H2 DESIGN OF ADAPTATION LAWS WITH CONSTANT GAINS

Mikael Sternad, Lars Lindbom and Anders Ahlén
Signals and Systems, Uppsala University, PO Box 528, SE-75120, Uppsala, Sweden.
{ms,aa}@signal.uu.se, Lars.Lindbom@ks.ericsson.se.

Keywords: adaptive estimation, adaptive filtering.

Abstract

We present a method for optimizing adaptation laws that are generalizations of the LMS algorithm. The proposed technique has been applied successfully for designing estimators of rapidly time-varying mobile radio channels. The estimators apply time-invariant filtering on the instantaneous gradient. Time-varying parameters of linear regression models are estimated in situations where the regressors are stationary or have slowly time-varying properties. The structure and gains of these adaptation laws are optimized in MSE for time-variations modeled as correlated stochastic processes. The aim is to systematically use such prior information to provide filtering, prediction or fixed lag smoothing estimates for arbitrary lags. Our design method is based on a novel transformation that recasts the adaptation problem into a Wiener filter design problem. The filter works in open loop for slow parameter variations while a time-varying closed loop is important for fast variations. In closed loop, the filter design is performed iteratively. The solution at one iteration can be obtained by a bilateral Diophantine polynomial matrix equation and a spectral factorization. For white noise, the Diophantine equation has a closed-form solution. When one filter is known, a set of predictors and smoothers, up to a predefined prediction horizon or smoothing lag, is obtained by analytical expressions.

1 Introduction

Adaptation algorithms that estimate time-varying parameters of linear regression models are fundamental tools in signal processing, control and digital communication. When prior information about the statistics of the variations are available, Kalman estimators minimize the tracking mean square error (MSE) among all linear estimators. Motivated by the need to develop channel estimators of low computational complexity for use in mobile radio systems, we have recently proposed a class of adaptation laws that for channel tracking attain close to the optimal Kalman performance, yet require a computational complexity close to that of LMS [1, 2, 11, 18]. An early version of the proposed algorithm has successfully been used on ANSI-136 channels [10, 13], and a case study on this application can be found in [19]. Instead of a time-varying gain computed via a Riccati update, these algorithms use a time-invariant filter that operates on the instantaneous gradient vector. The present paper discusses the H_2 design of such filters from a control perspective. Optimization of the adaptation law then corresponds to LQG or H_2 optimization of a time-invariant feedback regulator for a time-varying plant. Our tracking design solves this problem by a loop transformation, explained in Section 3, followed by the iterative use of an open-loop Wiener filter design, outlined in Section 4. The Wiener solution is presented in polynomial form, which provides important structural insights.

Notation: 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 denoted 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 $|z| \leq 1$.

2 Outline of the Problem

Consider discrete-time and possibly complex-valued measurements generated by a linear regression

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

where y_t is the measured signal with n_y elements, v_t is a noise while φ_t^* is a $n_y|n_h$ regression matrix, which is known at time t. The regressors are assumed to be persistently exciting, so that their covariance matrix

$$\mathbf{R} \stackrel{\Delta}{=} \mathrm{E}\,[\varphi_t \varphi_t^*] \tag{2}$$

is nonsingular. Furthermore, $\mathbf{R}$ is here assumed constant and known, while in practice it may be slowly time-varying. The linear regression could represent a FIR model or an orthogonal series expansion with fixed poles, such as a Laguerre or Kautz model. However, AR- and ARX models are here excluded, since the use of old measurements as regressors could make φ_t^* highly nonstationary when h_t is rapidly time-varying. The aim is to estimate the parameter vector

$$h_t = (h_{0,t} \dots h_{n_h-1,t})^T \quad , \tag{3}$$

assuming n_h to be known. The model structure is thus assumed to include the true system. With time-varying parameters, we face a parameter *tracking* problem. Models that constrain the assumed variation of h_t, sometimes called *hypermodels* [3], must then be introduced, since we would otherwise have unknowns $h_1, h_2 \dots$ with more elements than the available measurements $y_1, y_2 \dots$ whenever $n_y < n_h$. A commonly used alternative is to introduce a local polynomial approximation [4, 14, 15]

$$h_t = \sum_{j=0}^{n} b_j t^j \quad , \quad t \in [1, N] \tag{4}$$

where the $n+1$ vectors b_j are to be estimated over a time window of length N. The resulting least squares problem for the model (1),(4) has an adequate number of equations if $Nn_y > (n+1)n_h$, but high polynomial degrees n will lead to ill-conditioned and noise-sensitive solutions, due to the wide magnitude range of the regressors. Less numerically sensitive and more flexible designs are obtained by introducing stochastic hypermodels.

We here consider linear time-invariant stochastic models

$$h_t = \boldsymbol{\mathcal{H}}(q^{-1}) e_t \quad , \tag{5}$$

where e_t is white noise with covariance matrix $\mathbf{R}_e$ and $\boldsymbol{\mathcal{H}}(q^{-1})$ is a $n_h|n_h$ matrix of stable or marginally stable transfer operators. Let

$$\tilde{h}_{t+k|t} \triangleq h_{t+k} - \hat{h}_{t+k|t} \quad , \tag{6}$$

where $\hat{h}_{t+k|t}$ is an estimate of h_{t+k} obtained at time t by filtering ($k = 0$), prediction ($k > 0$) or fixed lag smoothing ($k < 0$). Kalman trackers can be designed based on state-space realizations of (5), with (1) representing the measurement equation. They are the linear estimators that minimizes the parameter error covariance matrix

$$\mathbf{P}_{k,t} \triangleq \mathrm{E}\,[\tilde{h}_{t+k|t}\tilde{h}^*_{t+k|t}] \quad , \tag{7}$$

where the expectation is taken with respect to e_t in (5) and v_t in (1). However, since φ_t^* in (1) is time-varying, the Kalman estimator will not converge to steady state. A time-varying gain matrix has to be updated via the Riccati equation, so the estimator will have relatively high computational complexity. For scalar y_t, Kalman predictors and filters can be represented by

$$\hat{h}_{t+k|t} = \boldsymbol{\mathcal{M}}_{k,t}(q^{-1})\varphi_t\varepsilon_t \quad , \tag{8}$$

where $\boldsymbol{\mathcal{M}}_{k,t}$ is a rational filter with time-varying coefficients, while

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

and $\varphi_t\varepsilon_t$ is the negative instantaneous gradient of $|\varepsilon_t|^2$ with respect to $\hat{h}_{t|t-1}$.

A class of adaptation laws related to (8) but with much lower complexity is obtained by using pre-designed linear time-invariant filters that operate on $\varphi_t\varepsilon_t$. The aim is to minimize the tracking error covariance matrix (7) asymptotically for $t \to \infty$ (after the initial transients), for arbitrary k. The filter should be designed to provide estimates with an appropriate amount of coupling and inertia. Based on (1),(5), the time-invariant filter $\boldsymbol{\mathcal{M}}_k(q^{-1})$ in

$$\varepsilon_t = y_t - \varphi_t^* \hat{h}_{t|t-1} \tag{9}$$

$$\hat{h}_{t+k|t} = \boldsymbol{\mathcal{M}}_k(q^{-1})\varphi_t\varepsilon_t \tag{10}$$

is thus to be optimized under various constraints and assumptions. The LMS algorithm

$$\hat{h}_{t+1|t} = \frac{\mu}{1-q^{-1}}\mathbf{I}\ \varphi_t\varepsilon_t \quad , \tag{11}$$

where $\mu > 0$ is a scalar gain, constitutes a simple special case. However, LMS algorithms can be shown to have optimal structure only when the parameters h_t behave as uncorrelated random walks and the regressor covariance matrix (2) is diagonal [18]. The general structure (10) provides considerably more flexibility.

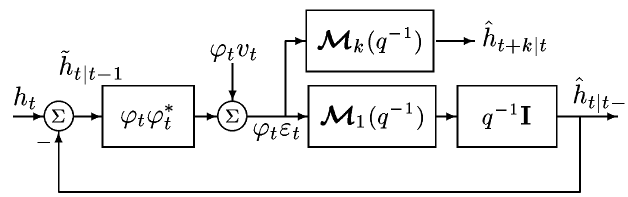

Figure 1: The one-step predictor could be seen as a linear time-invariant feedback regulator for a time-varying system.

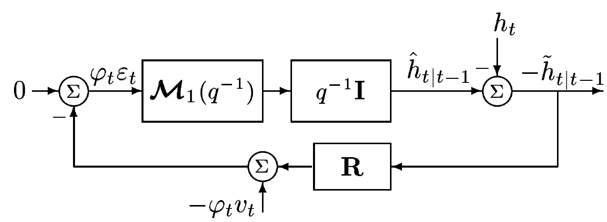

Figure 2: When $\varphi_t\varphi_t^*$ is approximated by $\mathbf{R}$, the design of $\boldsymbol{\mathcal{M}}_1(q^{-1})$ corresponds to the design of a minimum variance feedback regulator operating on a time-invariant system with measurement noise and a colored output disturbance $-h_t$.

For any desired k, a one-step predictor $\boldsymbol{\mathcal{M}}_1(q^{-1})$ must also be designed, due to the presence of $\hat{h}_{t|t-1}$ in (9). By (1), (9) and (10), this prediction problem corresponds to the design of a time-invariant feedback controller for a time-varying system, as illustrated by Figure 1. One approximate solution could be obtained by substituting the block $\varphi_t\varphi_t^*$ by its time-invariant average $\mathbf{R}$. We then have the minimum variance feedback regulator problem of Figure 2, where $-\tilde{h}_{t|t-1}$ is the vector to be controlled, $-h_t$ modeled by (5) plays the role of an output disturbance, $\mathbf{R}$ is a sensor matrix and $-\varphi_t v_t$ is measurement noise. This problem can be solved by state space LQG design or by polynomial solutions [7, 9, 17].

However, we have shown in [1] that the error resulting from basing an MSE-optimal design on the approximation $\varphi_t\varphi_t^* = \mathbf{R}$ would be negligible only if the *parameter variations are slow*[1]. If this is not the case, a design based on Figure 2 will be misleading and provide a filter with too high gain. The result may be instability of the tracking loop. In the next section we introduce a loop transformation that moves time-varying feedback effects to an outer perturbation loop. This perturbation is then taken into account by iterative filter adjustment.

3 The Loop Transformation

The algorithm (10) can be expressed as a stable and causal filter, denoted the *learning filter* $\mathcal{L}_k(q^{-1})$, that operates on the signal $f_t \triangleq \varphi_t\varepsilon_t + \mathbf{R}\hat{h}_{t|t-1}$. Since

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

$$\hat{h}_{t+k|t} = \mathcal{M}_k(q^{-1})(\mathbf{I}+q^{-1}\mathbf{R}\mathcal{M}_1(q^{-1}))^{-1}f_t \triangleq \mathcal{L}_k(q^{-1})f_t \; . \tag{12}$$

Consider the signal prediction error (9) and insert (1) describing y_t, to obtain

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

Adding and subtracting $\mathbf{R}\tilde{h}_{t|t-1}$ on the right-hand side of (13) gives

$$\varphi_t\varepsilon_t = \mathbf{R}\tilde{h}_{t|t-1} + (\varphi_t\varphi_t^* - \mathbf{R})\tilde{h}_{t|t-1} + \varphi_t v_t \quad . \tag{14}$$

Define

$$Z_t = \varphi_t\varphi_t^* - \mathbf{R} \tag{15}$$

$$\eta_t = Z_t\tilde{h}_{t|t-1} + \varphi_t v_t \tag{16}$$

which are called the *autocorrelation matrix noise* [6] and the *gradient noise*, respectively. The signal f_t , to be called the fictitious measurement, can then by (14), (15) and (16) be expressed as

$$f_t = \mathbf{R}h_t + Z_t\tilde{h}_{t|t-1} + \varphi_t v_t = \mathbf{R}h_t + \eta_t \quad , \tag{17}$$

so Figure 1 can now be transformed into Figure 3. The design problem for our adaptation law has now been transformed into a *Wiener filter design problem* for $\mathcal{L}_k$, where $\mathbf{R}h_t$ is a rotated signal while η_t plays the role of noise. The gradient noise η_t is affected by the term $Z_t\tilde{h}_{t|t-1}$, here called the *feedback noise*. It is shown in [1] that the feedback noise is negligible for slow variations, so the learning filter will then operate essentially in open loop, with $\eta_t \approx \varphi_t v_t$. The learning filter can then be designed directly as a Wiener filter, see Figure 4 and Section 3. With $Z_t\tilde{h}_{t|t-1}$ negligible, the approximation $\varphi_t\varphi_t^*\tilde{h}_{t|t-1} \approx \mathbf{R}\tilde{h}_{t|t-1}$ is valid by (15) and a Wiener design of $\mathcal{L}_1$ is just another way of designing the closed loop of Figure 2. Stability and convergence in MSE is then guaranteed by stability of the learning filter.[2]

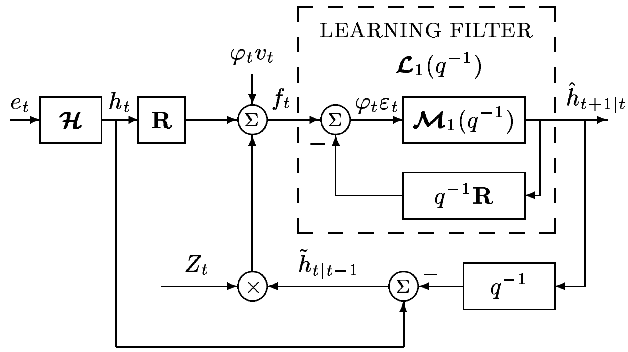

Figure 3: The one-step prediction learning filter operates in open loop for slow variations, when $Z_t\tilde{h}_{t|t-1}$ can be neglected. For fast variations, the feedback noise $Z_t\tilde{h}_{t|t-1}$ has to be taken into account.

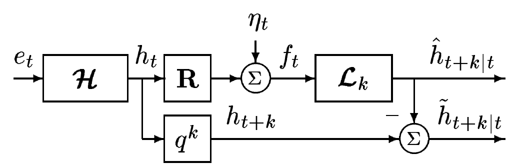

Figure 4: The filter design problem. The vector h_{t+k} is to be estimated from measurements f_t, such that the steady state tracking error covariance matrix is minimized.

For fast variations, the feedback may endanger stability. A sufficient but conservative condition for stability for a given design is provided by the small gain theorem [21]: If $\mathcal{L}_1 Z_t$ is causal and L_p-stable, stability is preserved if

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

Although the feedback noise will be dependent on old estimation errors, the *correlation* between the innovation sequence of η_t and previous estimation errors $\tilde{h}_{\tau|\tau-1}, \tau \leq t$ will in general be low. The reason is that multiplication by the matrix Z_t acts as a scrambler. The scrambling becomes more efficient for more rapidly time-varying regressor elements. When the correlation can be neglected, an optimal linear filter design can be performed as an open loop Wiener design, since higher order statistical dependencies will not affect the MSE-optimal linear design. However, the covariance matrix of the total gradient noise η_t, including the feedback noise, must be estimated iteratively, since the variance of the feedback noise will depend on the learning filter, which generates the estimation error $\tilde{h}_{t|t-1}$.

[1] See [12] and [1, 11] for quantifications of the degree of nonstationarity and the property of slow variations.

[2] For the LMS algorithm (11), stability of the learning filter (12) corresponds to the condition $0 < \mu < 2/\lambda_{\max}$ for convergence in the mean [8], where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{R}$.

An iterative design may proceed as follows.

1. Perform a one-step predictor design for slow time-variations, i.e. use $\eta_t = \varphi_t v_t$ to design $\mathcal{L}_1(q^{-1})$ by Theorem 1 below. Verify that the closed loop of Figure 3 is stable, so that the resulting error $\tilde{h}_{t|t-1}$ is stationary. If not, scale up the assumed covariance function of η_t to decrease the gain of $\mathcal{L}_1(q^{-1})$.

2. Based on a long simulation of $h_t = \mathcal{H}(q^{-1})e_t$ and on the corresponding estimate $\hat{h}_{t+1|t} = \mathcal{L}_1(q^{-1})f_t$, obtain an estimated gradient noise time series

$$\hat{\eta}_t = f_t - \mathbf{R}h_t = \varphi_t\varepsilon_t - \mathbf{R}(h_t - \hat{h}_{t|t-1}) \quad . \qquad (19)$$

Obtain an estimate of the covariance function of η_t by using sample averages over $\hat{\eta}_t$.[3]

3. Design a new estimator $\mathcal{L}_1(q^{-1})$ by Theorem 1.

Repeat steps 2. and 3. until the difference in consecutive estimates $\hat{h}_{t+1|t}$ becomes small. Then, construct an estimator for the desired lag k. For more details and a design example, see Section V of [18].

4 Learning Filter Optimization

The transfer operator $\mathcal{L}_k(q^{-1})$ will now be adjusted to minimize (7) for $t \to \infty$, when $\mathcal{H}(q^{-1})$ in (5) is assumed known and the properties of η_t are assumed given. The learning filter is designed under the constraint of stability, under the following assumptions:

Assumption A1. The sequence $\{\varphi_t^*\}$ is stationary and known up to time t, with a known nonsingular autocorrelation matrix $\mathbf{R}$ □
Assumption A2: The gradient noise η_t is stationary with zero mean and is modeled by a known vector ARMA process

$$\eta_t = \frac{1}{N(q^{-1})}\boldsymbol{M}(q^{-1})\nu_t \ , \qquad (20)$$

where $\boldsymbol{M}$ is a $n_h|n_h$ polynomial matrix of degree n_M, N is a stable polynomial of degree n_N, while $\mathrm{E}\,\nu_t\nu_t^* = \mathbf{I}$ □
Assumption A3: The correlation of the innovation sequence ν_t of the gradient noise with h_{t-i} and with $\hat{h}_{t-i|t-i-1}, i \geq 0$ is negligible □
Assumption A4. The linear regression coefficients are described by a stochastic process

$$h_t = \mathcal{H}(q^{-1})e_t = \boldsymbol{D}(q^{-1})^{-1}\boldsymbol{C}(q^{-1})e_t \ , \qquad (21)$$

where e_t is white, stationary and zero mean with nonsingular covariance matrix $\mathbf{R}_e$ and

$$\boldsymbol{D}(q^{-1}) \ = \ D_u(q^{-1})\boldsymbol{D}_s(q^{-1}) \ . \qquad (22)$$

[3] The gradient noise is often white, which simplifies the design.

Above, the polynomial matrix $\boldsymbol{C}(q^{-1})$ of degree n_C is assumed monic and stable, $\boldsymbol{D}(q^{-1})$ is a monic polynomial matrix with degree n_D, $D_u(q^{-1})$ is a polynomial with zeros on $|z| = 1$ and $\boldsymbol{D}_s(q^{-1})$ is a stable polynomial matrix □

Assumption A4 implies that e.g. random walks, integrated random walks and filtered random walk models can be considered, but that the unstable dynamics $D_u(q^{-1})$ must then affect all the elements of h_t. We are now ready to state the following result.

Theorem 1: *The optimal learning filter.* Under Assumptions A1-A4, the stable and causal learning filter minimizing the asymptotic covariance matrix (7) is

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

$$\hat{h}_{t+k|t} = \mathcal{L}_k^{opt} f_t = \boldsymbol{D}_s^{-1}\boldsymbol{Q}_k\boldsymbol{\beta}^{-1}N\boldsymbol{D}_s\mathbf{R}^{-1}f_t \ , \qquad (24)$$

where the polynomial matrix $\boldsymbol{\beta}(q^{-1})$ of dimension $n_h|n_h$ and degree $n_\beta = \max(n_C + n_N, n_D + n_M)$ is the stable spectral factor obtained from

$$\boldsymbol{\beta}\boldsymbol{\beta}_* = \boldsymbol{C}\mathbf{R}_e\boldsymbol{C}_*NN_* + \boldsymbol{D}\mathbf{R}^{-1}\boldsymbol{M}\boldsymbol{M}_*\mathbf{R}^{-1}\boldsymbol{D}_* \ . \qquad (25)$$

The unique solution to the bilateral Diophantine equation

$$q^k\boldsymbol{C}\mathbf{R}_e\boldsymbol{C}_*N_* = \boldsymbol{Q}_k\boldsymbol{\beta}_* + q\boldsymbol{D}\boldsymbol{L}_{k*} \qquad (26)$$

provides polynomial matrices $\boldsymbol{Q}_k(q^{-1})$ and $\boldsymbol{L}_{k*}(q)$ of dimension $n_h|n_h$, with generic degrees

$$n_Q = \max(n_C - k, n_D - 1) \ , \quad n_L = \max(n_C + n_N + k, n_\beta) - 1 \qquad (27)$$

respectively. The estimation error $\tilde{h}_{t+k|t}$ will be stationary with finite covariance matrix and zero mean.
Proof: See [18] or [11] □

Note that only two design equations (25), (26) are required. This would be far from obvious if the general solutions of [7, 9] were applied to the problem of Figure 2. The Diophantine equation (26) is guaranteed to be solvable also for marginally stable $\boldsymbol{D}(z^{-1})$. Under Assumption A4, $\boldsymbol{C}$ is assumed stable and $\mathbf{R}_e$ has full rank, so $\boldsymbol{C}(z^{-1})\mathbf{R}_e\boldsymbol{C}(z)$ will have full rank on $|z| = 1$. Therefore, the spectral factorization (25) is guaranteed to deliver a spectral factor $\boldsymbol{\beta}$ with a leading matrix β_0 of full rank and a stable and causal inverse.

The inverse of the regressor covariance matrix will appear as a right factor in all learning filters (24). If $\mathbf{R}$ is unknown, its inverse $\mathbf{R}^{-1}$ can be estimated recursively with well-known methods at the price of increasing the complexity to a level similar to that of RLS. It is important to note that the time-scales used in the estimation of h_t and in the estimation of the regressor covariance can and should be separated. Since $\mathbf{R}^{-1}$ is assumed constant or perhaps slowly time-varying, a long data window can be used for estimating it accurately even when the variations of h_t are fast.

From (23) (20) and (21), the spectral density of the fictitious signal f_t is, under Assumption A3, given by

$$\begin{aligned}\phi_f &= \mathbf{R}\boldsymbol{D}^{-1}\boldsymbol{C}\mathbf{R}_e\boldsymbol{C}_*\boldsymbol{D}_*^{-1}\mathbf{R} + \frac{1}{NN_*}\boldsymbol{M}\boldsymbol{M}_* \\ &= \mathbf{R}\boldsymbol{D}^{-1}N^{-1}\boldsymbol{\beta}\boldsymbol{\beta}_*N_*^{-1}\boldsymbol{D}_*^{-1}\mathbf{R} \quad (28)\end{aligned}$$

where (25) was used in the last equality. The innovations representation of f_t is thus given by

$$f_t = \mathbf{R}\boldsymbol{D}^{-1}N^{-1}\boldsymbol{\beta}\epsilon_t \;\Leftrightarrow\; \epsilon_t = \boldsymbol{\beta}^{-1}N\boldsymbol{D}\mathbf{R}^{-1}f_t \quad (29)$$

where the innovation sequence ϵ_t is white, with zero mean and unit covariance matrix. When $\boldsymbol{D}(q^{-1})$ has zeros on the unit circle, (29) corresponds to a generalized innovation model [16]. By defining the signal

$$\bar{\epsilon}_t \triangleq \frac{1}{D_u(q^{-1})}\epsilon_t \;=\; \boldsymbol{\beta}^{-1}N\boldsymbol{D}_s\mathbf{R}^{-1}f_t \quad (30)$$

the adaptation law (23), (24) can be expressed as

$$\varepsilon_t = y_t - \varphi_t^*\hat{h}_{t|t-1} \quad (31)$$

$$\bar{\epsilon}_t = \boldsymbol{\beta}^{-1}N\boldsymbol{D}_s(\mathbf{R}^{-1}\varphi_t\varepsilon_t + \hat{h}_{t|t-1}) \quad (32)$$

$$\hat{h}_{t+k|t} = \boldsymbol{D}_s^{-1}\boldsymbol{Q}_k\bar{\epsilon}_t \;. \quad (33)$$

The product $\mathbf{R}^{-1}\varphi_t$ can be updated efficiently, with a computational complexity proportional to n_h, for scalar FIR models with autoregressive inputs [5].

Corollary 1. *The Wiener optimized filter $\boldsymbol{\mathcal{M}}_k$.* The estimator (10) optimized by Theorem 1 is

$$\hat{h}_{t+k|t} = \boldsymbol{\mathcal{M}}_k(q^{-1})\varphi_t\varepsilon_t \;=\; \boldsymbol{D}_s^{-1}\boldsymbol{Q}_k\boldsymbol{\mathcal{R}}\mathbf{R}^{-1}\varphi_t\varepsilon_t \quad (34)$$

where the causal rational matrix $\boldsymbol{\mathcal{R}}(q^{-1})$ is given by

$$\boldsymbol{\mathcal{R}} = [\boldsymbol{\beta} - q^{-1}N\boldsymbol{Q}_1]^{-1}N\boldsymbol{D}_s = \frac{1}{D_u(q^{-1})}\boldsymbol{X}_1^{-1}N\boldsymbol{D}_s \quad (35)$$

where $\boldsymbol{X}_1(q^{-1})$ is a polynomial matrix which solves

$$\boldsymbol{\beta} - q^{-1}\boldsymbol{Q}_1N = D_u\boldsymbol{X}_1 \;. \quad (36)$$

Proof. Multiply both sides of (32) from the left by $\boldsymbol{\beta}$ and then substitute the expression for $q^{-1}\hat{h}_{t+1|t}$, obtained from (33) with $k=1$, into (32). We obtain

$$\boldsymbol{\beta}\bar{\epsilon}_t = N\boldsymbol{D}_s\mathbf{R}^{-1}\varphi_t\varepsilon_t + q^{-1}N\boldsymbol{Q}_1\bar{\epsilon}_t \;.$$

Thus,

$$\bar{\epsilon}_t = \boldsymbol{\mathcal{R}}\mathbf{R}^{-1}\varphi_t\varepsilon_t \;.$$

The use of this expression in (33) gives (34)-(35). The equation (36) is derived in Appendix A of [18] □

Note that $\mathbf{R}^{-1}$ will always be a right factor of the optimal $\boldsymbol{\mathcal{M}}_k$ and that $\boldsymbol{D}_s^{-1}$ will be a left factor. While the learning filter $\boldsymbol{\mathcal{L}}_k(q^{-1})$ must be stable, the matrix $\boldsymbol{\mathcal{M}}_1(q^{-1})$ in Figure 1 need not be stable, since it works within a feedback loop. In fact, by the last expression of (35), D_u^{-1} will be present in all elements of $\boldsymbol{\mathcal{R}}$ in accordance with the internal model principle.

5 Recursive and Analytical Solutions to the Wiener Equations

The solution for $k=1$ will always be required. When several estimation horizons are of interest, we need to solve the equations in Theorem 1 for one value of k only. We believe this property of Diophantine equations used in Wiener filter design to be of independent interest.

Corollary 2. Let $\boldsymbol{Q}_k(q^{-1})$ and $\boldsymbol{L}_{k*}(q)$ solve (26) for lag k, having leading coefficients $\mathbf{Q}_0^k$ and $\mathbf{L}_0^{k*}$. Then,

$$\boldsymbol{Q}_{k+1}(q^{-1}) \;=\; q\left(\boldsymbol{Q}_k(q^{-1}) - \boldsymbol{D}(q^{-1})\mathbf{Q}_0^k\right) \quad (37)$$

$$\boldsymbol{L}_{k+1*}(q) \;=\; q\boldsymbol{L}_{k*}(q) + \mathbf{Q}_0^k\boldsymbol{\beta}_*(q) \quad (38)$$

solve the Diophantine equation (26) for lag $k+1$ and

$$\boldsymbol{Q}_{k-1}(q^{-1}) = q^{-1}\boldsymbol{Q}_k(q^{-1}) + \boldsymbol{D}(q^{-1})\mathbf{L}_0^{k*}(\beta_0^*)^{-1} \quad (39)$$

$$\boldsymbol{L}_{k-1*}(q) = q^{-1}\left(\boldsymbol{L}_{k*}(q) - \mathbf{L}_0^{k*}(\beta_0^*)^{-1}\boldsymbol{\beta}_*(q)\right) \quad (40)$$

constitute the solution to (26) for lag $k-1$, where β_0^* is the leading coefficient of $\beta_*(q)$.

Proof: It follows from the Diophantine equation (26) that $\boldsymbol{Q}_{k+1}$ and $\boldsymbol{L}_{k+1*}$ should satisfy

$$q^{k+1}\boldsymbol{C}\mathbf{R}_e\boldsymbol{C}_*N_* = \boldsymbol{Q}_{k+1}\boldsymbol{\beta}_* + q\boldsymbol{D}\boldsymbol{L}_{k+1*} \;. \quad (41)$$

Multiplying (41) by q^{-1} and using (37) yields

$$\begin{aligned}q^k\boldsymbol{C}\mathbf{R}_e\boldsymbol{C}_*N_* &= (\boldsymbol{Q}_k - \boldsymbol{D}\mathbf{Q}_0^k)\boldsymbol{\beta}_* + \boldsymbol{D}\boldsymbol{L}_{k+1*} \\ &= \boldsymbol{Q}_k\boldsymbol{\beta}_* + \boldsymbol{D}(\boldsymbol{L}_{k+1*} - \mathbf{Q}_0^k\boldsymbol{\beta}_*) \;.\end{aligned}$$

The use of (38) reduces this equation to the Diophantine equation for lag k, which is by definition satisfied by $\boldsymbol{Q}_k(q^{-1})$, $\boldsymbol{L}_{k*}(q)$. Equations (39) and (40) are verified in the same way. □

Remark. Note that since $\boldsymbol{D}$ is monic, the leading coefficient matrix of the right hand side of (37) and of (40) will cancel. No positive powers of q are be present in $\boldsymbol{Q}_{k+1}(q^{-1})$ and no negative powers of q will be present in $\boldsymbol{L}_{k-1*}(q)$. □

Corollary 3. For white gradient noise η_t with covariance matrix $\mathbf{R}_\eta$, the Diophantine equation (26) has a closed-form solution for $k=1$, given by

$$\boldsymbol{Q}_1(q^{-1}) \;=\; q(\boldsymbol{\beta}(q^{-1}) - \boldsymbol{D}(q^{-1})\beta_0) \quad (42)$$

$$\boldsymbol{L}_{1*}(q) \;=\; \beta_0\boldsymbol{\beta}_*(q) - \mathbf{R}^{-1}\mathbf{R}_\eta\mathbf{R}^{-1}\boldsymbol{D}_*(q) \quad (43)$$

where β_0 is the leading coefficient matrix of $\boldsymbol{\beta}(q^{-1})$.

Proof. With $\boldsymbol{M}\boldsymbol{M}_* = \mathbf{R}_\eta$ and $N=1$, (25) becomes

$$\boldsymbol{\beta}\boldsymbol{\beta}_* = \boldsymbol{C}\mathbf{R}_e\boldsymbol{C}_* + \boldsymbol{D}\mathbf{R}^{-1}\mathbf{R}_\eta\mathbf{R}^{-1}\boldsymbol{D}_*$$

and with $k=1$ and $N=1$, the equation (26) becomes

$$q\boldsymbol{C}\mathbf{R}_e\boldsymbol{C}_* = \boldsymbol{Q}_1\boldsymbol{\beta}_* + q\boldsymbol{D}\boldsymbol{L}_{1*} \;. \quad (44)$$

By substituting the expressions (42) and (43) into the right hand side of (44), the result is verified □

6 Conclusions

We have outlined the design of a class of adaptation laws which are generalizations of LMS. Compared to Kalman tracking of linear regression parameters, a main advantage with the proposed class of algoritms is their lower computational complexity. Another advantage is that it becomes more straightforward to design fixed-lag smoothing estimators. A disadvantage is that our Wiener design is a steady-state solution, which could lead to worse transient properties than for a Kalman estimator. Improved transients could be obtained by using an increased adaptation gain at the beginning of the time series, which then decays to the steady-state value.

The hypermodel (21) is in practice never exactly known, but it may be known to belong to a set of possible models. A robust design which minimizes the *average* performance can then be obtained by averaging the hypermodels in the frequency domain and performing the design for this averaged model. See [22] for general robust design methods, [11, 20] for uncertain hypermodels and [20] for a specialization to fading mobile radio channels parametrized by uncertain Doppler frequencies.

References

[1] A. Ahlén, L. Lindbom and M. Sternad, "Tracking of time-varying systems, Part II: Analysis of stability and performance of adaptation algorithms with time-invariant gains". Report, Signals and Systems group, Uppsala University. Submitted.
www.signal.uu.se/Publications/abstracts/r002.html

[2] A. Ahlén, L. Lindbom and M. Sternad "Channel tracking with WLMS algorithms: High performance at LMS computational load," *IEEE Vehicular Technology Conf. VTC2000-Spring*, Tokyo, Japan, May 15-18, 2000.
www.signal.uu.se/Publications/abstracts/r003.html

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

[4] G. Davidov, A. Shavit and Y. Koren, "Estimation of dynamical-varying parameters by the internal model principle," *IEEE Trans. on Automatic Control*, **vol. 37**, pp. 498-503, 1992.

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

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

[7] M.J. Grimble, "Two-degrees of freedom feedback and feedforward optimal control of multivariable stochastic systems," *Automatica*, **vol. 24**, pp. 809-817, 1988.

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

[9] K.J. Hunt, M. Šebek and V. Kučera, "H2 control problems," in K Hunt, ed: Polynomial Methods in Optimal Control and Filtering, Control Engineering Series, Peter Peregrinus, London, 1993.

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

[11] 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

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

[13] 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.

[14] 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.

[15] M. Niedźwiecki, *Identification of Time-varying Systems.* Wiley, Chichester, 2000.

[16] 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.

[17] M. Sternad and A. Ahlén, "A Novel Derivation Methodology for Polynomial-LQ Controller Design." *IEEE Trans. on Automatic Control*, **vol. 38**, pp 116-121, 1993.

[18] M. Sternad, L Lindbom, and A. Ahlén "Tracking of time-varying systems. Part I: Wiener design of algorithms with constant gains." Report, Signals and Systems group, Uppsala University. Submitted.
www.signal.uu.se/Publications/abstracts/r001.html

[19] M. Sternad, L. Lindbom and A.Ahlén "Tracking of time-varying mobile radio channels with WLMS Algorithms: A case study on D-AMPS 1900 channels," *IEEE Vehicular Technology Conf. VTC2000-Spring*, Tokyo, Japan, May 15-18, 2000.
www.signal.uu.se/Publications/abstracts/r004.html

[20] 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 29-31, 2001.

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

[22] K. Öhrn, A. Ahlén and M. Sternad, "A probabilistic approach to multivariable robust filtering and open-loop control." *IEEE Trans. on Automatic Control*, **vol. 40**, pp. 405-417, 1995.