tag{13}$$

where $\hat{h}_t$ denotes the filtering estimate $\hat{h}_{t|t}$, $\hat{h}_{t+1|t}$ is the one-step prediction estimate, μ is the scalar adaptation gain and ε_t is the prediction error:

1. The update direction in (12) is modified to $\mathbf{R}^{-1} \varphi_t \varepsilon_t$, as in the LMS/Newton law [46][2].

2. Equation (13) must be modified by filtering $\hat{h}_t$ to obtain the prediction and smoothing estimates for arbitrary lags k, see (16) below.[3] When the parameter dynamics differs from a random walk, the filter estimate $\hat{h}_t$ will not be an optimal predictor.

The modified algorithm is then described by

$$\varepsilon_t = y_t - \varphi_t^* \hat{h}_{t|t-1} \tag{14}$$
$$\hat{h}_t = \hat{h}_{t|t-1} + \mu \mathbf{R}^{-1} \varphi_t \varepsilon_t \tag{15}$$
$$\hat{h}_{t+k|t} = \boldsymbol{\mathcal{P}}_k(q^{-1}) \hat{h}_t \quad , \tag{16}$$

where $\boldsymbol{\mathcal{P}}_k(q^{-1})$ is a causal and stable rational matrix, which may provide prediction or smoothing estimates for any horizon k. We shall refer to $\boldsymbol{\mathcal{P}}_k(q^{-1})$ as the *coefficient smoothing-prediction filter*. The appropriate tuning of $\boldsymbol{\mathcal{P}}_k(q^{-1})$ will depend on the dynamics of h_t and on the SNR. Note that the one-step prediction estimate

[2] Of course, this presupposes that $\mathbf{R}^{-1}$ is known, or is estimated separately. If $\mathbf{R}$ is unknown, its inverse $\mathbf{R}^{-1}$ can be estimated on line, with well-known methods [17], at the price of an increase in complexity to a level similar to that of RLS. Since $\mathbf{R}^{-1}$ is assumed constant or perhaps slowly time-varying, a long data window can be used for estimating it even when h_t varies quickly.

[3] The use of one-step *coefficient prediction filters* for this purpose has been proposed by Kubin in [23, 24] and equivalent filter structures were introduced in [7] by Clark.

$\hat{h}_{t+1|t} = \boldsymbol{\mathcal{P}}_1(q^{-1})\hat{h}_t$ must always be made available, since $\hat{h}_{t|t-1}$ is required in (14) and (15).

In the *Wiener LMS* (or WLMS) algorithm discussed in this paper, $\boldsymbol{\mathcal{P}}_k(q^{-1})$ will be constrained to a *diagonal rational matrix*, with equal stable and causal transfer functions along the diagonal. With a diagonal $\boldsymbol{\mathcal{P}}_k(q^{-1})$, the required number of computations grows only linearly with the number of parameters n_h, when $\mathbf{R}$ is diagonal. For nondiagonal $\mathbf{R}$, the product $\mathbf{R}^{-1}\varphi_t$ can be updated with a computational complexity proportional to n_h, for scalar FIR models with autoregressive inputs [11].

To summarize, the time-varying parameters h_t in

$$\begin{aligned} h_t &= \boldsymbol{\mathcal{H}}(q^{-1})e_t \\ y_t &= \varphi_t^* h_t + v_t \end{aligned} \tag{17}$$

are to be estimated by minimizing the criterion (10) within the class of tracking algorithms represented by (14)-(16) with $\boldsymbol{\mathcal{P}}_k(q^{-1})$ being diagonal, assuming φ_t^*, $\mathbf{R}$, $\boldsymbol{\mathcal{H}}(q^{-1})$ and $\mathbf{R}_e = \mathrm{E}\, e_t e_t^*$ to be known.

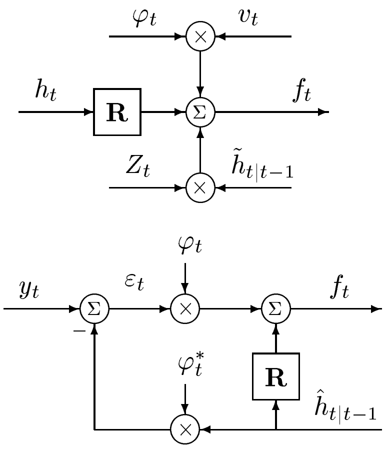

Figure 1: Two equivalent representations of the fictitious measurement f_t. The lower diagram constructs f_t via (22) from available signals when $\mathbf{R} = \mathrm{E}\,[\varphi_t\varphi_t^*]$ is known.

III. Wiener LMS Design

A time-invariant filter, operating on a fictitious measurement signal, will be shown to be equivalent to the adaptation law (14)-(16). The optimization of the adaptation law will then be solved by an open loop linear Wiener design, which gives several advantages: It provides a systematic design technique, a numerically safe implementation and an opportunity for using tools and design intuition from Wiener filtering.

A. The Fictitious Measurements

Consider the signal prediction error (14) and insert (2) describing y_t, to obtain

$$\begin{aligned} \varepsilon_t &= \varphi_t^*(h_t - \hat{h}_{t|t-1}) + v_t \\ \varphi_t\varepsilon_t &= \varphi_t\varphi_t^*\tilde{h}_{t|t-1} + \varphi_t v_t \quad . \end{aligned} \tag{18}$$

By adding and subtracting $\mathbf{R}\tilde{h}_{t|t-1}$ and defining

$$Z_t = \varphi_t\varphi_t^* - \mathbf{R} \tag{19}$$

$$\eta_t = Z_t\tilde{h}_{t|t-1} + \varphi_t v_t \tag{20}$$

$$f_t = \mathbf{R}h_t + \eta_t \quad , \tag{21}$$

the vector (18) is now reformulated as

$$\begin{aligned} \varphi_t\varepsilon_t &= \mathbf{R}h_t - \mathbf{R}\hat{h}_{t|t-1} + Z_t\tilde{h}_{t|t-1} + \varphi_t v_t \\ &= f_t - \mathbf{R}\hat{h}_{t|t-1} \quad . \end{aligned} \tag{22}$$

Here, f_t can be regarded as a fictitious measurement, with $\mathbf{R}h_t$ and η_t in (21) being the signal and the noise, respectively. It can be constructed from known signals as depicted in the lower diagram of Figure 1. In the sequel, the noise terms η_t and $Z_t\tilde{h}_{t|t-1}$ will be referred to as the *gradient noise* and the *feedback noise*, respectively. The matrix Z_t, of dimension $n_h|n_h$, has zero mean by definition. This matrix was introduced by Gardner [13] and was referred to as the *autocorrelation matrix noise.*

B. Tracking Regarded as Filtering

Based on the relations (19)-(22), we can design a time-invariant stable rational matrix $\boldsymbol{\mathcal{L}}_k(q^{-1})$ that operates on f_t and provides an estimate of h_{t+k}

$$f_t = \mathbf{R}\hat{h}_{t|t-1} + \varphi_t\varepsilon_t = \mathbf{R}h_t + \eta_t \tag{23}$$

$$\hat{h}_{t+k|t} = \boldsymbol{\mathcal{L}}_k(q^{-1})f_t \quad . \tag{24}$$

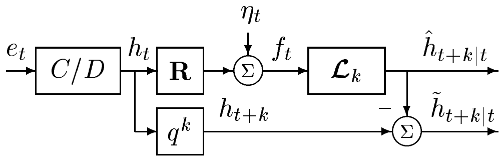

Figure 2: A filter design problem. The vector h_{t+k} is to be estimated from measurements f_t, such that the steady state mean square tracking error $E|\tilde{h}_{t+k|t}|^2$ is minimized.

The filter $\boldsymbol{\mathcal{L}}_k(q^{-1})$ will be referred to as the *learning filter.* If (14)-(16) is the preferred implementation, a corresponding gain μ and a stable filter $\boldsymbol{\mathcal{P}}_k(q^{-1})$ can be readily calculated.

For the LMS algorithm, the learning filter

$$\boldsymbol{\mathcal{L}}_1(q^{-1}) = (\mathbf{I} - (\mathbf{I} - \mu\mathbf{R})q^{-1})^{-1}\mu \tag{25}$$

is obtained by inserting $\varphi_t\varepsilon_t$ from (22) in (12),(13). The stability requirement on this LMS learning filter corresponds to the condition for convergence in the mean square [17]

$$0 < \mu < \frac{2}{\lambda_{\max}} \tag{26}$$

where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{R}$. This is readily shown through an eigenvalue decomposition of $\mathbf{R}$ in (25).

Three factors influence the tracking performance via f_t in (24),(21). The scaled and rotated parameters $\mathbf{R}h_t$, representing the useful signal, the noise $\varphi_t v_t$ and old tracking errors, via the feedback noise $Z_t\tilde{h}_{t|t-1}$ cf. (20).

The estimation error follows from (21) and (24) as

$$\tilde{h}_{t+k|t} = (q^k\mathbf{I} - \boldsymbol{\mathcal{L}}_k(q^{-1})\mathbf{R})h_t - \boldsymbol{\mathcal{L}}_k(q^{-1})\eta_t \quad . \qquad (27)$$

The first term of (27) is for $k=0$ usually referred to as the *lag error*. It could be eliminated completely by using the learning filter $\boldsymbol{\mathcal{L}}_0(q^{-1}) = \mathbf{R}^{-1}$, but this would amplify the noise η_t by $\mathbf{R}^{-1}$. The aim of our design will be to obtain an MSE-optimal balance between the lag error and the influence of the gradient noise η_t, by minimizing (10).

If the innovations sequence of the noise η_t is uncorrelated with $\hat{h}_{t-i|t-i-1}, i \geq 0$, then an open-loop Wiener design of $\boldsymbol{\mathcal{L}}_k$ in (24) can be performed. If it is uncorrelated with h_{t-i}, an open-loop design is further simplified. These conditions are not always fulfilled but hold approximately under many practically important circumstances, since the multiplication by Z_t in (20) acts as a scrambler.

In the following filter design, η_t is assumed white and its uncorrelatedness with h_{t-i} and $\hat{h}_{t-i|t-i-1}$ will be stated as an assumption, under which $\boldsymbol{\mathcal{L}}_k(q^{-1})$ will be optimized by just treating η_t in (27) as an additive noise with known properties.

C. Optimal Filter Design

The diagonality constraint imposed on $\boldsymbol{\mathcal{P}}_k(q^{-1})$ in the implementation (14)-(16) will correspond to a related constraint on the structure of the learning filter. The so constrained learning filter will now be optimized for parameter variations described by (5). The main reason for imposing the constraint (5) on the hypermodel is that this leads to a *single* scalar design equation.

The design assumptions are formalized below.

Assumption A1: The signal y_t is described by (2), where $\mathbf{R}$ is time-invariant, known and nonsingular. The second order moments of h_t are described by

$$h_t = \frac{C(q^{-1})}{D(q^{-1})}\mathbf{I}\; e_t = \frac{C(q^{-1})}{D_s(q^{-1})D_u(q^{-1})}\mathbf{I}\; e_t \qquad (28)$$

with known and monic polynomials, with zeros of $C(z^{-1})$ and of $D_s(z^{-1})$ in $|z|<1$, and zeros of $D_u(z^{-1})$ on $|z|=1$. The vector e_t is white and stationary, with zero mean and known $\mathbf{R}_e = \mathrm{E}\, e_t e_t^*$ □

Assumption A2: The learning filter (24) of dimension $n_h|n_h$ is constrained to have the structure

$$\boldsymbol{\mathcal{L}}_k(q^{-1}) = \frac{Q_k(q^{-1})}{\beta(q^{-1})}\; \mathbf{R}^{-1} \qquad (29)$$

with the polynomial $\beta(z^{-1})$ having all zeros in $|z|<1$ □

Assumption A3: The gradient noise η_t is uncorrelated with h_{t-i} and $\hat{h}_{t-i|t-i-1}, i \geq 0$. It is stationary and white, with zero mean and covariance matrix $\mathbf{R}_\eta$ □

Assumption A4: The parameter-drift-to-noise ratio, defined by

$$\gamma \;\stackrel{\Delta}{=}\; \mathrm{tr}\,\mathbf{R}_e/\mathrm{tr}\,\mathbf{R}_{\bar{\eta}} = \mathrm{tr}\,\mathbf{R}_e/\mathrm{tr}\,(\mathbf{R}^{-1}\mathbf{R}_\eta\mathbf{R}^{-1}) \qquad (30)$$

is nonzero, known and limited $(0 < \gamma < \infty)$ □

In A2, the inverse of the regressor covariance matrix is included *a priori* in the constrained learning filter (29). Among other possible structural constraints, such as $\boldsymbol{\mathcal{L}}_k = Q_k/\beta\;\mathbf{I}$ or $\boldsymbol{\mathcal{L}}_k = Q_k/\beta\;\mathbf{R}^{-1/2}$, the choice (29) is made for two reasons. First, the learning filter formulation (24) becomes equivalent to the algorithm (15) only if $\mathbf{R}^{-1}$ is a right factor of $\boldsymbol{\mathcal{L}}_k$. Second, it assures that the constrained algorithm can attain perfect tracking, with $\boldsymbol{\mathcal{L}}_0 = \mathbf{R}^{-1}$, in the noise-free case.

Assuming the gradient noise to be white in A3 simplifies the design equations.

The assumptions A1-A4 need not be fulfilled exactly in practice. For example, due to a time-varying disturbance environments, scattering geometrics and vehicle speeds, the fading model and the gradient noise will have slowly time-varying statistics. A design based on our stationarity assumptions will then still be a good approximation if these properties vary much more slowly than the time constants of the tracking loop.

Furthermore, the covariance matrix $\mathbf{R}_\eta$ of η_t will in practice depend on the actual choice of estimator. This means that the scalar γ has to be adjusted iteratively. We will return to this issue at the end of the subsection, after having discussed the optimal design.

Theorem 1: The WLMS Learning filter. Under Assumptions A1 to A4, the optimal constrained learning filter (29) minimizing (10) is unique. The polynomial $\beta(q^{-1})$ is the stable and monic solution to the polynomial spectral factorization

$$r\beta\beta_* = \gamma CC_* + DD_* \qquad (31)$$

with r being a real-valued positive scalar. The unique solution to the scalar Diophantine equation

$$q^k\gamma CC_* = rQ_k\beta_* + qDL_{k*} \qquad (32)$$

provides polynomials $Q_k(q^{-1})$ and $L_{k*}(q)$ with degrees

$$\begin{aligned} nQ_k &= \max(n_c - k, n_D - 1) \\ nL_{k*} &= \max(n_c + k, n_\beta) - 1 \quad . \end{aligned} \qquad (33)$$

The error $\tilde{h}_{t+k|t}$ is stationary with zero mean □

Proof: See Appendix A.

Learning filters determined by the design equations (31) and (32), all have the same denominator polynomial $\beta(q^{-1})$ for any lag k. Since $\beta(q^{-1})$ is a stable spectral factor, the learning filters are causal and stable[4]. The spectral factorization can be solved by computing the roots

[4] By Assumption A1, the right-hand side terms of (31) cannot have common zeros on $|z|=1$, since $C(z^{-1})$ must have all zeros inside $|z|=1$. Thus, the spectral factor β is stable for $\gamma > 0$.

of the right-hand side of (31) and forming $\beta(q^{-1})$ from the factors with stable roots. There also exist several iterative Newton algorithms for spectral factorization, see e.g. [19], [9].

Polynomial spectral factors in general appear in Wiener filters as part of the whitening filter, the inverse of the innovations model for the measurement vector [1, 2], which is obtained when all contributing signal sources are combined into one stochastic process. The equation (31) does not have this intuitive interpretation, since our problem is a *constrained* Wiener optimization problem. The polynomial $\beta(q^{-1})$ could be interpreted as the numerator of an innovations model for a *scalar* measurement $\bar{f}_t$ of a scalar signal $(C(q^{-1})/D(q^{-1}))\bar{e}_t$, where $\bar{e}_t$ has variance $\mathrm{tr}\,\mathbf{R}_e$, disturbed by a scalar noise $\bar{\eta}_t$ with variance $\mathrm{tr}\,(\mathbf{R}^{-1}\mathbf{R}_\eta\mathbf{R}^{-1})$. The innovations model of $\bar{f}_t$ would then have β as numerator polynomial. (When $D(z^{-1})$ has zeros on $|z|=1$, we would have a generalized innovations model [41].)

For predictors, $k>0$, the computational complexity of the estimator (the degree of $Q_k(q^{-1})$ by (33)) is independent of the prediction horizon.

Equation (32) is a polynomial Diophantine equation. If equated for equal powers of q and q^{-1}, it constitutes a linear system of equations, with equal number of unknowns and equations. It can always be solved with respect to the polynomial coefficients of $Q_k(q^{-1})$ and $L_{k*}(q)$ [1, 2].[5] This operation can be simplified, since a closed-form solution exists.

Corollary 1. Consider the Diophantine equation (32). For one-step prediction $k=1$, its solution is

$$\begin{aligned} Q_1(q^{-1}) &= q\left(\beta(q^{-1})-D(q^{-1})\right) && (34)\\ L_{1*}(q) &= r\beta_*(q)-D_*(q)\ . && (35)\end{aligned}$$

For $k=0$, the filtering solution is given by

$$\begin{aligned} Q_0(q^{-1}) &= \beta(q^{-1})-D(q^{-1})\frac{1}{r} = L_1(q^{-1})\frac{1}{r} && (36)\\ L_{0*}(q) &= q^{-1}(\beta_*(q)-D_*(q)) \ = \ Q_{1*}(q)\ . && (37)\end{aligned}$$

If a solution is known for some k, solutions for higher k can be obtained through the forward recursions

$$\begin{aligned} Q_{k+1}(q^{-1}) &= q\left(Q_k(q^{-1})-D(q^{-1})Q_0^k\right) && (38)\\ L_{k+1*}(q) &= qL_{k*}(q)+r\beta_*(q)Q_0^k\ , && (39)\end{aligned}$$

whereas for lower k, the solutions can be obtained through the backward recursions

$$\begin{aligned} Q_{k-1}(q^{-1}) &= q^{-1}Q_k(q^{-1})+D(q^{-1})\frac{L_0^{*k}}{r} && (40)\\ L_{k-1*}(q) &= q^{-1}\left(L_{k*}(q)-L_0^{*k}\beta_*(q)\right)\ . && (41)\end{aligned}$$

[5] Since $\beta(z^{-1})$ is stable, it has all zeros in $|z|<1$, so $\beta_*(z)$ has all zeros in $|z|>1$. Since $D(z^{-1})$ has zeros in $|z|\le 1$, $\beta_*(z)$ and $zD(z^{-1})$ will be coprime, so (32) will have a solution for any left-hand side.

Above, Q_0^k and L_0^{*k} are the leading coefficients of $Q_k(q^{-1})$ and $L_{k*}(q)$, respectively □

Proof: See Appendix B.

Remark 1: For $k=0$, $Q_0^0=1-(1/r)$. Since both $\beta(q^{-1})$ and $D(q^{-1})$ are monic, their leading coefficients cancel in (34) so no positive powers of q will appear in $Q_1(q^{-1})$. Note also that the leading coefficients inside the parentheses in (38) and in (41) cancel □

The filter in (27) determining the lag error for $k=0$ will by (29) be given by

$$\mathbf{I}-\boldsymbol{\mathcal{L}}_o(e^{-i\omega})\mathbf{R} = \frac{\beta(e^{-i\omega})-Q_0(e^{-i\omega})}{\beta(e^{-i\omega})}\mathbf{I} = \frac{D(e^{-i\omega})}{r\beta(e^{-i\omega})}\mathbf{I} \qquad (42)$$

in the frequency domain, where the second equality follows from (36). When elements of the parameter vector h_t are time-invariant, their estimates will be biased, unless the hypermodel contains an integrator, that is, when $D(z^{-1})=D_1(z^{-1})(1-z^{-1})$ for some $D_1(z^{-1})$. The lag error gain (42) will then vanish at $\omega=0$ or $e^{-i\omega}=1$ since $D(1)=0$.

We now return to the implementation (14)-(16).

Lemma 1: For a learning filter designed according to Theorem 1, the adaptive filter implementation (14)-(16) is equivalent to filtering by (24). The optimal adaptation gain is

$$\mu = Q_0^0 = 1-\frac{1}{r} \qquad (43)$$

where Q_0^0 constitutes the leading coefficient of the polynomial $Q_0(q^{-1})$, and r originates from (31) □

Proof: See Appendix B.

Remark 2: In (31), $\gamma\in]0,\infty[\ \Rightarrow r\in]1,\infty[$ which implies that $\mu\in]0,1[$.[6] The step length increases with an increased γ: When $\gamma\to\infty, r\to\gamma$. Thus, $\mu\to 1$ and $\beta\to C, Q_0\to C$ in (31), (32), so $\boldsymbol{\mathcal{L}}_0\to\mathbf{R}^{-1}$. On the other hand, a vanishing parameter drift-to-noise ratio leads to a vanishing adaptation gain. When $\gamma\to 0,\beta\to D$ and $r\to 1$. Thus, $\mu\to 0$ and $\boldsymbol{\mathcal{L}}_0\to\mathbf{0}$ □

In order to determine the coefficient smoothing-prediction filter corresponding to $\boldsymbol{\mathcal{L}}_k(q^{-1})$, we note that, by (16) and (24),

$$\hat{h}_{t+k|t} = \boldsymbol{\mathcal{P}}_k(q^{-1})\hat{h}_t \quad\text{and}\quad \hat{h}_t = \boldsymbol{\mathcal{L}}_0(q^{-1})f_t\ .$$

The k-step estimate may thus be expressed by

$$\hat{h}_{t+k|t} \ \stackrel{\Delta}{=}\ \boldsymbol{\mathcal{L}}_k(q^{-1})f_t = \boldsymbol{\mathcal{P}}_k(q^{-1})\boldsymbol{\mathcal{L}}_0(q^{-1})f_t\ .$$

Thus, $\boldsymbol{\mathcal{P}}_k(q^{-1})$ will be obtained as the diagonal matrix

$$\boldsymbol{\mathcal{P}}_k(q^{-1}) = \boldsymbol{\mathcal{L}}_k(q^{-1})\boldsymbol{\mathcal{L}}_0^{-1}(q^{-1}) = \frac{Q_k(q^{-1})}{Q_0(q^{-1})}\mathbf{I}\ , \qquad (44)$$

[6] Here, $]a,b[$ represents the interval with the boundaries excluded.

with $Q_k(q^{-1})$ and $Q_0(q^{-1})$ obtained from Corollary 1.

One iteration of (16) requires $n_h(nQ_k+1+nQ_0)$ multiplications. The steps (14),(15) require $n_y(2n_h+1)$ multiplications with $\mathbf{R}$ diagonal and $n_h^2+2n_yn_h$ multiplications otherwise, in general. See however [11] for an efficient realization of $\mathbf{R}^{-1}\varphi_t$ when the elements of φ_t are autoregressive processes with known statistics.

Since $\boldsymbol{\mathcal{P}}_k(q^{-1})$ must be stable if the implementation (14)-(16) is to be useful, $Q_0(q^{-1})$ must have all its zeros in $|z|<1$. This property is not obvious from the Wiener results of Theorem 1, but is verified below.

Lemma 2: The zeros of $Q_0(z^{-1})$ are all located in $|z|<1$ for any parameter-drift-to-noise ratio $\gamma>0$ □

Proof: See Appendix B.

Summary of WLMS design steps:

1. Calculate the spectral factor β and the scalar r via equation (31).
2. Let $Q_1 = q(\beta - D)$ and $Q_0 = \beta - D/r$. Obtain other required Q_k from Corollary 1 or (32).
3. Use the realization (24),(29), or use (14)-(16) with $\boldsymbol{\mathcal{P}}_k = (Q_k/Q_0)\mathbf{I}$ and $\mu = Q_0^0 = 1-(1/r)$.

How is this procedure to be applied in practice, when the properties of the gradient noise η_t may be hard to know in advance? An obvious approach is to iterate the design a few times, see Section V of [29]: A design may first be based on preliminary assumptions on the gradient noise level $\mathrm{tr}\,\mathbf{R}_\eta$, e.g. by neglecting the feedback noise and assuming $\eta_t=\varphi_t v_t$. A better estimate of the actual gradient noise level can then be obtained. In some situations, exact analytical expressions for the tracking MSE can be used, see the next subsection. In others, the gradient noise must be investigated by simulation. We then compute $\mathbf{R}_\eta$ from (20) and calculate a modified tracking algorithm.

Since the bandwidth of the learning filter is controlled by one scalar parameter γ, an alternative is to just use it as a scalar tuning knob, to obtain a desired tradeoff between noise sensitivity and tracking ability.

D. Stability

Under assumptions A1 and A3, the variance of the estimation error $\tilde{h}_{t|t-1}$ is bounded whenever the learning filter $\boldsymbol{\mathcal{L}}_1 = \boldsymbol{\mathcal{P}}_1\boldsymbol{\mathcal{L}}_0$ is stable and $D_u h_t$ and η_t have fixed and bounded variances. However, due to the feedback loop via (20),(21),(24) in Figure 3, the feedback noise $Z_t\tilde{h}_{t|t-1}$ will not be *independent* of $\hat{h}_{t-i|t-i-1}$ when $Z_t\neq 0$, even when there is no *correlation*. The loop could become unstable if the learning filter has too high gain, and this must be taken into account in any stability analysis. For bounded regressors the small gain theorem [44] provides (conservative) sufficient conditions for stability. Other, less conservative, conditions are presented in [30].

In [30],[28], three important scenarios are discussed in which an exact stability and performance analysis can be performed when assuming v_t, φ_t and e_t to be independent. These results are summarized below.

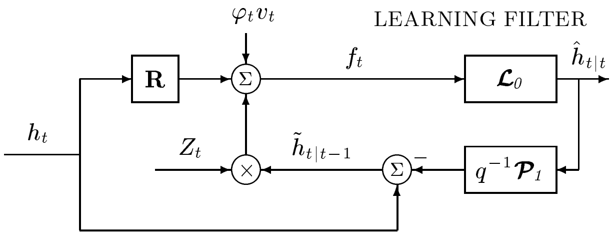

Figure 3: The feedback loop via the feedback noise $Z_t\tilde{h}_{t|t-1}$ affects the variance of the fictitious measurement f_t, and may cause dependence with $\tilde{h}_{t|t-1}$.

"Slowly" varying parameters (vanishing feedback noise). We then have a true open-loop situation. When the power of the noise e_t which drives the parameters h_t becomes small relative to the power of $\varphi_t v_t$, then the impact of the feedback noise $Z_t\tilde{h}_{t|t-1}$ on our obtained tracking MSE vanishes. This situation occurs either when the parameters h_t vary slowly, or when the noise level is high. Then, $\eta_t\approx\varphi_t v_t$ and η_t will be white whenever v_t or φ_t is a white sequence.

Independent consecutive regression matrices. If φ_t^* and φ_s^* are independent for $t\neq s$, then the feedback noise $Z_t\tilde{h}_{t|t-1}$ will be white with zero mean, and its covariance can be derived exactly. Stability is verified by evaluating the stability of a scalar transfer function [30].

While commonly used, see e.g. [13, 17, 20, 32, 45], the assumption of independent regressor matrices is unfortunately quite restrictive, as pointed out e.g. by Macchi [33]. In particular, it does not apply to the modeling of dynamic systems, for example the FIR channel of Example 1. It does hold approximately in many array applications [20] and for flat fading channels.

FIR channels with white real or circular complex scalar inputs. This type of problem will appear in the case study of Part II of this paper [31]. Assume that

$$y_t = h_{0,t}u_t + h_{1,t}u_{t-1} + v_t$$

and that u_t is a white sequence. The performance and stability of the WLMS algorithm can then be predicted without approximations. See Lemma 1 in [31]. Stability is checked by evaluating the stability of a scalar transfer function. Under mild assumptions, the stability and performance results can be used as good approximations also for FIR channels of arbitrary order with white circular complex inputs. See [30].

IV. Simplified Wiener LMS

We now confine the class of hypermodels (5) to the set (6) of *second order autoregressive (integrated)* models with real-valued coefficients. Spectral factors can then be calculated analytically [39], as outlined by Lemma 3 in Appendix C. We then obtain the following result:

Theorem 2: The simplified WLMS (SWLMS) algorithm. Consider the WLMS algorithm (14)-(16), with μ and $\boldsymbol{\mathcal{P}}_k(q^{-1})$ given by Theorem 1 and (43), (44). Assume that h_t is described by (6). Then, $\mu = Q_0^0 = 1 - (1/r)$, with r obtained from the explicit solution (C.2) in Appendix C to the spectral factorization (31) and

$$Q_k(q^{-1}) = \mu(1 \; q^{-1}) \begin{pmatrix} -d_1 & 1 \\ -d_2 & 0 \end{pmatrix}^k \begin{pmatrix} 1 \\ p \end{pmatrix} \tag{45}$$

for $k \geq 0$, where p is defined as

$$p = \frac{d_1 d_2 (1-\mu)}{1 + d_2(1-\mu)} \tag{46}$$

□

Proof: See Appendix C.

For $k = 0$ and $k = 1$, we obtain from (45)

$$Q_0(q^{-1}) = \mu(1 + pq^{-1}) \tag{47}$$

$$Q_1(q^{-1}) = \mu\left(\frac{-d_1}{1 + d_2(1-\mu)} - d_2 q^{-1}\right) \; . \tag{48}$$

For smoothing ($k < 0$), the polynomials $Q_k(q^{-1})$ are determined from the recursions (40) and (41), using $Q_0(q^{-1})$ and $L_{0*}(q) = Q_{1*}(q)$ as the initial solution pair. The coefficient predictor or smoother is readily obtained from (44) as a filter with first order denominator. For example, the one-step predictor is given by

$$\hat{h}_{t+1|t} = -p\hat{h}_{t|t-1} - \frac{d_1}{1 + d_2(1-\mu)}\hat{h}_t - d_2\hat{h}_{t-1} \; . \tag{49}$$

Remark 3: For random walk parameter variations ($d_2 = 0, d_1 = -1$), the optimal gain $\mu = 1 - (1/r)$ can be expressed in terms of γ defined in (30) as

$$\mu = \frac{1}{2}\left(\sqrt{1 + \frac{4}{\gamma}} - 1\right)\gamma \tag{50}$$

This expression is obtained from (C.2)-(C.4) in Appendix C in which $1/r = 1/\xi = 2/(2 + \gamma + \gamma m)$, where $m = (1 + 4/\gamma)^{1/2}$. When the parameter drift is slow, $\gamma \ll 1$, (50) is well approximated by

$$\mu \approx \sqrt{\gamma} \; . \tag{51}$$

□

Example 2: When is an LMS structure appropriate? Consider a first order low-pass dynamics for h_t,

$$h_t = a h_{t-1} + e_t \tag{52}$$

where $|a| \leq 1$. This corresponds to $d_1 = -a$ and $d_2 = 0$ in (6). The one-step predictor (49) becomes $\hat{h}_{t+1|t} = a\hat{h}_t$ so the optimal filter estimate (15) is given by

$$\begin{aligned} \varepsilon_t &= y_t - \varphi_t^* a \hat{h}_{t-1} \\ \hat{h}_t &= a\hat{h}_{t-1} + \mu \mathbf{R}^{-1}\varphi_t \varepsilon_t \; . \end{aligned} \tag{53}$$

For diagonal $\mathbf{R}$ and $a < 1$, this is LMS with leakage [43, 46] which for *random walk dynamics* ($a = 1$) reduces to the ordinary LMS algorithm.

An LMS algorithm (with leakage) will thus be the optimal WLMS algorithm if and only if h_t is first order autoregressive and the regressors in φ_t^* are white □

Example 3: Tracking based on integrated random walk models. Using $d_1 = -2$ and $d_2 = 1$ in (49), we obtain the one-step coefficient predictor as

$$\hat{h}_{t+1|t} = \frac{1-\mu}{1-0.5\mu}\hat{h}_{t|t-1} + \frac{1}{1-0.5\mu}\hat{h}_t - \hat{h}_{t-1} \; . \tag{54}$$

If the parameter drift-to-noise ratio γ is known or estimated, μ is obtained via Lemma 3 in Appendix C □

Several works have suggested improvements of LMS algorithms by estimation of the derivative of the parameter drift, using various difference approximations [7],[14]. In our formalism, such schemes are closely related to the use of integrated random walk models. For example, when the (arbitrary) tuning parameter α in [14] is set to $\alpha = \mu/(2-\mu)$, that algorithm reduces to SWLMS with the predictor (54). Likewise, if the tuning parameters of the "degree-1 least squares fading memory predictor" of [7, 16] are set to $\theta_1 = \mu/(2-\mu), \theta_2 = 1$, then that algorithm equals SWLMS with the predictor (54).

Example 4: A performance comparison for AR_2 models. In a scalar FIR system $y_t = h_{0,t}u_t + h_{1,t}u_{t-1} + v_t$ where v_t is white noise with variance 0.03, the parameter evolution is described by

$$\begin{pmatrix} h_{0,t} \\ h_{1,t} \end{pmatrix} = \frac{1}{1 - 2\rho\cos\omega_0 q^{-1} + \rho^2 q^{-2}} \begin{pmatrix} e_{0,t} \\ e_{1,t} \end{pmatrix} \tag{55}$$

where $\mathbf{R}_e = 10^{-5}\mathbf{I}$, $\rho = 0.995$ and $\omega_0 = 0.015$ (SNR 21dB). The steady-state tracking performance for $k = 0$ has been compared by simulation for the time-varying Kalman filter, the simplified WLMS algorithm, LMS and RLS with exponential forgetting. The 4-state Kalman estimator and the WLMS algorithms are both based on the known hypermodel (55) and a known $\mathbf{R}$. The step-size in LMS and the forgetting factor in RLS were optimized by simulation. The regressors u_t with unit variance are either white and binary (B) or Gaussian (G). For Gaussian signals, we investigate two cases: white u_t, resulting in $\mathbf{R} = \mathbf{I}$ and colored regressors, resulting in a covariance matrix with eigenvalue spread $\chi(\mathbf{R}) = 10$.

The results are summarized in Table 1, where we also compare the number of real multiplications per time

	Eigenvalue spread $\chi(\mathbf{R})$	Kalman	WLMS	LMS	RLS
B	1	0.011	0.0115	0.020	0.026
G	1	0.012	0.015	0.032	0.038
G	10	0.026	0.038	0.085	0.075
	#mult./step	214	44	18	72
	$\mathbf{R}^{-1} = \mathbf{I}$:	214	30	18	72

Table 1: Steady state mean square tracking error and number of real multiplications per time step obtained by optimized Kalman, WLMS, LMS and RLS adaptation algorithms, for binary (B) and Gaussian (G) regressors. Last line: Complexity for $\mathbf{R}^{-1} = \mathbf{I}$.

step, using the implementation (14)-(16) and (44) for WLMS.[7] It is well-known for LMS that a large regressor eigenvalue spread will reduce the attainable performance in nonstationary environments. A less known phenomenon, explained in [30], is that the use of binary regressors, instead of Gaussian regressors improves the steady state tracking performance.

It can be noted that the WLMS design attains almost the same performance as the optimal time-varying Kalman estimator at a much lower computational complexity, not much above that of the LMS algorithm □

V. Concluding Remarks

We have here presented a novel way of designing adaptation algorithms with constant gains. The resulting algorithm is close to LMS complexity when $\mathbf{R}$ is diagonal or when FIR models have AR regressors, yet it has the power to capture a range of time-variations of considerable practical importance.

In the simplest case, when assuming autoregressive parameter variations of second order or lower, only simple algebraic expressions have to be evaluated in the design. Even when no prior information on the time variations is available, the three parameters d_1, d_2 and γ (or μ) can be adjusted rather straightforwardly. While d_1, d_2 tailor the algorithm to the nature of the time-variations, γ (or μ) trades noise sensitivity against tracking ability.

The WLMS algorithm can also be generalized to structures with more degrees of freedom, which may offer higher performance in some applications, and are better equipped to handle large spreads in the properties of the elements of h_t. These generalizations, which remove the structural constraints on learning filters and allow for colored gradient noise η_t, are presented in [28] and [29].

[7] Multiplications between complex numbers are counted as four real multiplications, while multiplications or divisions between a real and a complex number are counted as two real multiplications.

A. Proof of Theorem 1

The estimate $\hat{h}_{t+k|t}$ in (24) will be perturbed by a variational term ξ_t which is based on the same constraints and the same measurement data as is available for $\hat{h}_{t+k|t} = \boldsymbol{\mathcal{L}}_k f_t$ [1],[2]. In other words, all admissible stationary variations of the estimate can be represented by $\hat{h}_{t+k|t} + \xi_t$, where

$$\xi_t = \boldsymbol{\mathcal{T}}(\mathbf{R}h_t + \eta_t) = \boldsymbol{\mathcal{T}}\left(\mathbf{R}\frac{C}{D_u D_s}e_t + \eta_t\right) \quad . \quad \text{(A.1)}$$

Since $\boldsymbol{\mathcal{L}}_k$ is constrained by (29), $\boldsymbol{\mathcal{T}}$ is also required to be the product of a stable and causal diagonal rational matrix with equal elements and $\mathbf{R}^{-1}$. Thus,

$$\boldsymbol{\mathcal{T}} = D_u \mathcal{T}' \mathbf{R}^{-1} \quad , \quad \text{(A.2)}$$

where $\mathcal{T}'$ is an arbitrary stable scalar transfer function. The factor D_u must be present in (A.2) to assure the stationarity of (A.1). By adding the variational term ξ_t to the proposed optimal estimate $\hat{h}_{t+k|t}$, a modified criterion value is obtained

$$\begin{aligned} \operatorname{tr}\mathbf{P}_\xi &= \operatorname{tr}\mathrm{E}\,(\tilde{h}_{t+k|t} - \xi_t)(\tilde{h}_{t+k|t} - \xi_t)^* \\ &= \operatorname{tr}\mathbf{P}_k - 2\mathrm{Re}(\operatorname{tr}\mathrm{E}\,\tilde{h}_{t+k|t}\xi_t^*) + \operatorname{tr}\mathrm{E}\,\xi_t\xi_t^* \quad \text{(A.3)} \end{aligned}$$

The minimal criterion value is obtained by adjusting the structurally constrained filter $\boldsymbol{\mathcal{L}}_k$ so that orthogonality is attained between $\tilde{h}_{t+k|t}$ and ξ_t. Then, the cross term in (A.3) will vanish, so no admissible modification ξ_t can improve the estimate. Thus, $\xi_t = 0$ will minimize $\operatorname{tr}\mathbf{P}_\xi$.

By invoking assumptions A2 and A3 and making use of equations (21),(29), (28) and (A.1), the cross-term in (A.3) can be expressed as

$$\begin{aligned} &\operatorname{tr}\mathrm{E}\,\tilde{h}_{t+k|t}\xi_t^* \\ &= \operatorname{tr}\mathrm{E}\left(\left(q^k\mathbf{I} - \frac{Q_k}{\beta}\mathbf{R}^{-1}\mathbf{R}\right)\frac{C}{D}e_t - \frac{Q_k}{\beta}\mathbf{R}^{-1}\eta_t\right) \\ &\quad \left(\boldsymbol{\mathcal{T}}\left(\mathbf{R}\frac{C}{D}e_t + \eta_t\right)\right)^* \quad \text{(A.4)} \end{aligned}$$

By using Parsevals formula [1],[38], this cross term is eliminated when

$$\operatorname{tr}\mathrm{E}\,\tilde{h}_{t+k|t}\xi_t^* = \frac{1}{2\pi j}\oint_{|z|=1}\operatorname{tr}\phi_{\tilde{h}_{t+k|t}\xi_t^*}\frac{dz}{z} = 0 \quad , \quad \text{(A.5)}$$

where $\operatorname{tr}\phi_{\tilde{h}_{t+k|t}\xi_t^*}$ is the trace of the cross spectral density matrix between $\tilde{h}_{t+k|t}$ and the variation ξ_t. This scalar function is readily obtained from (A.4), using (A.2) and Assumption A3, as

$$\begin{aligned} &\operatorname{tr}\left(\left(z^k\mathbf{I} - \frac{Q_k}{\beta}\right)\frac{C}{D}\mathbf{R}_e\frac{C_*}{D_*} - \frac{Q_k}{\beta}\mathbf{R}^{-1}\mathbf{R}_\eta\mathbf{R}^{-1}\right)D_{u*}\mathcal{T}'_* \\ &= \operatorname{tr}\frac{1}{D}[z^kC\mathbf{R}_eC_* - \frac{Q_k}{\beta}(C\mathbf{R}_eC_* + D\mathbf{R}^{-1}\mathbf{R}_\eta\mathbf{R}^{-1}D_*)]\frac{1}{D_{s*}}\mathcal{T}'_* \\ &\quad \text{(A.6)} \end{aligned}$$

since $D_{u*} = D_*/D_{s*}$. By invoking Assumption A4 and making use of the spectral factorization (31), (A.6) can thus be reformulated as

$$\begin{aligned}
&\phi_{\tilde{h}_{t+k|t}\xi_t^*} \\
&= \frac{1}{D}\left(z^k\gamma CC_* - \frac{Q_k}{\beta}(\gamma CC_* + DD_*)\right)\frac{\operatorname{tr}\mathbf{R}_{\bar{\eta}}}{D_{s*}}\mathcal{T}'_* \\
&= \frac{1}{D}(z^k\gamma CC_* - Q_k r\beta_*)\frac{\operatorname{tr}\mathbf{R}_{\bar{\eta}}}{D_{s*}}\mathcal{T}'_* \quad . \qquad \text{(A.7)}
\end{aligned}$$

In (A.7), we used that $\operatorname{tr} a\mathbf{X} = a\operatorname{tr}\mathbf{X}$ for scalars a and matrices $\mathbf{X}$. Orthogonality is attained if there are no poles inside the integration path $|z| = 1$ in (A.5). Since D_s and $\mathcal{T}'$ are stable, $(\operatorname{tr}\mathbf{R}_{\bar{\eta}}/D_{s*})\mathcal{T}'_*$ will have no poles in $|z| < 1$. The orthogonality requirement (A.5) is therefore fulfilled if

$$\frac{1}{D}\left(z^k\gamma CC_* - Q_k r\beta_*\right)\frac{1}{z} = L_{k*} \qquad \text{(A.8)}$$

where L_{k*} is a polynomial in positive powers of z only. Evidently, (A.8) coincides with (32).

The Diophantine equation (32) will always have a solution since $\beta_*(z)$ has zeros only in $|z| > 1$ while $z^{nD}D(z^{-1})$ is stable or marginally stable and thereby will have no factors in common with $\beta_*(z)$ [1]. Let $(\bar{Q}_k, \bar{L}_{k*})$ be one solution pair to (32). Every solution to (32) can then be expressed as

$$(Q_k, L_{k*}) = (\bar{Q}_k - qDX, \bar{L}_{k*} + Xr\beta_*) \qquad \text{(A.9)}$$

where $X(q^{-1}, q)$ is an arbitrary polynomial. Since $\boldsymbol{\mathcal{L}}_k$ must be causal, $Q_k(q^{-1})$ is required to be a polynomial in powers of q^{-1} while L_{k*} from (A.8) is required to be a polynomial in q. Thus, $X = 0$ is the only choice, so (32) has a unique solution. The degrees (33) of this solution are determined by the matching of the highest powers in q^{-1} and q on both sides of (32).

Finally we shall verify that the estimation error

$$\tilde{h}_{t+k|t} = \left(q^k - \frac{Q_k}{\beta}\right)\frac{C}{D_sD_u}\mathbf{I}\,e_t - \frac{Q_k}{\beta}\mathbf{R}^{-1}\eta_t \qquad \text{(A.10)}$$

is stationary and zero mean with a finite value of $\operatorname{tr}\mathbf{P}_k$, even when the hypermodel is marginally stable. Since the learning filter $\boldsymbol{\mathcal{L}}_k^{opt} = (Q_k/\beta)\mathbf{R}^{-1}$ is stable and the noise η_t is assumed to be stationary, the last term of (A.10) has finite variance and zero mean. By substituting z for q and observing that

$$z^k\gamma CC_*\big|_{z_{D_u}} = rQ_k\beta_*\big|_{z_{D_u}} \;;\; r\beta\beta_*\big|_{z_{D_u}} = \gamma CC_*\big|_{z_{D_u}}$$

when evaluating (32) and (31) at the zeros of D_u, we obtain

$$z^k - \frac{Q_k}{\beta}\bigg|_{z_{D_u}} = z^k - \frac{1}{\beta}z^k\frac{\gamma CC_*}{r\beta_*}\bigg|_{z_{D_u}} = 0 \;. \qquad \text{(A.11)}$$

Thus, the poles on the unit circle of the first term of (A.10) are canceled by zeros. In other words, $q^k\beta - Q_k = D_uX$, for some polynomial $X(q^{-1})$. The tracking error (A.10) will therefore be generated by a stationary noise fed through a stable linear time-invariant system, so it will be stationary and have finite variance *if* $D_u(z^{-1})$ is exactly known ∎

B. Proof of Corollary 1, Lemma 1 and Lemma 2

Proof of Corollary 1: The solutions (34)-(37) are verified by direct substitution into (32). The recursions (38) and (39) are verified by noting that Q_{k+1} and L_{k+1*} will satisfy (32) for $k+1$

$$q^{k+1}\gamma CC_* = rQ_{k+1}\beta_* + qDL_{k+1*} \quad . \qquad \text{(B.1)}$$

Multiplying the left and the right hand sides of (B.1) by q^{-1} and making use of the assumed relation (38) yields

$$\begin{aligned}
q^k\gamma CC_* &= r(Q_k - DQ_0^k)\beta_* + DL_{k+1*} \\
&= rQ_k\beta_* + D(L_{k+1*} - rQ_0^k\beta_*) \;. \qquad \text{(B.2)}
\end{aligned}$$

The use of (39) in (B.2) reduces this equation to the Diophantine equation for lag k, which is by definition satisfied by $Q_k(q^{-1})$, $L_{k*}(q)$. Equations (40) and (41) can be verified in the same way ∎

Proof of Lemma 1: From (22),(24) and (29) we obtain an expression for the fictitious measurement

$$f_t = \mathbf{R}\hat{h}_{t|t-1} + \varphi_t\varepsilon_t = q^{-1}\frac{Q_1}{\beta}f_t + \varphi_t v_t$$

$$(\beta - q^{-1}Q_1)f_t = \beta\varphi_t\varepsilon_t$$

which by use of (34) becomes

$$f_t = \frac{\beta}{D}\varphi_t\varepsilon_t \quad . \qquad \text{(B.3)}$$

The k-step estimate (24),(29) can then be expressed as

$$\hat{h}_{t+k|t} = \boldsymbol{\mathcal{L}}_k f_t = \frac{Q_k}{D}\mathbf{R}^{-1}\varphi_t\varepsilon_t \quad . \qquad \text{(B.4)}$$

By evaluating (B.4) for $k = 0$ and $k = 1$ and subtracting, use (40) for $k = 1$ and noting from (35) that $L_0^{*1} = r - 1$, we obtain

$$\hat{h}_t - \hat{h}_{t|t-1} = \frac{Q_0 - q^{-1}Q_1}{D}\mathbf{R}^{-1}\varphi_t\varepsilon_t = \frac{r-1}{r}\mathbf{R}^{-1}\varphi_t\varepsilon_t \quad ,$$

which is (15) with $\mu = (r-1)/r$. The relation $1 - 1/r = Q_0^0$ follows from (36) ∎

Proof of Lemma 2: We will use the small gain theorem, see e.g. [47] and observe that the inverse of the estimator (24), for $k = 0$ can be expressed as

$$f_t = \mathbf{R}\frac{\beta}{Q_0}\hat{h}_{t|t} = \mathbf{R}\frac{r\beta}{r\beta - D}\hat{h}_{t|t} = \mathbf{R}\frac{1}{1 - \frac{D}{r\beta}}\hat{h}_{t|t} \quad ,$$

in which $D/r\beta$ is always stable. According to the small gain theorem, the inverse estimator will be stable if

$$\left|\frac{D(e^{\jmath\omega})}{r\beta(e^{\jmath\omega})}\right| < 1 \quad \forall\ \omega$$

Utilizing that β is a spectral factor, it readily follows from (31) that

$$|r\beta|^2 = r\left(\gamma|C|^2 + |D|^2\right) > |D|^2\ ,\ |z| = 1\ , \qquad \text{(B.5)}$$

since $\gamma|C(z^{-1})|^2 > 0$ and $r > 1$ for all $\gamma > 0$. Thus, all zeros of $Q_0(z^{-1})$ are in $|z| < 1$ ∎

C. Proof of Theorem 2

Lemma 3 [39]. Consider the following second order spectral factorization, with real-valued coefficients α_0, α_1, α_2.

$$\begin{aligned} r\beta(q^{-1})\beta_*(q) &= r(1+\beta_1 q^{-1}+\beta_2 q^{-2})(1+\beta_1 q+\beta_2 q^2) \\ &= \alpha_0 + \alpha_1(q+q^{-1}) + a_2(q^2+q^{-2}) \quad \text{(C.1)} \end{aligned}$$

where r is a positive scalar. Then, the solution r and $\beta(q^{-1})$ to (C.1) with zeros in $|z| \leq 1$ is given by

$$r = \frac{\xi + \sqrt{\xi^2 - 4\alpha_2^2}}{2} \qquad \beta_1 = \frac{\alpha_1}{r+\alpha_2} \qquad \beta_2 = \frac{\alpha_2}{r} \quad \text{(C.2)}$$

where

$$\xi = \frac{\alpha_0}{2} - \alpha_2 + \sqrt{\left(\frac{\alpha_0}{2}+\alpha_2\right)^2 - \alpha_1^2} \qquad \text{(C.3)}$$

∎

With $D(q^{-1}) = 1 + d_1 q^{-1} + d_2 q^{-2}$ and $C(q^{-1}) = 1$, the right-hand side of (C.1) is by (31) obtained as

$$\gamma + (1+d_1 q^{-1}+d_2 q^{-2})(1+d_1 q+d_2 q^2)$$

$$= \gamma + 1 + d_1^2 + d_2^2 + d_1(1+d_2)(q+q^{-1}) + d_2(q^2+q^{-2})\ .$$

From (C.2), we then obtain

$$\begin{aligned} \beta(q^{-1}) &= 1 + \frac{d_1(1+d_2)}{r+d_2}q^{-1} + \frac{d_2}{r}q^{-2} \qquad \text{(C.4)} \\ &= 1 + \frac{d_1(1+d_2)(1-\mu)}{1+d_2(1-\mu)}q^{-1} + d_2(1-\mu)q^{-2} \end{aligned}$$

where we in the last line utilized $r = 1/(1-\mu)$ from (43). The expression (45) for $Q_k(q^{-1})$ follows from (38). With $nQ_k = 1$ from (33), we obtain

$$Q_{k+1}(q^{-1}) = q(Q_k(q^{-1}) - D(q^{-1})Q_0^k) = Q_1^k - d_1 Q_0^k - d_2 Q_0^k q^{-1}$$

$$= \begin{pmatrix} 1 & q^{-1} \end{pmatrix} \begin{pmatrix} -d_1 & 1 \\ -d_2 & 0 \end{pmatrix} \begin{pmatrix} Q_0^k \\ Q_1^k \end{pmatrix}. \qquad \text{(C.5)}$$

Clearly, from (C.5) we have

$$\begin{pmatrix} Q_0^{k+1} \\ Q_1^{k+1} \end{pmatrix} = \begin{pmatrix} -d_1 & 1 \\ -d_2 & 0 \end{pmatrix} \begin{pmatrix} Q_0^k \\ Q_1^k \end{pmatrix} \qquad \text{(C.6)}$$

and by iterating (C.6) j times we obtain, for $k = 0$,

$$\begin{pmatrix} Q_0^j \\ Q_1^j \end{pmatrix} = \begin{pmatrix} -d_1 & 1 \\ -d_2 & 0 \end{pmatrix}^j \begin{pmatrix} Q_0^0 \\ Q_1^0 \end{pmatrix} \qquad \text{(C.7)}$$

which provides the coefficients for $Q_j(q^{-1})$ expressed in the coefficients of $Q_0(q^{-1})$. We proceed by deriving an explicit expression for $Q_0(q^{-1})$. From (36) we obtain

$$Q_0(q^{-1}) = \left(1 - \frac{1}{r}\right) + \left(\beta_1 - \frac{d_1}{r}\right)q^{-1} + \left(\beta_2 - \frac{d_2}{r}\right)q^{-2}\ . \qquad \text{(C.8)}$$

Substitution of the coefficients of $\beta(q^{-1})$ from the first line of (C.4) into (C.8) yields

$$Q_0(q^{-1}) = \left(1 - \frac{1}{r}\right) + \frac{d_1 d_2(r-1)}{(r+d_2)r}q^{-1}\ . \qquad \text{(C.9)}$$

The relation (43) gives $r = 1/(1-\mu)$, so with (46), we obtain $Q_0^0 = \mu, Q_1^0 = \mu p$. Used in (C.7) with k substituted for j, this gives (45) ∎

References

[1] A. Ahlén and M. Sternad, "Wiener filter design using polynomial equations," *IEEE Trans. on Signal Proc.*, vol. 39, pp. 2387-2399, 1991.

[2] A. Ahlén and M. Sternad, "Derivation and design of Wiener filters using polynomial equations," In C T Leondes ed., *Control and Dynamic Systems,* vol 64: Stochastic Techniques in Digital Signal Processing, Academic Press, New York, 1994, pp. 353-418.

[3] B. D. O. Anderson and J. B. Moore, *Optimal Filtering.* Prentice Hall, Englewood Cliffs, NJ, 1979.

[4] A. Benveniste, M. Métivier and P. Priouret, *Adaptive Algorithms and Stochastic Approximations.* Springer-Verlag, Berlin Heidelberg, 1990.

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

[6] M.-C. Chiu and C. Chao, "Analysis of LMS-adaptive MLSE equalization on multipath fading channels," *IEEE Trans. on Commun.*, vol. 44, pp. 1684-1692, 1996.

[7] A. P. Clark, *Adaptive Detectors for Digital Modems.* Pentech Press, London, 1989.

[8] L. Davis, I. Collings and R. Evans, "Coupled estimators for equalization of fast-fading mobile channels," *IEEE Trans. on Commun.* vol. 46, pp. 1262-1265, 1998.

[9] C. J. Demeure and C. T. Mullis, "A Newton-Raphson method for moving average spectral factorization using the Euclid algorithm," *IEEE Trans. on Acoustics, Speech and Signal Proc.*, vol 38, pp. 1697-1709, 1990.

[10] E. Eleftheriou and D. D. Falconer, "Tracking properties and steady-state performance of RLS adaptive filter algorithms," *IEEE Trans. on Acoustics, Speech, and Signal Proc.*, vol. 34, pp. 1097-1109, 1986.

[11] B. Farhang-Boroujeny, "Fast LMS/Newton algorithms based on autoregressive modeling and their application to acoustic echo cancellation," *IEEE Transactions on Signal Processing*, vol. 45, pp. 1987-2000, 1997.

[12] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas." *Bell Labs Tech. J.* vol. 2, pp. 41-59, 1996

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

[14] S. Gazor, "Prediction in LMS-type adaptive algorithms for smoothly time-varying environments," *IEEE Trans. on Signal Proc.*, vol 47, pp. 1735-1739, 1999.

[15] Y. Grenier, "Time-dependent ARMA modeling of nonstationary signals," *IEEE Trans. on Acoustics, Speech and Signal Proc.*, vol. 31, pp. 899-911, 1983.

[16] S. Hariharan and A. P. Clark, "HF channel estimation using a fast transversal filter algorithm," *IEEE Trans. on Acoustics, Speech, and Signal Proc.*, vol. 38, pp. 1353-1362, 1990.

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

[18] S. Haykin, A. H. Sayed, J. R. Zeidler, P. Yee, P. C. Wei, "Adaptive tracking of linear time-variant systems by extended RLS algorithms," *IEEE Trans. on Signal Proc.*, vol 45, pp. 1118-1128, 1997.

[19] "Polynomial toolbox, version 2.5," See http://www.polyx.com.

[20] 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 Proc.*, vol. 29, pp. 722-735, 1981.

[21] K. Jamal, G. Brismark and B. Gudmundson, "Adaptive MLSE performance on the D-AMPS 1900 channel," *IEEE Trans. on Vehicular Technol.*, vol. 46, pp. 634-641, 1997.

[22] G. Kitagawa and W. Gersch, "A smoothness priors time–varying AR coefficient modeling of nonstationary covariance time series," *IEEE Trans. on Automatic Control*, vol. 30, pp. 48-56, 1985.

[23] G. Kubin, "Adaptation in rapidly time–varying environments using coefficient filters," *IEEE ICASSP*, April 1991, Toronto, Ontario, Canada, vol. 3, pp. 2097-2100.

[24] G. Kubin, "Coefficient filtering - a common framework for the adaptation in time-varying environments," *Proc. of Adaptive Algorithms: Applications and non classical schemes,* pp. 91-110, March 1991, Madrid, Spain.

[25] H. J. Kushner and A. Shwartz, "Weak convergence and asymptotic properties of adaptive filters with constant gain," *IEEE Trans. on Information Theory*, vol. 30, pp. 177-182, 1984.

[26] Y-T. Lee and H. F. Silverman, "On a general time-varying model for speech signals," *IEEE ICASSP*, April 1988, New York, NY, vol. 1, pp. 95-98.

[27] L. Lindbom, "Simplified Kalman estimation of fading mobile radio channels: high performance at LMS computational load," *IEEE ICASSP* 1993, Minneapolis, MN, USA, April 27-30, vol. 3, pp. 352-355.

[28] L Lindbom, *A Wiener filtering approach to the design of tracking algorithms: with application in mobile radio communications.* PhD Thesis, Department of Technology, Uppsala University, Sweden, 1995. www.signal.uu.se/Publications/abstracts/a951.html

[29] M. Sternad, L. Lindbom and A. Ahlén, "Tracking of time-varying systems, Part I: Wiener design of algorithms with time-invariant gains," Report, www.signal.uu.se/Publications/abstracts/r001.html

[30] A. Ahlén, L. Lindbom and M. Sternad, "Tracking of time-varying systems, Part II: Analysis of stability and performance of adaptive algorithms with time-invariant gains," Report, www.signal.uu.se/Publications/abstracts/r002.html

[31] L. Lindbom, A. Ahlén, M. Sternad and M. Falkenström, "Tracking of time-varying mobile radio channels. Part II: A case study." *IEEE Transactions on Communications,* vol. 49, pp. xx-yy, 2001. www.signal.uu.se/Publications/abstracts/r004.html

[32] L. Ljung and S. Gunnarsson, "Adaptation and tracking in system identification-A survey," *Automatica*, vol. 26, pp. 7-21, 1990.

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

[34] P. S. Maybeck, *Stochastic Models, Estimation and Control, Vol I-III,* Academic Press, New York, 1979.

[35] K. J. Molnar and G. E. Bottomley, "Adaptive array processing MLSE receivers for TDMA digital cellular PCS communications." *IEEE J. Selected Areas in Commun.*, vol. 16, pp. 1340-1351, 1998.

[36] M. Niedźwiecki, "Recursive functional series modeling estimators for identification of time-varying plants–More bad news than good?," *IEEE Trans. on Automatic Control*, vol. 35, pp. 610-616, 1990.

[37] M. Niedźwiecki, *Identification of Time-varying Processes*, Wiley, Chichester, UK, 2000.

[38] A. V. Oppenheim and R. W. Shafer, *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1975.

[39] V. Peterka, "Predictor-based self-tuning control," *Automatica*, vol. 20, pp. 39-50, 1984.

[40] A. H. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering," *IEEE Signal Processing Magazine*, vol 11, no. 3, pp. 18-60, 1994.

[41] U. Shaked. "A generalized transfer function approach to linear stationary filtering and steady-state optimal control problems," *International Journal of Control*, vol. 29, pp. 741-770, 1976.

[42] C. Tidestav, M. Sternad and A. Ahlén "Reuse Within a Cell - Interference Rejection or Multiuser Detection?" *IEEE Trans. on Commun.*, vol 47, pp. 1511-1522, 1999.

[43] J. R. Treichler, C. R. Johnson, Jr., and M. G. Larimore, *Theory and Design of Adaptive Filters.* Wiley, NY, 1987.

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

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

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

[47] K. J. Åström and B. Wittenmark, *Adaptive Control.* third ed. Addison-Wesley, 1996.