Wiener Design of Adaptation Algorithms with Time-invariant Gains

Mikael Sternad*, Lars Lindbom and Anders Ahlén*

April 7, 2002

Abstract- **A design method is presented that extends LMS adaptation of time-varying parameters by including general linear time-invariant filters that operate on the instantaneous gradient vector. The aim is to track time-varying parameters of linear regression models, in situations where the regressors are stationary or have slowly time-varying properties. The adaptation law is optimized with respect to the steady-state parameter error covariance matrix for time-variations modeled as vector-ARIMA processes. The design method systematically uses prior information about time-varying parameters to provide filtering, prediction or fixed lag smoothing estimates for arbitrary lags. The method is based on a transformation of 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 has to be considered for fast variations. In the latter case, the filter design is performed iteratively. The general form of the solution at each iteration is obtained by a bilateral Diophantine polynomial matrix equation and a spectral factorization. For white gradient noise, the Diophantine equation has a closed form solution. Further structural constraints result in very simple design equations. Under certain model assumptions, the Wiener designed adaptation laws reduce to LMS adaptation. Compared to Kalman estimators, the channel tracking performance becomes nearly the same in mobile radio applications, while the complexity is in general much lower.**

Keywords: Tracking, adaptive estimation, adaptive filtering, channel modeling, Least mean squares.

EDICS: 2-ADAPT

I. Introduction

When tracking time-varying parameters of linear regression models, LMS is one of the simplest adaptation laws, while Kalman algorithms are the most powerful linear estimators. A third, intermediate, alternative is proposed here: The integration of the instantaneous gradient vector used in LMS is substituted by general linear time-invariant filtering. Well-tuned filters then provide estimates with an appropriate amount of coupling and inertia, resulting in high performance at low computational complexity.

We will here present a novel approach to the design of such adaptation laws that is based on a polynomial matrix approach to Wiener filtering [1].

The difficult problem of accurately tracking time-varying radio channels in IS-136 TDMA cellular systems was an original motivating application for this work. For such systems, LMS and windowed RLS adaptation provide inadequate performance while the use of Kalman algorithms has so far been precluded, due to their computational complexity. An early version of our present approach ("KLMS") was reported in [24] and it has subsequently been used for rapidly time-varying IS-136 1900MHz channels. A thorough case study on this application is found in [27]. See also [17, 29].

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

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

where all terms may be complex-valued. The known regression matrix sequence $\{\varphi_t^*\}$, of dimension $n_y \times n_h$, is stationary with zero mean and a covariance matrix

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

that is assumed to be nonsingular. The elements of φ_t^* may consist of filtered known signals, for example in Laguerre models of IIR systems or in filtered-X LMS-like adaptation of inverse filters. The noise v_t is assumed to be uncorrelated with φ_t^* and to be stationary.

Our aim is to estimate the time-varying $n_h \times 1$ parameter vector h_t in an environment with stationary (or slowly time-varying) statistics of the regressors and the noise. Linear regressions (1) with delayed measurements y_t used as regressors are here excluded; The properties of φ_t^* could then become highly nonstationary when h_t is rapidly time-varying.

Without further assumptions, we cannot for $n_y < n_h$ determine the sequence of parameters 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. 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, models that represent assumptions on the relationship between h_t and h_τ for $\tau \neq t$ must be introduced.

Models of time-varying parameters, sometimes denoted *hypermodels* [5, 6], may be deterministic [7, 8,

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

13, 23, 30, 31] or stochastic [12, 21]. A large variety of parameter dynamics can be described by linear time-invariant stochastic hypermodels

$$h_t = \boldsymbol{\mathcal{H}}(q^{-1})e_t \quad , \tag{3}$$

where e_t is white noise with covariance matrix $\mathbf{R}_e$, $\boldsymbol{\mathcal{H}}(q^{-1})$ is a matrix of stable or marginally stable transfer operators of dimension $n_h \times n_h$ and q^{-1} is the backward shift operator ($q^{-1}x_t = x_{t-1}$). Such models are used here and represent either prior information or design assumptions.

Define the tracking error vector

$$\tilde{h}_{t+k|t} \;\stackrel{\Delta}{=}\; h_{t+k} - \hat{h}_{t+k|t} \tag{4}$$

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 ($k < 0$). Tracking performance will be measured by the steady state covariance matrix

$$\mathbf{P}_k \stackrel{\Delta}{=} \lim_{t\to\infty} \mathbf{P}_{t+k|t} = \lim_{t\to\infty} \mathrm{E}\, \tilde{h}_{t+k|t}\tilde{h}^*_{t+k|t} \quad , \tag{5}$$

where the expectation is taken with respect to e_t in (3) and v_t in (1).

Among all adaptation laws which perform linear operations on y_t, the Kalman filter will minimize $\mathbf{P}_{t+k|t}$ if $\varphi_t^*, \boldsymbol{\mathcal{H}}(q^{-1})$ and $\mathbf{R}_e$ in (1),(3) are known [4, 18]. Kalman estimators are based on a state-space model of (3), with (1) as the measurement equation. Kalman-based adaptive filters discussed in the literature are mostly based on first order autoregressive parameter models [14, 15, 10, 36], but can, of course, be designed for more complicated linear models.

The computational complexity of Kalman estimators is relatively high, due to the on-line Riccati equation update. This may preclude their use in high-speed applications or when the number of parameters is high. A commonly used alternative of much lower complexity is the LMS algorithm [44, 45]

$$\begin{aligned} \varepsilon_t &= y_t - \varphi_t^* \hat{h}_{t|t-1} && (6)\\ \hat{h}_{t+1|t} &= \hat{h}_{t|t-1} + \mu\varphi_t\varepsilon_t && (7) \end{aligned}$$

or, in shift operator notation,

$$\hat{h}_{t+1|t} = \frac{\mu}{1-q^{-1}}\mathbf{I}\ (\varphi_t\varepsilon_t) \tag{8}$$

where μ is the scalar gain and ε_t is the prediction error.

Our aim will be to propose design rules for a class of algorithms that require much lower computational complexity as compared to Kalman tracking, while attaining close to the same performance. They utilize stochastic hypermodels (3), and deliver filtering, prediction and smoothing estimates for arbitrary values of k.

The class of estimators generalize LMS by substituting a time-invariant matrix of linear causal transfer operators for the LMS operator $[\mu/(1-q^{-1})]\,\mathbf{I}$ in (8). An IIR filter matrix $\boldsymbol{\mathcal{M}}_k(q^{-1})$ that operates on $\varphi_t\varepsilon_t$

$$\begin{aligned} \varepsilon_t &= y_t - \varphi_t^* \hat{h}_{t|t-1} && (9)\\ \hat{h}_{t+k|t} &= \boldsymbol{\mathcal{M}}_k(q^{-1})(\varphi_t\varepsilon_t) && (10) \end{aligned}$$

is to be designed based on (1),(3), so that (5) is minimized, under various structural constraints and assumptions. An aim is to clarify how design assumptions are reflected in the complexity of the resulting algorithm.

Related work has been presented by Benveniste and co-workers [5, 6] who used state-space models to perform an interesting analysis of multi-step adaptation laws with constant gains. However, that work, as well as most other analysis of LMS, RLS and Kalman-based tracking, has focused exclusively on cases with slowly time-varying dynamics, since only then can tools of weak convergence theory and various methods of averaging [22, 28] be used.

A design methodology that can handle also practically important classes of problems with fast parameter variations is derived here, by formulating the tracking problem in a novel way. In Section II, the adaptation law is expressed as a stable Wiener filter which is applied on a signal that can be constructed from y_t. In Section III, the conditions for open loop Wiener design are specified, and the filter is thereafter optimized using the polynomial approach outlined in [1, 2]. Section IV introduces constraints that lead to simpler algorithms, such as the Wiener LMS (WLMS) structure, which was introduced and derived from a constained MSE optimization problem in [26]. Such schemes in general have much lower computational complexity than the more general case (10), but may pay a price for this in performance. Section V summarizes the proposed iterative design process and illustrates it with examples.

Remarks on the notation. A superscript asterisk represents a conjugate transpose. For polynomial matrices $\boldsymbol{P}(q^{-1})$ and rational matrices $\boldsymbol{\mathcal{R}}(q^{-1})$ [19], conjugate matrices $\boldsymbol{P}_*(q)$ or $\boldsymbol{\mathcal{R}}_*(q)$ (denoted by subscript asterisks) are obtained by conjugating coefficients, transposing and substituting the forward shift operator q for q^{-1}. To simplify notation, the arguments q or q^{-1} are often omitted in Section III. Scalar polynomials $P(q^{-1})$ are denoted by non-boldface capitals.

The degree of a polynomial matrix is the highest degree of any polynomial element.

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 $|z| \leq 1$ □

II. The Transformed Problem

A. The Fictitious Measurement

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

$$\varepsilon_t \;=\; \varphi_t^*(h_t - \hat{h}_{t|t-1}) + v_t$$

$$\varphi_t \varepsilon_t \;=\; \varphi_t \varphi_t^* \tilde{h}_{t|t-1} + \varphi_t v_t \quad . \tag{11}$$

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

$$\mathbf{Z}_t \;=\; \varphi_t \varphi_t^* - \mathbf{R} \tag{12}$$
$$\eta_t \;=\; \mathbf{Z}_t \tilde{h}_{t|t-1} + \varphi_t v_t \tag{13}$$
$$f_t \;=\; \mathbf{R} h_t + \eta_t \quad , \tag{14}$$

the vector $\varphi_t \varepsilon_t$ in (11) is now reformulated as

$$\begin{aligned} \varphi_t \varepsilon_t &= \mathbf{R} h_t - \mathbf{R}\hat{h}_{t|t-1} + \mathbf{Z}_t \tilde{h}_{t|t-1} + \varphi_t v_t \\ &= f_t - \mathbf{R}\hat{h}_{t|t-1} \quad . \end{aligned} \tag{15}$$

The signal f_t defined in (14) can be regarded as a fictitious measurement, with $\mathbf{R}h_t$ and η_t being the signal and the noise, respectively. In the sequel, the noise terms η_t and $\mathbf{Z}_t \tilde{h}_{t|t-1}$ will be referred to as the *gradient noise* and the *feedback noise*, respectively. The matrix $\mathbf{Z}_t$, of dimension $n_h \times n_h$, has zero mean by definition. This matrix was introduced by Gardner [11] and was referred to as the *autocorrelation matrix noise.*

B. Tracking Regarded as Time-invariant Filtering

By (10),

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

Substitution of this expression into (15) and use of the resulting expression for $\varphi_t \varepsilon_t$ in (10) gives

$$\hat{h}_{t+k|t} = \boldsymbol{\mathcal{M}}_k(q^{-1})(\mathbf{I} + q^{-1}\mathbf{R}\boldsymbol{\mathcal{M}}_1(q^{-1}))^{-1} f_t \;\stackrel{\Delta}{=}\; \boldsymbol{\mathcal{L}}_k(q^{-1}) f_t \, . \tag{16}$$

We may therefore design a time-invariant stable rational matrix $\boldsymbol{\mathcal{L}}_k(q^{-1})$ that operates on f_t to estimate h_{t+k}

$$f_t \;=\; \mathbf{R}\hat{h}_{t|t-1} + \varphi_t \varepsilon_t \;=\; \mathbf{R} h_t + \eta_t \tag{17}$$
$$\hat{h}_{t+k|t} \;=\; \boldsymbol{\mathcal{L}}_k(q^{-1}) f_t \quad . \tag{18}$$

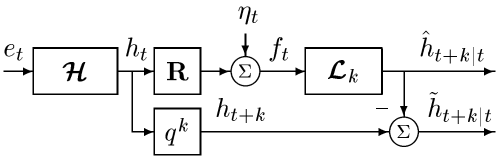

Figure 1: The filter design problem. The vector h_{t+k} is to be estimated from measurements f_t, such that the steady state tracking error covariance matrix is minimized.

This causal filter $\boldsymbol{\mathcal{L}}_k(q^{-1})$ will be referred to as the *learning filter.* For a given $\boldsymbol{\mathcal{L}}_1(q^{-1})$, $k=1$ in (16) gives

$$\boldsymbol{\mathcal{M}}_1(q^{-1}) = (\mathbf{I} - q^{-1}\boldsymbol{\mathcal{L}}_1(q^{-1})\mathbf{R})^{-1} \boldsymbol{\mathcal{L}}_1(q^{-1}) \quad . \tag{19}$$

With $\boldsymbol{\mathcal{L}}_k(q^{-1})$ and $\boldsymbol{\mathcal{M}}_1(q^{-1})$, $\boldsymbol{\mathcal{M}}_k(q^{-1})$ can then be obtained from (16).

As seen by (13)-(14), three terms influence the tracking performance via f_t: 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 $\mathbf{Z}_t \tilde{h}_{t|t-1}$.

The estimation error follows from (14) and (18) as

$$\tilde{h}_{t+k|t} = (q^k \mathbf{I} - \boldsymbol{\mathcal{L}}_k(q^{-1})\mathbf{R}) h_t - \boldsymbol{\mathcal{L}}_k(q^{-1}) \eta_t \quad , \tag{20}$$

where $q^k h_t = h_{t+k}$. The first right-hand term is for $k=0$ usually called the *lag error.*

An open-loop Wiener design, yielding $\boldsymbol{\mathcal{L}}_k(q^{-1})$, can now be performed. Such a design would be based on the assumption that the innovations of η_t are uncorrelated with $\hat{h}_{t-i|t-i-1}, i \geq 0$. (Possible higher-order statistical dependencies due to the outer feedback loop in Figure 2 do not affect an MSE-optimal linear design.) If the innovations of η_t are furthermore uncorrelated with the signal h_{t-i}, such an open-loop design is simplified. Although these conditions are not always fulfilled they hold exactly or approximately in many practically important circumstances, since the multiplication by $\mathbf{Z}_t$ in (13) acts as a scrambler, see e.g. [3, 26, 27].

Uncorrelatedness holds when the time-variations are slow or when consecutive regressor matrices are independent. It also turns out to be a good approximation even for fast parameter variations in FIR systems with white regressors and it holds exactly in the second order case for regressors with constant modulus [3]. In problems with colored regressors and fast variations, the scrambling by $\mathbf{Z}_t$ may become less effective. The correlation properties of the innovation sequence of η_t should then be investigated after the design has been performed, in the way indicated in Section V by Figure 6.

Uncorrelatedness of the innovations of η_t with h_{t-i} and $\hat{h}_{t-i|t-i-1}$ will in Section III below be stated as a design assumption, under which $\boldsymbol{\mathcal{L}}_k(q^{-1})$ will be optimized by just treating η_t in (20) as an additive noise, with known properties.

C. Properties of the Gradient Noise

The feedback noise $\mathbf{Z}_t \tilde{h}_{t|t-1}$ will not be *independent* of $\hat{h}_{t-i|t-i-1}$, due to the outer feedback loop in Figure 2. This feedback could cause instability. Therefore, the gain of $\boldsymbol{\mathcal{L}}_1(q^{-1})$ cannot be allowed to be too large.

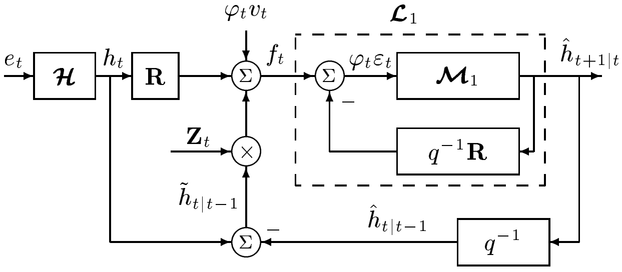

Figure 2: The feedback loop via the feedback noise $\mathbf{Z}_t \tilde{h}_{t|t-1}$ may significantly affect the variance of the fictitious measurement f_t, and causes dependence with $\tilde{h}_{t|t-1}$.

Since the properties of η_t depend on $\boldsymbol{\mathcal{L}}_1(q^{-1})$, a tracking design will in general require a few iterations, as outlined in Section V. After each iteration, we may have to estimate the properties of η_t by simulation. However, in [3], three important scenarios are discussed in which an analytical performance evaluation is possible by assuming v_t, φ_t and e_t to be mutually independent:

1. *Slowly varying parameters (vanishing feedback noise).* We then have a true open-loop situation, in which the outer feedback loop in Figure 2 can be neglected. This will indeed be the case in many applications of adaptive filtering. When a Wiener design is performed in a situation with stable $\mathcal{H}(q^{-1})$ and bounded regressors, it can be shown that the impact of the feedback noise $\mathbf{Z}_t\tilde{h}_{t|t-1}$ on the tracking MSE will vanish when the power of e_t becomes sufficiently small relative to the power of v_t. The feedback noise becomes negligible either when the parameters h_t vary slowly, or when the noise level is high[1]. Then, $\eta_t \approx \varphi_t v_t$. In [3], we propose negligible feedback noise as the defining characteristic of the concept of "slow variations".

2. *Independent consecutive regression matrices.* If φ_t^* and φ_s^* are independent for $t \neq s$, then $\mathbf{Z}_t\tilde{h}_{t|t-1}$ is uncorrelated with old estimates. It is white with zero mean and its covariance can be derived exactly.

3. *FIR models with white zero mean regressors.* The performance can then be predicted with good accuracy from theoretical expressions valid for arbitrarily fast variations of h_t.

III. The Wiener Solution

The transfer operator $\mathcal{L}_k(q^{-1})$ will now be adjusted to minimize (5), when $\mathcal{H}(q^{-1})$ is known and the properties of η_t are assumed given. Minimization implies that any alternative estimator provides a covariance matrix, say $\mathbf{P}_k^o$, for which $\mathbf{P}_k^o - \mathbf{P}_k$ will be positive semidefinite. A minimization of $\mathbf{P}_k$ will also minimize its trace, the sum of componentwise tracking MSEs, or the mean square deviation (MSD).

A. Main Result

The learning filter $\mathcal{L}_k(q^{-1})$ is designed under the constraints of stability and causality and 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. It has a known rational spectral density $\phi_\eta(e^{\jmath\omega})$ modeled by a vector ARMA process

$$\eta_t = \frac{1}{N(q^{-1})}\boldsymbol{M}(q^{-1})\nu_t \quad ; \quad \mathrm{E}\,\nu_t\nu_t^* = \mathbf{I} \quad , \qquad (21)$$

where $\boldsymbol{M}$ is an $n_h \times n_h$ polynomial matrix of degree n_M, N is a stable polynomial of degree n_N □

Assumption A3: The innovation sequence ν_t of the gradient noise is uncorrelated with h_{t-i} and with $\hat{h}_{t-i|t-i-1}, i \geq 0$ □

Assumption A4. The linear regression coefficients are described by a stochastic process

$$h_t = \mathcal{H}(q^{-1})e_t = \boldsymbol{D}(q^{-1})^{-1}\boldsymbol{C}(q^{-1})e_t \quad , \qquad (22)$$

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

$$\begin{aligned} \boldsymbol{D}(q^{-1}) &= D_u(q^{-1})\boldsymbol{D}_s(q^{-1}) \qquad (23)\\ &= \mathbf{I} + \mathbf{D}_1 q^{-1} + \ldots \mathbf{D}_{n_D} q^{-n_D} \\ \boldsymbol{C}(q^{-1}) &= \mathbf{I} + \mathbf{C}_1 q^{-1} + \ldots \mathbf{C}_{n_C} q^{-n_C} \end{aligned}$$

are time-invariant[2]. Above, $\boldsymbol{C}(q^{-1})$ is assumed stable, with full rank on $|z| = 1$. $D_u(q^{-1})$ is a polynomial with zeros on the unit circle 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 marginally stable dynamics $D_u(q^{-1})$ must then affect all the elements of h_t.

We are now ready to state the following main result.

Theorem 1: The optimal learning filter. Under Assumptions A1-A4, the stable and causal learning filter in (18) minimizing (5) is given by

$$\mathcal{L}_k^{opt} = \boldsymbol{D}_s^{-1}\boldsymbol{Q}_k\beta^{-1}N\boldsymbol{D}_s\mathbf{R}^{-1} \quad , \qquad (24)$$

where the polynomial matrix

$$\beta(q^{-1}) = \beta_0 + \beta_1 q^{-1} + \ldots + \beta_{n_\beta} q^{-n\beta}$$

of dimension $n_h \times n_h$ and degree $n_\beta = \max(n_C + n_N, n_D + n_M)$ is the stable spectral factor obtained from

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

The unique solution to the bilateral Diophantine equation

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

provides polynomial matrices

$$\begin{aligned} \boldsymbol{Q}_k(q^{-1}) &\triangleq \mathbf{Q}_0^k + \mathbf{Q}_1^k q^{-1} + \ldots + \mathbf{Q}_{n_Q}^k q^{-n_Q} \\ \boldsymbol{L}_{k*}(q) &\triangleq \mathbf{L}_0^{k*} + \mathbf{L}_1^{k*} q + \ldots + \mathbf{L}_{n_L}^{k*} q^{n_L} \end{aligned}$$

of dimension $n_h \times n_h$, with generic degrees

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

respectively. The estimation error $\tilde{h}_{t+k|t}$ will be stationary, with finite covariance matrix and zero mean □

Proof: See Appendix A.

[1] Another case when $\mathbf{Z}_t\tilde{h}_{t|t-1}$ vanishes completely is when φ_t^* is scalar with constant modulus. Then, $\mathbf{Z}_t = 0$. This is the case e.g. when tracking flat fading channels in mobile radio systems using PSK symbol alphabets.

[2] While we assume $\mathcal{H}(q^{-1})$ to be time-invariant, it can in practice be allowed to be slowly time-varying, as long as the variations are much slower than those of h_t.

B. Remarks and Generalizations

Solvability of the equations. For a discussion of multivariable Wiener filtering problems solved by Diophantine equations and spectral factorizations, see [1, 2, 35, 38]. The Diophantine equation (26) is guaranteed to be solvable and it corresponds to a linear system of equations, with equal number of unknowns and equations.

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, (25) has full rank on $|z|=1$, resulting in a stable spectral factor $\boldsymbol{\beta}$ with a leading matrix $\boldsymbol{\beta}_0$ of full rank. Thus, $\boldsymbol{\beta}^{-1}$ in (24) is causal and stable.

Algorithms for solving polynomial matrix spectral factorizations and bilateral Diophantine equations are presented in [34, 20].

The learning filters have real-valued coefficients when $\boldsymbol{C}, \boldsymbol{D}, \mathbf{R}_e, \boldsymbol{M}, N$ and $\mathbf{R}$ have real-valued coefficients. Optimal learning filters (24) for different lags k differ only in $\boldsymbol{Q}_k$, since $\boldsymbol{\beta}$ is unaffected by k.

The limiting cases of high and low gradient noise. If the gradient noise has a spectral peak at $\omega=\omega_1$ described by a zero of N close to the unit circle, all elements of $\boldsymbol{\mathcal{L}}_k^{opt}$ will have a notch at ω_1 since $N(e^{-j\omega_1})\approx 0$.

When the gradient noise is negligible, $\boldsymbol{M}\approx 0$. Equations (25), (26) are then for $k\leq 0$ solved by

$$\boldsymbol{\beta}\approx N\boldsymbol{C}\mathbf{R}_e^{1/2} \quad , \quad \boldsymbol{Q}_k\approx q^k\boldsymbol{C}\mathbf{R}_e^{1/2} \quad , \quad \boldsymbol{L}_{k*}\approx 0 \ .$$

The lag error in (20) then vanishes since $\boldsymbol{\mathcal{L}}_k^{opt}\approx q^k\mathbf{R}^{-1}$ and this estimator attains $\mathbf{P}_k\approx 0$ for $k\leq 0$.

Recursive computation of estimators for different smoothing lags. The solution for $k=1$ will always be required, since $\hat{h}_{t|t-1}$ appears in (17). When several estimation horizons are of interest, we need to solve equation (26) for one value of k only. It is shown in Appendix B that solutions for any k can then be obtained recursively from one of the solutions.

Robust design. The hypermodel (22) 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 of (5) can then be obtained by averaging the hypermodels in the frequency domain and performing the design for this averaged model. See [25] for details, [46] for general methods and [39] for a specialization to fading mobile radio channels parametrized by uncertain Doppler frequencies.

C. Realizations and Interpretations

The estimator defined by (17), (18) and (24) can be realized as it stands. We can however give one internal signal a special meaning. From (17) (21) and (22), the spectral density of f_t is under A2-A4 given by

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

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

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

where the innovation sequence ϵ_t is white, with zero mean and unit covariance matrix. When $\boldsymbol{D}$ has zeros on the unit circle, (29) corresponds to a generalized innovation model [37]. By defining the signal

$$\bar{\epsilon}_t \;\stackrel{\Delta}{=}\; \frac{1}{D_u(q^{-1})}\mathbf{I}\,\epsilon_t \;=\; \boldsymbol{\beta}^{-1}N\boldsymbol{D}_s\mathbf{R}^{-1}f_t \qquad (30)$$

we obtain the realization of (17),(18),(24) (Figure 3):

$$\begin{aligned}\varepsilon_t &= y_t-\varphi_t^*\hat{h}_{t|t-1} &(31)\\ s_t &= \mathbf{R}^{-1}f_t=\hat{h}_{t|t-1}+\mathbf{R}^{-1}\varphi_t\varepsilon_t &(32)\\ \bar{\epsilon}_t &= \boldsymbol{\beta}^{-1}N\boldsymbol{D}_s s_t &(33)\\ \hat{h}_{t+k|t} &= \boldsymbol{D}_s^{-1}\boldsymbol{Q}_k\bar{\epsilon}_t \ . &(34)\end{aligned}$$

The realization (31)-(34) has good numerical properties and all involved filters are internally stable.

The optimal tracking filter $\boldsymbol{\mathcal{M}}_k$ can be calculated from (16), (19) once $\boldsymbol{\mathcal{L}}_k$ and $\boldsymbol{\mathcal{L}}_1$ are derived. A compact expression for $\boldsymbol{\mathcal{M}}_k$ is given by the following corollary.

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

$$\hat{h}_{t+k|t}=\boldsymbol{\mathcal{M}}_k(\varphi_t\varepsilon_t) \;=\; \boldsymbol{D}_s^{-1}\boldsymbol{Q}_k\boldsymbol{\mathcal{R}}(\mathbf{R}^{-1}\varphi_t\varepsilon_t) \qquad (35)$$

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

$$\boldsymbol{\mathcal{R}} \;\stackrel{\Delta}{=}\; [\boldsymbol{\beta}-q^{-1}N\boldsymbol{Q}_1]^{-1}N\boldsymbol{D}_s \qquad (36)$$

□

Proof. Multiply both sides of (33) from the left by $\boldsymbol{\beta}$ and then substitute the expression for $q^{-1}\hat{h}_{t+1|t}$, obtained from (34) with $k=1$, into (32),(33). We obtain

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

Thus,

$$\bar{\epsilon}_t=\boldsymbol{\mathcal{R}}(\mathbf{R}^{-1}\varphi_t\varepsilon_t) \ . \qquad (37)$$

The use of this expression in (34) gives (35)-(36) □

Note that $\mathbf{R}^{-1}$ will always be a right factor of the optimal $\boldsymbol{\mathcal{M}}_k$. Our estimator can thus be seen as a generalization of the LMS-Newton adaptation law [45].

While $\boldsymbol{\mathcal{L}}_k$ must be stable, $\boldsymbol{\mathcal{M}}_k$ need not be stable, since it is a block in a feedback loop. In fact, any marginally stable model denominator D_u^{-1} is present in all elements of $\boldsymbol{\mathcal{M}}_k$. See (A.7) in Appendix A.

D. White Gradient Noise

By assuming the gradient noise η_t to be white, with zero mean and with a known covariance matrix $\mathbf{R}_\eta$, both the design and the implementation is simplified. Since

$$\eta_t=\mathbf{Z}_t\tilde{h}_{t|t-1}+\varphi_t v_t \ \ ,$$

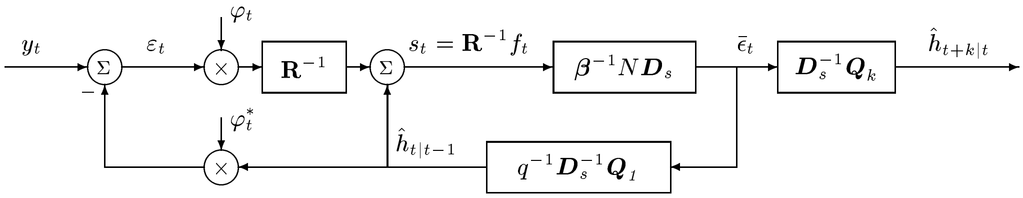

Figure 3: The optimized tracking algorithm with time-invariant gains, realized as in (31)-(34)

the feedback noise is in such cases assumed to be white or negligible and the noise vector $\varphi_t v_t$ is white. The last assumption is true whenever the noise v_t is white and independent of the regressors. Negligible feedback noise is characteristic of situations with slow variations. The feedback noise is white under the restrictive assumption of independent regression matrices but it is otherwise not in general white for colored regressors. It will be white for second order FIR channels with white regressor elements with constant modulus [3], a case of practical significance in mobile radio channel tracking as discussed in [27] and Section V.

For white gradient noise, there exists a closed-form solution to the Diophantine equation (26). The solution for one-step prediction is presented in the following lemma. The iterations yielding filters for arbitrary lags k are presented in Corollary 2 in Appendix B.

Lemma 1. For white gradient noise η_t with covariance matrix $\mathbf{R}_\eta$, let

$$\mathbf{R}_\gamma \triangleq \mathbf{R}^{-1}\mathbf{R}_\eta\mathbf{R}^{-1} \quad . \tag{38}$$

The solution to the Diophantine equation (26) for $k = 1$ is then given by

$$\begin{aligned} \boldsymbol{Q}_1(q^{-1}) &= q(\boldsymbol{\beta}(q^{-1}) - \boldsymbol{D}(q^{-1})\boldsymbol{\beta}_0) && (39) \\ \boldsymbol{L}_{1*}(q) &= \boldsymbol{\beta}_0\boldsymbol{\beta}_*(q) - \mathbf{R}_\gamma \boldsymbol{D}_*(q) && (40) \end{aligned}$$

where $\boldsymbol{\beta}_0$ is the leading coefficient matrix of $\boldsymbol{\beta}(q^{-1})$ □

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

$$\boldsymbol{\beta}\boldsymbol{\beta}_* = \boldsymbol{C}\,\mathbf{R}_e\boldsymbol{C}_* + \boldsymbol{D}\,\mathbf{R}_\gamma\boldsymbol{D}_* \tag{41}$$

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

$$q\boldsymbol{C}\,\mathbf{R}_e\boldsymbol{C}_* = \boldsymbol{Q}_1\boldsymbol{\beta}_* + q\boldsymbol{D}\,\boldsymbol{L}_{1*} \quad . \tag{42}$$

By substituting the expressions (39) and (40) into the right hand side of (42), the lemma is verified □

The implementation of the tracker is also simplified since by (39) and $N = 1$ in (36),

$$\begin{aligned} \boldsymbol{\mathcal{R}} &= (\boldsymbol{\beta} - q^{-1}\boldsymbol{Q}_1)^{-1}\boldsymbol{D}_s \\ &= (\boldsymbol{\beta} - (\boldsymbol{\beta} - \boldsymbol{D}\boldsymbol{\beta}_0))^{-1}\boldsymbol{D}_s = \frac{1}{D_u}\boldsymbol{\beta}_0^{-1} && (43) \end{aligned}$$

which simplifies the realizations of $\boldsymbol{\mathcal{M}}_k(q^{-1})$ in (35). By using (43) in (37), the innovation processes are

$$\bar{\epsilon}_t = \frac{1}{D_u}\boldsymbol{\beta}_0^{-1}\mathbf{R}^{-1}\varphi_t\varepsilon_t \;\; ; \;\; \epsilon_t = \boldsymbol{\beta}_0^{-1}\mathbf{R}^{-1}\varphi_t\varepsilon_t \quad . \tag{44}$$

Example 1. First order models and LMS-like algorithms. Assume that (22) is a vector of coupled first order autoregressive or random walk parameters

$$\boldsymbol{D}(q^{-1}) = (1 - aq^{-1})\mathbf{I} \;\; ; \;\; \boldsymbol{C}(q^{-1}) = \mathbf{I}$$

where a is a scalar with $|a| \leq 1$ and $\mathbf{R}_e$ is given. The gradient noise is white, with $\mathbf{R}_\eta$ known. The spectral factorization (41) then becomes

$$(\boldsymbol{\beta}_0+\boldsymbol{\beta}_1q^{-1})(\boldsymbol{\beta}_0^*+\boldsymbol{\beta}_1^*q) = \mathbf{R}_e+(1-aq^{-1})\mathbf{R}_\gamma(1-aq) \tag{45}$$

giving

$$\begin{aligned} \boldsymbol{\beta}_0\boldsymbol{\beta}_0^* + \boldsymbol{\beta}_1\boldsymbol{\beta}_1^* &= \mathbf{R}_e + (1+a^2)\mathbf{R}_\gamma \\ \boldsymbol{\beta}_1\boldsymbol{\beta}_0^* = \boldsymbol{\beta}_0\boldsymbol{\beta}_1^* &= -a\mathbf{R}_\gamma \quad . \end{aligned}$$

These expressions represent a set of coupled second order equations in the elements of $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$.

If we are interested in the one-step predictor, then the solution (39) for $k = 1$ directly gives $\boldsymbol{Q}_1(q^{-1})$ as

$$\boldsymbol{Q}_1(q^{-1}) = q(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1q^{-1} - \boldsymbol{\beta}_0 + a\boldsymbol{\beta}_0q^{-1}) \;=\; \boldsymbol{\beta}_1 + a\boldsymbol{\beta}_0 \quad .$$

From (35) and (43),

$$\hat{h}_{t+1|t} = \boldsymbol{D}_s^{-1}\boldsymbol{Q}_1\frac{1}{D_u}\boldsymbol{\beta}_0^{-1}\mathbf{R}^{-1}\varphi_t\varepsilon_t = \boldsymbol{D}^{-1}\boldsymbol{Q}_1\boldsymbol{\beta}_0^{-1}\mathbf{R}^{-1}\varphi_t\varepsilon_t$$

Thus, we obtain a generalized LMS update equation

$$\hat{h}_{t+1|t} = a\hat{h}_{t|t-1} + (\boldsymbol{\beta}_1 + a\boldsymbol{\beta}_0)\boldsymbol{\beta}_0^{-1}\mathbf{R}^{-1}\varphi_t\varepsilon_t \; . \tag{46}$$

The update (46) is similar to the LMS/Newton law [45] in that the instantaneous gradient is rotated by $\mathbf{R}^{-1}$. It also contains leakage [42, 45] whenever $|a| < 1$. Furthermore, it has a matrix gain instead of the scalar gain in (7) of LMS. The algorithm reduces to LMS when $\mathbf{R} = \sigma_u^2\mathbf{I}$ (white regressors), $a = 1, \mathbf{R}_e = \sigma_e^2\mathbf{I}$ (random walk model with uncorrelated parameters) and if the elements of the gradient noise are uncorrelated and have equal variance $\mathbf{R}_\eta = c\mathbf{I}$. Then, $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ from (45) become diagonal and have all diagonal elements equal, so

$$(\boldsymbol{\beta}_1 + a\boldsymbol{\beta}_0)\boldsymbol{\beta}_0^{-1}\mathbf{R}^{-1} = \mu\mathbf{I}$$

for some scalar μ □

IV. Low Complexity Designs

The design and implementation can be simplified further, at the price of some performance degradation, by placing successively harder restrictions on the hypermodel and on the learning filter.

A. Generalized Wiener LMS

This algorithm is obtained by minimizing the trace of the tracking covariance matrix (the mean square deviation) for possibly colored gradient noise (21) and for hypermodels in common denominator form

$$h_t = \frac{1}{D(q^{-1})}\boldsymbol{C}\,(q^{-1})e_t \quad . \tag{47}$$

The structure of the learning filter is constrained to

$$\hat{h}_{t+k|t} = \boldsymbol{\mathcal{S}}_k(q^{-1})\mathbf{R}^{-1}f_t \quad , \tag{48}$$

where $\boldsymbol{\mathcal{S}}_k(q^{-1})$ is a *diagonal* stable rational matrix. The design equations for this filter are derived and presented in [25]. They consist of n_h separate polynomial spectral factorizations and n_h scalar Diophantine equations.

B. Wiener LMS Algorithms (WLMS)

A further simplification is obtained by minimizing the trace of the tracking covariance matrix for *white* gradient noise with covariance matrix $\mathbf{R}_\eta$, and for diagonal hypermodels with equal elements

$$h_t = \frac{C(q^{-1})}{D(q^{-1})}\mathbf{I}\; e_t \quad . \tag{49}$$

The learning filter is restricted to (48) but with all filters on the diagonal of $\boldsymbol{\mathcal{S}}_k(q^{-1})$ being equal:

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

$$\hat{h}_{t+k|t} = \frac{Q_k(q^{-1})}{\beta(q^{-1})}\mathbf{R}^{-1}f_t = \frac{Q_k(q^{-1})}{\beta(q^{-1})}(\hat{h}_{t|t-1} + \mathbf{R}^{-1}\varphi_t\varepsilon_t) \quad . \tag{50}$$

Compared to Generalized Wiener LMS, this filter structure has reduced ability to handle situations where different elements of h_t have differing dynamical properties, but it is still a useful special case.

The resulting Wiener LMS (WLMS) algorithm can be optimized for a given parameter-drift-to-noise ratio

$$\gamma \;\stackrel{\Delta}{=}\; \mathrm{tr}\,\mathbf{R}_e / \mathrm{tr}\,(\mathbf{R}^{-1}\mathbf{R}_\eta\mathbf{R}^{-1}) \quad . \tag{51}$$

The denominator polynomial $\beta(q^{-1})$ in (50) minimizing $\mathrm{tr}\,\mathbf{P}_k$ is then obtained from the solution to one scalar polynomial spectral factorization

$$\beta\beta_* = \gamma C C_* + D D_* \quad .$$

Then $Q_k(q^{-1})$ is obtained from scalar versions of (26) or Lemma 1, where 1 is substituted for $\mathbf{R}_\gamma$ and γ for $\mathbf{R}_e$.

The WLMS algorithm is derived in [26] and design software can be found on the webpage of [26]. It is applied to the tracking of fading mobile radio channels in [27].

C. Computational Complexity

Let us compare the computational demands of the various discussed algorithms. We assume a complex linear regression (1) with n_y outputs, complex-valued regressors and n_h parameters. These parameters are described by a vector ARIMA model (22) with real-valued coefficients and diagonal denominator $\boldsymbol{D}\,(q^{-1})$, with diagonal polynomial elements of degree n_D.

The comparison concerns one-step predictors $\hat{h}_{t+1|t}$, and is expressed by the number of real multiplications per step. (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.)[3]

We compare the following algorithms.

- A Kalman predictor based on a state space model of (22) with $n_x = n_D n_h$ states, with each element of h_t modeled by n_D states

$$x_{t+1} = \mathbf{F}x_t + \mathbf{G}e_t \tag{52}$$

$$h_t = \mathbf{H}x_t \quad ; \quad y_t = \varphi_t^*\mathbf{H}x_t + v_t \quad .$$

  Here, $\mathbf{F}$ and $\mathbf{H}$ are assumed to be real-valued and block diagonal, with $n_D \times n_D$ blocks in $\mathbf{F}$ and $1 \times n_D$-blocks in $\mathbf{H}$. The noise v_t is assumed white. The block structure is used to reduce the number of operations, but the Riccati update is otherwise performed in the conventional way.

- The Wiener algorithm for $k = 1$, implemented as in Figure 3 or (31)-(34). The gradient noise may be colored and described by (21). The polynomial matrices $\boldsymbol{\beta}(q^{-1})$ and $\boldsymbol{Q}_1(q^{-1})$ have $n_h \times n_h$ complex-valued polynomial elements each. Their degrees are $n_\beta = n_D + n_M$ and $n_Q = n_D - 1$. We also specialize to white gradient noise, using (31), (44) and (34).

- The Wiener LMS algorithm (50) based on (49), with $\beta(q^{-1})$ of degree n_D and $Q_1(q^{-1})$ of degree $n_D - 1$. (Generalized WLMS has similar complexity.)

- The LMS algorithm (6), (7), with a real-valued μ.

Table 1 displays the results. Here, r denotes the complexity of the operation $\mathbf{R}^{-1}\varphi_t$. It is zero for white regressors. It is proportional to n_h when the regressors are moving average processes, so that $\mathbf{R}$ becomes multidiagonal. For scalar FIR models with autoregressively generated inputs, the product can also be updated with a computational complexity proportional to n_h [9]. For an arbitrary covariance matrix, $r = 4n_h^2 - 2n_h$.

[3]The added complexity in cases when $\mathbf{R}^{-1}$ needs to be estimated is not considered here. In the Wiener design, this would add to the complexity proportional to n_h^2, to become similar to RLS. Since n_h is smaller than the number of states whenever $n_D > 1$, the complexity would then still be considerably below that of a Kalman update.

Method	Number of real multiplications	Ex.
Kalman:	$n_h^2\{4n_D^3 + 8n_D^2 n_y\} + n_h\{2n_D^2(n_y+1) + n_D(8n_y^2 + 6n_y) + 4n_y\} + 12n_y^2$	7152
Wiener:	$n_h^2\{4n_D + 2n_M + 2\} + n_h\{4n_D + 2n_N + 8n_y\} + r$	864
Wiener, white η_t	$n_h^2\{2n_D + 2\} + n_h\{2n_D + 8n_y\} + r$	554
WLMS:	$n_h(4n_D + 8n_y) + r$	224
LMS:	$8n_h n_y + 2n_y$	132

Table 1: Computational complexity when tracking the n_h parameters of a complex linear regression model with n_y outputs, n_D states per parameter. Right-hand example for 16 states, $n_h = 8, n_D = 2, n_y = 2, r = 0, n_M = 0$ and $n_N = 2$.

V. Iterative Design

For slow time-variations, the feedback noise is by our definition negligible. so we may perform a one-shot design using $\eta_t = \varphi_t v_t$. When the noise v_t is white, the solution for white gradient noise can be used, with $\mathbf{R}_\eta = \mathbf{R}\sigma_v^2$, if $\mathrm{E}\, v_t v_t^* = \sigma_v^2 \mathbf{I}$ and if φ_t and v_t are independent.

Otherwise, the design can be performed iteratively, by using long simulation runs to estimate the covariance function element matrices

$$\mathbf{R}_{\eta\eta}(j) \;\stackrel{\Delta}{=}\; \mathrm{E}\, \eta_t \eta_{t-j}^* \quad . \tag{53}$$

In a model (21) with $N = 1$, the covariance function of the gradient noise can be represented by

$$\boldsymbol{M}(q^{-1})\boldsymbol{M}_*(q) = \sum_{j=-n_M}^{n_M} \mathbf{R}_{\eta\eta}(j) q^j \quad . \tag{54}$$

Note that only the total covariance function (54), not $\boldsymbol{M}$, is needed in the design equation (25).

We proceed as follows:

1. Perform a one-step predictor design for slow time-variations, i.e. assume $\eta_t = \varphi_t v_t$ in the design $\boldsymbol{\mathcal{L}}_1(q^{-1})$. Verify that the closed loop of Figure 2 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 $\boldsymbol{\mathcal{L}}_1(q^{-1})$.

2. Obtain an estimated gradient noise time series from (17) based on a long simulation of $h_t = \boldsymbol{\mathcal{H}}(q^{-1})e_t$, φ_t and v_t, and on the corresponding estimate $\hat{h}_{t+1|t} = \boldsymbol{\mathcal{L}}_1(q^{-1}) f_t$, as

$$\hat{\eta}_t = f_t - \mathbf{R} h_t = \varphi_t \varepsilon_t - \mathbf{R}(h_t - \hat{h}_{t|t-1}) \quad . \tag{55}$$

Obtain an estimate $\hat{\mathbf{R}}_{\eta\eta}(j)$ of the covariance function (53),(54) by using sample averages over $\hat{\eta}_t$.

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

Repeat steps 2. and 3. until the difference in the estimates $\hat{h}_{t+1|t}$ becomes small for consecutive estimators. Then, construct an estimator for the desired lag k.

It will be possible to find an initial stable solution under mild conditions. If $\boldsymbol{\mathcal{H}}$ is stable, then $\boldsymbol{\mathcal{L}}_1(\omega) \to 0 \;\forall\, \omega$ when the assumed noise power is increased. If $\mathbf{Z}_t$ has bounded elements, then the small gain theorem [43] will imply that the outer closed loop of Figure 2 can be stabilized by assuming a sufficiently high noise power in the design of $\boldsymbol{\mathcal{L}}_1$. See Appendix C.

The covariance function estimate provides additional information. If $\hat{\mathbf{R}}_{\eta\eta}(0)$ does not differ much from the covariance matrix of $\varphi_t\varepsilon_t$, then the time-variations can be regarded as slow, and step 1 above turns out to be sufficient. This will occur when e_t has low power relative to $\varphi_t v_t$, i.e. either when the increments of the parameters h_t are small, or when the noise level is high [3].

If $\mathrm{tr}\,(\hat{\mathbf{R}}_{\eta\eta}(j)) << \mathrm{tr}\,(\hat{\mathbf{R}}_{\eta\eta}(0))$ for all $j \neq 0$, then the gradient noise can be regarded as white so the design of Section III.D can be used, with $\mathbf{R}_\eta = \hat{\mathbf{R}}_{\eta\eta}(0)$.

Iterative tuning becomes much simpler for the WLMS algorithm than in the general case. We then have to tune only one scalar parameter, the noise ratio γ in (51). This parameter could alternatively be used as an on-line tuning knob, to provide an appropriate balance between tracking ability and noise sensitivity.

It should be emphasized that the design methodology assumes a good hypermodel. With incorrect models, there is no reason to believe that the iterations will minimize the tracking MSE. However, robustification according to [25, 39, 46] would alleviate the effect of incorrect models.

Example 2. Iterative design and a comparison to Kalman and LMS tracking. Consider the uplink of a TDMA-based mobile cellular communication system in which two mobile users transmit at the same frequency in the same time slot [40, 41]. One of the users could represent a strong out-of-cell co-channel interferer. A receiver with two diversity branches (multiple antennas or polarization diversity branches) detects both users u_t^1 and u_t^2 simultaneously. We model the situation by

$$\begin{pmatrix} y_t^1 \\ y_t^2 \end{pmatrix} = \begin{pmatrix} B_t^{11}(q^{-1}) & B_t^{12}(q^{-1}) \\ B_t^{21}(q^{-1}) & B_t^{22}(q^{-1}) \end{pmatrix} \begin{pmatrix} u_t^1 \\ u_t^2 \end{pmatrix} + \begin{pmatrix} v_t^1 \\ v_t^2 \end{pmatrix} , \tag{56}$$

where y_t^i is the sampled baseband signal at receiver i. In IS-136 systems [33], delay spreads of up to one symbol can be encountered. To illustrate the tracking performance, we here restrict the discussion to symbol-spaced two-tap channels, with taps of equal variance. Thus,

$$B_t^{ij}(q^{-1}) = b_{0,t}^{ij} + b_{1,t}^{ij} q^{-1} \quad . \tag{57}$$

The model (56),(57) can be expressed in the linear regression form (1) where

$$\varphi_t^* = \begin{pmatrix} u_t^1 & u_{t-1}^1 & u_t^2 & u_{t-1}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & u_t^1 & u_{t-1}^1 & u_t^2 & u_{t-1}^2 \end{pmatrix}$$

and

$$h_t = (\; b_{0,t}^{11} \;\; b_{1,t}^{11} \;\; b_{0,t}^{12} \;\; b_{1,t}^{12} \;\; b_{0,t}^{21} \;\; b_{1,t}^{21} \;\; b_{0,t}^{22} \;\; b_{1,t}^{22} \;)^T \; . \quad (58)$$

Here, the symbols $u_{t-\tau}^j$ are assumed known. (In reality the unknown parts of the received symbol sequences have to be estimated.) Assume $\{u_t^j\}$ to be white complex-valued QPSK symbols with $\mathbf{R} = \mathbf{I}_8$, while the noise $v_t = [v_t^1 \; v_t^2]^T$ is white with variance $\sigma_v^2 \mathbf{I}_2$.

The messages are transmitted from moving mobile terminals, so the channel taps $b_{\ell,t}^{ij}$ will be time-varying (fading). The second order statistics of fading radio channels can be approximated by autoregressive models, here assumed to be of second order,

$$\frac{1}{1 - 2\rho\cos(\omega_{\mathrm{D},j}T/\sqrt{2})q^{-1} + \rho^2 q^{-2}} = \frac{1}{D(q^{-1}, \omega_{\mathrm{D},j}T)} \; . \quad (59)$$

According to the discussion of [27], (59) provides a reasonable approximation to Rayleigh fading statistics [16], if

$$\omega_{\mathrm{D},j} = 2\pi v_0/\lambda \qquad \text{(rad/s)}$$

is the maximal Doppler angular frequency at the carrier wavelength λ for mobile number j, moving at velocity v_0. The pole radius ρ should then be selected as $\rho = 0.999 - 0.1\omega_{\mathrm{D}}T$ for $\omega_{\mathrm{D}}T \leq 0.10$. The sampling time (symbol length) T is set to 41.15μs and $\lambda = 16$cm ($\sim$1900 MHz) as in IS-136 systems. We investigate $\omega_{\mathrm{D}} \in [0.02 \; 0.10]$, corresponding to vehicle speeds from 45km/h to 225km/h.

The discussion is here simplified by assuming the model (59) to be correct and known when designing the tracker. More realistic situations, with model structure mismatch and estimation errors in the estimated Doppler frequencies, are discussed in [27].

If the two vehicles have different velocities, corresponding to $\omega_{\mathrm{D},1}$ and $\omega_{\mathrm{D},2}$ respectively, and if the channels to different receivers are assumed uncorrelated, an appropriate hypermodel (22) is given by

$$\boldsymbol{D}(q^{-1})h_t = e_t \quad (60)$$

with a diagonal auto-regression matrix

$$\boldsymbol{D}(q^{-1}) = \mathrm{diag}[\, \boldsymbol{D}_{11}(q^{-1}) \;\; \boldsymbol{D}_{12}(q^{-1}) \;\; \boldsymbol{D}_{21}(q^{-1}) \;\; \boldsymbol{D}_{22}(q^{-1}) \,] \quad (61)$$

$$\boldsymbol{D}_{ij}(q^{-1}) = D(q^{-1}, \omega_{\mathrm{D},j}T)\mathbf{I}_2 \quad (62)$$

and a block-diagonal covariance matrix for e_t

$$\mathbf{R}_e = \mathrm{diag}[\; \mathbf{R}_{e11} \;\; \mathbf{R}_{e12} \;\; \mathbf{R}_{e21} \;\; \mathbf{R}_{e22} \;]$$

where

$$\mathbf{R}_{eij} = \begin{pmatrix} \sigma_{ij0} & g_j \\ g_j & \sigma_{ij1} \end{pmatrix} .$$

All σ_{ijk} are assumed equal. The receiver is assumed to be synchronized to mobile 1, resulting in zero correlation in the taps from mobile 1 ($g_1 = 0$). We assume correlation 0.8 in the taps from mobile 2 ($g_2 = 0.8$) and fix the velocity of mobile 1 to 45km/h, while the velocity of mobile 2 is varied. The SNR is equal for both users.

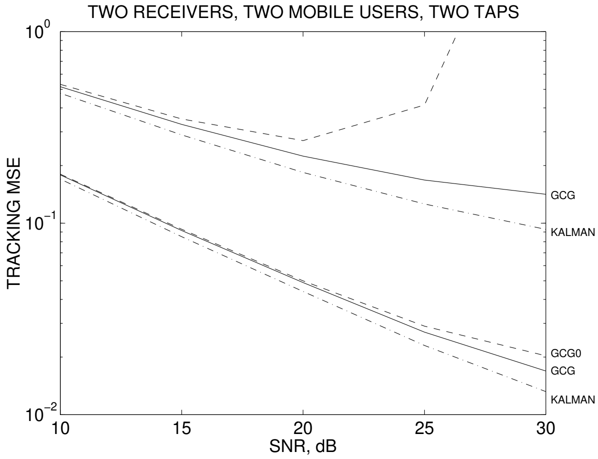

Figure 4: The sum of squared four-step channel tap prediction errors $\mathrm{tr}\,\mathbf{P}_4$ in Example 2 when the first mobile moves at 45km/h while the second mobile has velocity 45km/h (lower curves) and 225km/h (upper curves). Results for one-shot designs assuming $\eta_t = \varphi_t v_t$ (dashed), full iterative design (solid) and the Kalman estimator (dash-dotted).

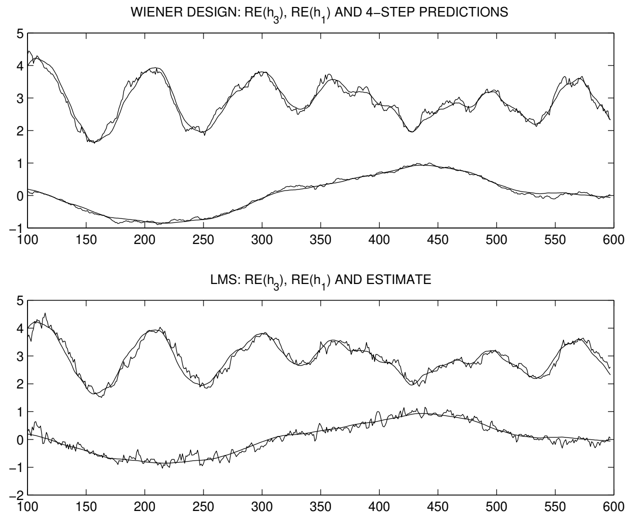

Figure 5: Tracing performance at SNR 20dB, $\omega_{\mathrm{D},1}T = 0.02$, $\omega_{\mathrm{D},2}T = 0.10$ (top), compared to LMS tracking (bottom).

Prediction estimates of the channel taps are required in equalizers. We here design four-step prediction estimators ($k = 4$) according to the iterative scheme outlined above for the two cases $\omega_{\mathrm{D},2}T = 0.02$ and $\omega_{\mathrm{D},2}T = 0.10$, and for an SNR per channel in the range 10dB-30dB. Figure 4 displays the tracking MSE $\mathrm{tr}\,\mathbf{P}_4$ for two designs: A non-iterative design assuming slow time-variations (dashed curves) and the full iterative de-

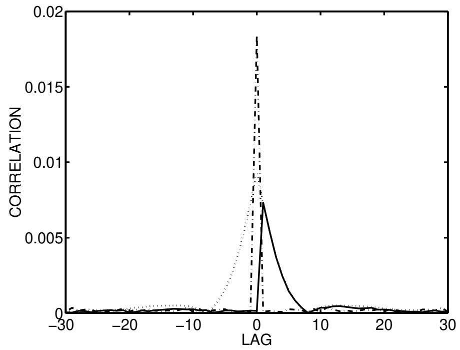

Figure 6: Absolute values of correlations in Example 2 at SNR 30dB, $\omega_{\mathrm{D},2}T = 0.10$, estimated based on 10000 data. Correlation functions for element (3,3) of $E(\eta_t \eta_{t+\tau}^*)$ (dash-dotted), $E(\eta_t \tilde{h}_{t+\tau}^*)$ (solid) and $E(\tilde{h}_{t|t-1}\tilde{h}_{t+\tau|t+\tau-1}^*)$ (dotted).

sign (solid curves), measured from simulations of (60) of length 10000. Only a single iteration was required at all design points except at 30dB in the upper curves.

The performance of the constant-gain tracker is close to that of the Kalman estimator at all operating points. This performance can be well approximated at many, but not all, operating points by the non-iterative design for slow parameter variations. The exceptions are high vehicle speeds at high SNR's: In the upper curve of Figure 4, the use of $\eta_t = \varphi_t v_t$ at SNR 30dB results in instability. A design theory based on slow time-variations simply cannot handle such situations. However, when the covariance matrix for η_t is scaled up in the first iteration, our iterative design is completed successfully.

As illustrated by Figure 6, the gradient noise is white. Furthermore, there was no significant correlation between the innovations sequence of η_t (which here equals η_t) and old tracking errors, as required by Assumption A3. This is true even at the difficult design point, SNR 30dB and $\omega_{\mathrm{D},2}T = 0.10$ (solid line for lags $\tau \leq 0$).

In Table 2, we compare the tracking MSE for Kalman predictors, the Wiener design, here denoted the general constant gain algorithm (GCG) as well as a robustly designed WLMS algorithm [26, 27], exponentially windowed RLS and an LMS estimator. We also compare their computational complexity, as measured by the number of real-valued multiplications per step[4].

The Kalman predictor is designed based on a state-space realization (52) of (60) with 16 complex-valued states. The Wiener LMS algorithm (50) is not equipped to handle elements of h_t with differing dynamics. However, it was in [27] found to be robust against variations of the Doppler frequency if $\omega_{\mathrm{D},j}T$ is set at the high end of its uncertainty range, and if integration is introduced (AR_2I-modelling). We thus design (50) for a third order model (49) with $C = 1$ and $D = D(q^{-1}, \omega_{\mathrm{D},2}T)(1-q^{-1})$, with $D(q^{-1}, \omega_{\mathrm{D},2}T)$ from (59).

From Table 2, it is evident that the GCG Wiener design attains nearly the same performance as the Kalman estimator, at much lower complexity.

The here presented GCG algorithm outperforms the simpler WLMS scheme, at the price of a somewhat higher complexity. At $\omega_{\mathrm{D},2}T = 0.10$, this is due to the better tuning of GCG to differing tap dynamics. At $\omega_{\mathrm{D},2}T = 0.02$, the difference is essentially caused by the ability of GCG to take the tap correlation for mobile number 2 into account.

Note that the use of RLS would in this example give *both* bad performance and a high computational load.

SNR	$\omega_{\mathrm{D},2}T$	Kalm.	GCG	WLMS	RLS	LMS
10	0.10	0.477	0.516	1.045	1.43	1.58
30	0.10	0.093	0.142	0.488	0.82	1.00
10	0.02	0.170	0.179	0.247	0.33	0.413
30	0.02	0.013	0.017	0.028	0.077	0.115
	#mult.	5396	416	272	1564	132

Table 2: Steady state sum of mean square tracking errors $\mathrm{tr}\,\mathbf{P}_4$ and number of real multiplications per time step in Example 2, obtained by optimized Kalman tracking, the general constant gain algorithm (GCG), WLMS, RLS and LMS adaptation algorithms.

VI. Concluding Remarks

Within the class of constant gain algorithms presented here, we can control the level of design complexity and computational complexity by selecting models for the parameters h_t and the gradient noise η_t.

The general constant-gain algorithm is based on linear time-invariant models of the parameters and of the gradient noise. If the gradient noise is assumed white, we obtain both a simpler design and a simpler implementation. Finally, the generalized WLMS and WLMS algorithms of Section IV are the simplest alternatives.

For fast time-varying parameters, the feedback noise contribution to the gradient noise η_t cannot be neglected, so an iterative design has to be performed. An alternative is to assume white gradient noise with diagonal covariance matrix $\mathbf{R}_\eta$ and use the diagonal elements as tuning knobs. For WLMS, we then have only one scalar tuning knob, the parameter drift-to-noise ratio.

Compared to Kalman adaptation laws, 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 as compared to a Kalman estimator.

One topic under present investigation is to what extent the design assumptions for open-loop Wiener design

[4] The complexities are similar, but not identical, to the ones displayed in the example in Table 1. This is partly due to other conditions (k = 4-step prediction, use of $n_D = 3$ in WLMS). In GCG, the diagonal structure of $\boldsymbol{D}$ and $\mathbf{R}_\eta$ and the block-diagonal structure of $\mathbf{R}_e$, results in 2×2 block-diagonal $\boldsymbol{\beta}$ and $\boldsymbol{Q}_4$.

are satisfied in problems with both fast variations and colored regressors.

Another interesting problem for further research is to generalize the proposed class of algorithms to handle also IIR model structures of output error, AR and ARX type.

Appendix A: Proof of Theorem 1

To prove Theorem 1, the variational approach for the derivation of polynomial design equations for Wiener filters [1],[2],[38] is adopted. Consider the filtering problem depicted in Figure 1. The estimation error $\tilde{h}_{t+k|t}$ is optimal if and only if no admissible variation ξ_t, subtracted from $\tilde{h}_{t+k|t}$, can improve the estimate.

Consider the covariance matrix of the so perturbed estimation error

$$\begin{aligned}\mathbf{P}_\xi &= \mathrm{E}\,(\tilde{h}_{t+k|t}-\xi_t)(\tilde{h}_{t+k|t}-\xi_t)^* \\ &= \mathbf{P}_k + \mathrm{E}\,\xi_t\xi_t^* - \mathrm{E}\,\tilde{h}_{t+k|t}\xi_t^* - \mathrm{E}\,\xi_t\tilde{h}^*_{t+k|t}\end{aligned} \quad (A.1)$$

If $\boldsymbol{\mathcal{L}}_k$ is adjusted so that the cross-terms vanish, then the optimal ξ_t must be zero and the covariance $\mathbf{P}_k$ obtained with the unperturbed estimator is minimal.

Derivation of the design equations: All admissible variations can be represented by

$$\xi_t = \boldsymbol{\mathcal{T}}(\mathbf{R}h_t+\eta_t) = \boldsymbol{\mathcal{T}}(\mathbf{R}\boldsymbol{D}^{-1}\boldsymbol{C}\,e_t+\eta_t) \quad , \qquad (A.2)$$

where $\boldsymbol{\mathcal{T}}$ is a stable and causal rational matrix. Since the signal ξ_t must be stationary, the factor D_u^{-1} in $\boldsymbol{D}^{-1}$ must be canceled by $\boldsymbol{\mathcal{T}}$. Thus, we require that $\boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{T}}_s D_u$, with $\boldsymbol{\mathcal{T}}_s$ being some stable and causal rational matrix. With $\tilde{h}_{t+k|t}$ given by (20), the first cross-term of (A.1) is expressed as

$$\mathrm{E}\,\tilde{h}_{t+k|t}\xi_t^*$$

$$= \mathrm{E}\,\left((q^k\mathbf{I}-\boldsymbol{\mathcal{L}}_k\mathbf{R})\boldsymbol{D}^{-1}\boldsymbol{C}\,e_t - \boldsymbol{\mathcal{L}}_k\eta_t\right)\left(\boldsymbol{\mathcal{T}}(\mathbf{R}\boldsymbol{D}^{-1}\boldsymbol{C}\,e_t+\eta_t)\right)^* .$$

We now use Parsevals formula, cf. [32], to convert the orthogonality requirement of $\tilde{h}_{t+k|t}$ and ξ_t into the frequency domain relation

$$\mathrm{E}\,\tilde{h}_{t+k|t}\xi_t^* = \frac{1}{2\pi\jmath}\oint_{|z|=1}\phi_{\tilde{h}\xi^*}\frac{dz}{z} = 0 \quad , \qquad (A.3)$$

where $\phi_{\tilde{h}\xi^*}$, the cross-spectral density between the estimation error and the variational term, is, by Assumptions A2, A3 and A4 given by

$$\left((z^k\mathbf{I}-\boldsymbol{\mathcal{L}}_k\mathbf{R})\boldsymbol{D}^{-1}\boldsymbol{C}\,\mathbf{R}_e\boldsymbol{C}_*\boldsymbol{D}_*^{-1} - \boldsymbol{\mathcal{L}}_k\frac{\boldsymbol{M}\boldsymbol{M}_*}{NN_*}\mathbf{R}^{-1}\right)\mathbf{R}\boldsymbol{\mathcal{T}}_*$$

$$= \left(z^k\boldsymbol{D}^{-1}\boldsymbol{C}\,\mathbf{R}_e\boldsymbol{C}_*N_* - \boldsymbol{\mathcal{L}}_k\mathbf{R}\boldsymbol{D}^{-1}N^{-1}\boldsymbol{\beta}\boldsymbol{\beta}_*\right)N_*^{-1}\boldsymbol{D}_*^{-1}\mathbf{R}\boldsymbol{\mathcal{T}}_* \qquad (A.4)$$

where we utilized (25) in the last equality. The orthogonality requirement is fulfilled for all admissible $\boldsymbol{\mathcal{T}}_*$ if and only if the integrand is made analytic inside the integration path. For a formal proof of this property, see e.g. Lemma A1 in Appendix A of [41].

This implies that in every element of the integrand, all poles in $|z|\leq 1$ must be canceled by zeros. We first cancel what can be canceled directly by $\boldsymbol{\mathcal{L}}_k$. Thus, let

$$\boldsymbol{\mathcal{L}}_k = \boldsymbol{D}^{-1}\boldsymbol{Q}_k\boldsymbol{\beta}^{-1}N\boldsymbol{D}\mathbf{R}^{-1} \quad , \qquad (A.5)$$

with $\boldsymbol{Q}_k$ being an undetermined causal polynomial matrix. The filter $\boldsymbol{\mathcal{L}}_k$, as expressed by (A.5), contains the marginally stable polynomial D_u in $\boldsymbol{D} = D_u\boldsymbol{D}_s$ as a common factor of all elements. After eliminating these factors, the stable expression (24) is obtained. With (A.5) inserted into (A.4), the integrand of (A.3) becomes

$$\phi_{\tilde{h}\xi^*}\frac{1}{z} = \boldsymbol{D}^{-1}\left(z^k\boldsymbol{C}\,\mathbf{R}_e\boldsymbol{C}_*N_* - \boldsymbol{Q}_k\boldsymbol{\beta}_*\right)\boldsymbol{D}_*^{-1}N_*^{-1}\mathbf{R}\boldsymbol{\mathcal{T}}_*\frac{1}{z}$$

Since $\boldsymbol{\mathcal{T}}_s$, $\boldsymbol{D}_s^{-1}$ and N^{-1} are all assumed to be stable, and $\boldsymbol{\mathcal{T}} = D_u\boldsymbol{\mathcal{T}}_s$ is required to cancel the marginally stable polynomial factor D_u of $\boldsymbol{D}$,

$$N_*(z)^{-1}\boldsymbol{D}_*(z)^{-1}\mathbf{R}\boldsymbol{\mathcal{T}}_*(z) = N_*^{-1}\boldsymbol{D}_{s*}^{-1}\mathbf{R}\boldsymbol{\mathcal{T}}_{s*}$$

will have no poles inside or on the unit circle. In order to achieve orthogonality, it is thus sufficient and necessary to require that

$$\boldsymbol{D}^{-1}\left(z^k\boldsymbol{C}\,\mathbf{R}_e\boldsymbol{C}_*N_* - \boldsymbol{Q}_k\boldsymbol{\beta}_*\right)\frac{1}{z} = \boldsymbol{L}_{k*} \quad , \qquad (A.6)$$

where $\boldsymbol{L}_{k*}(z)$ is a polynomial matrix in z only. This is equation (26). Thus, by the residue theorem,

$$\mathrm{E}\,\tilde{h}_{t+k|t}\xi_t^* = \frac{1}{2\pi\jmath}\oint_{|z|=1}\boldsymbol{L}_{k*}\boldsymbol{D}_*^{-1}N_*^{-1}\mathbf{R}\boldsymbol{\mathcal{T}}_*dz = 0 \quad .$$

Unique solvability of the Diophantine equation. The Diophantine equation (26) will always have a solution, since the invariant polynomials of $\boldsymbol{\beta}_*(q)$ are all unstable, while those of $\boldsymbol{D}(q^{-1})$ are stable or marginally stable [1],[35]. Let $\boldsymbol{Q}_k^o, \boldsymbol{L}_{k*}^o$ be one solution pair. Every solution to (26) can then be expressed as

$$(\boldsymbol{Q}_k, \boldsymbol{L}_{k*}) = (\boldsymbol{Q}_k^o - q\boldsymbol{D}\,\boldsymbol{X}, \boldsymbol{L}_{k*}^o + \boldsymbol{X}\boldsymbol{\beta}_*) \quad ,$$

where the polynomial matrix $\boldsymbol{X}$ is undetermined. Since $\boldsymbol{Q}_k$ is required to be a polynomial matrix in q^{-1} while $\boldsymbol{L}_{k*}$ is required to be a polynomial matrix in q, $\boldsymbol{X}=0$ is the only choice. Consequently, the solution to (26) is unique. The degrees (27) of $\boldsymbol{Q}_k$ and $\boldsymbol{L}_{k*}$ are determined by the requirement that the maximum powers of q^{-1} and q are covered on both sides of (26).

Stationarity of the estimation error. The estimator $\boldsymbol{\mathcal{L}}_k^{opt}$ in (24) is stable, and the noise η_t is assumed to be stationary. Thus, the last term of (20) will be stationary, with finite variance. To verify stationarity and finite variance of the estimation error $\tilde{h}_{t+k|t}$, it remains to show that $(q^k\mathbf{I}-\boldsymbol{\mathcal{L}}_k^{opt}\mathbf{R})h_t$ has finite variance even when the hypermodel contains $D_u(q^{-1})\neq 1$. This term can be expressed as

$$\left(q^k\mathbf{I} - \boldsymbol{D}_s^{-1}\boldsymbol{Q}_k\boldsymbol{\beta}^{-1}N\boldsymbol{D}_s\right)\boldsymbol{D}^{-1}\boldsymbol{C}\,e_t$$

$$= q^k \boldsymbol{D}_s^{-1} \left(\beta - q^{-k}\boldsymbol{Q}_k N\right) \beta^{-1} \boldsymbol{D}_s \boldsymbol{D}^{-1} \boldsymbol{C} e_t \ .$$

The output from this filter will be stationary with finite variance if marginally stable poles of $\boldsymbol{D}_s\boldsymbol{D}^{-1} = 1/D_u$ are canceled by the transfer function matrix

$$\beta - q^{-k}\boldsymbol{Q}_k N = D_u \boldsymbol{X}_k \tag{A.7}$$

for some polynomial matrix $\boldsymbol{X}_k$. This condition is verified by right-multiplying the left-hand side of (A.7) by β_* (which has zeros only in $|z| > 1$) and evaluating at the zeros of D_u, denoted $\{z_j\}$. We first notice that when (25) and (26) are evaluated at $\{z_j\}$, their most right hand terms vanish when $D_u \neq 1$. Thus,

$$\beta\beta_* = \boldsymbol{C}\ \mathbf{R}_e \boldsymbol{C}_* N N_* \ \ ; \ \ z^k \boldsymbol{C}\ \mathbf{R}_e \boldsymbol{C}_* N_* = \boldsymbol{Q}_k \beta_*$$

at $z = z_j$. This directly gives $\beta\beta_* - z^{-k}\boldsymbol{Q}_k N \beta_*\big|_{z=z_j} = 0$. Thus, since $\beta_*(z)$ has full rank on $|z| = 1$, $\beta - z^{-k}\boldsymbol{Q}_k N\big|_{z=z_j} = 0$, so (A.7) holds □

Appendix B: Recursive Computation of Estimators with Differing Smoothing Lags

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,

$$\boldsymbol{Q}_{k+1}(q^{-1}) = q\left(\boldsymbol{Q}_k(q^{-1}) - \boldsymbol{D}(q^{-1})\mathbf{Q}_0^k\right) \tag{B.1}$$

$$\boldsymbol{L}_{k+1*}(q) = q\boldsymbol{L}_{k*}(q) + \mathbf{Q}_0^k \beta_*(q) \tag{B.2}$$

constitute the solution to the Diophantine equation (26) for lag $k+1$ and

$$\boldsymbol{Q}_{k-1}(q^{-1}) = q^{-1}\boldsymbol{Q}_k(q^{-1}) + \boldsymbol{D}(q^{-1})\mathbf{L}_0^{k*}(\beta_0^*)^{-1} \tag{B.3}$$

$$\boldsymbol{L}_{k-1*}(q) = q^{-1}\left(\boldsymbol{L}_{k*}(q) - \mathbf{L}_0^{k*}(\beta_0^*)^{-1}\beta_*(q)\right) \tag{B.4}$$

constitute the solution to (26) for lag $k-1$. □

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

$$q^{k+1}\boldsymbol{C}\ \mathbf{R}_e \boldsymbol{C}_* N_* = \boldsymbol{Q}_{k+1}\beta_* + q\boldsymbol{D}\,\boldsymbol{L}_{k+1*} \ . \tag{B.5}$$

Multiplying both sides of (B.5) by q^{-1} and using the assumed relation (B.1) yields

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

The use of (B.2) 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 (B.3) and (B.4) are verified in the same way, with $k-1$ substituted for $k+1$ in (B.5), multiplying by q and inserting (B.3) and (B.4) □

Remark. Since $\boldsymbol{D}$ is monic and the leading coefficient of β_* is β_0^*, the leading coefficient matrix of the right hand side of (B.1) and of (B.4) will cancel. No positive powers of q are present in $\boldsymbol{Q}_{k+1}(q^{-1})$ and no negative powers of q are present in $\boldsymbol{L}_{k-1*}(q)$ □

Appendix C: Stability of Wiener Adaptation for Slow Variations

The effect of the the Wiener design when the power of the gradient noise η_t is assumed to become very large relative to the power of the noise e_t driving h_t, can equivalently be represented by fixing $\boldsymbol{M}(q^{-1})$ and letting $\mathbf{R}_e \rightarrow 0$ in (21), (22). The left-hand side of the Diophantine equation (26) will then vanish, so a limiting solution is $\boldsymbol{Q}_k(q^{-1}) = 0$,$\boldsymbol{L}_{k*}(q) = 0$. This solution is unique, see Appendix A. Since $\boldsymbol{Q}_k(q^{-1}) \rightarrow 0$ while all other factors in (24) remain bounded for stable models, $||\boldsymbol{\mathcal{L}}_k^{opt}||_2^2 \rightarrow 0$ with a decreasing signal-to-noise ratio.

The small gain theorem, see e.g. [43], implies that if $\boldsymbol{\mathcal{L}}_1(q^{-1})$ is causal and L_p-stable, stability of the outer time-varying feedback loop of Figure 2 is preserved if

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

If φ_t, and thus also $\mathbf{Z}_t$, is assumed to have bounded elements, (C.1) will be fulfilled when we perform a Wiener design (that will always produce a stable $\boldsymbol{\mathcal{L}}_1(q^{-1})$), if the assumed noise level is sufficiently high, so that $||\boldsymbol{\mathcal{L}}_1(q^{-1})||_p||\mathbf{Z}_t||_p < 1$ □

References

[1] A. Ahlén and M. Sternad, "Wiener filter design using polynomial equations," *IEEE Transactions on Signal Processing*, 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] A. Ahlén, L. Lindbom and M. Sternad, "Analysis of stability and performance of adaptation algorithms with time-invariant gains". Submitted. www.signal.uu.se/Publications/abstracts/r002.html

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

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

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

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

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

[9] 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.

[10] A. W. Fuxjaeger and R. A. Iltis, "Adaptive parameter estimation using parallel Kalman filtering for spread spectrum code and Doppler tracking," *IEEE Transactions on Communications*, vol. 42, pp. 2227-2230, 1994.

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

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

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

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

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

[16] W. C. Jakes, "Multipath Interference". In W. C. Jakes ed. *Microwave Mobile Communications.* Wiley, New York, NY, 1974. Reissued by IEEE, Piscataway NJ.

[17] 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.

[18] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.

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

[20] J. Ježek and V. Kůcera, "Efficient algorithm for matrix spectral factorization," *Automatica*, vol. 21, pp. 663-669, 1985.

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

[22] 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.

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

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

[25] 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

[26] L. Lindbom, M. Sternad and A. Ahlén, "Tracking of time-varying mobile radio channels. Part I: The Wiener LMS algorithm". *IEEE Trans. on Commun*, vol 49, pp. 2207-2217, Dec. 2001.
www.signal.uu.se/Publications/abstracts/r003.html

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

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

[29] 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.

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

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

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

[33] "PCS IS136-based air-interface compatibility 1900MHz standard, Part I and II," SP 3388-1, SP 3388-2, J-STD-011.

[34] Polynomial toolbox for use with Matlab, version 2.5. See www.polyx.com

[35] A. P. Roberts and M. M. Newmann, "Polynomial approach to Wiener filtering," *International Journal of Control*, vol. 47, pp. 681-696, 1988.

[36] 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.

[37] 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.

[38] M. Sternad and A. Ahlén, "H2 design of nominal and robust discrete time filters," In M. J. Grimble and V Kůcera , eds., *Polynomial Methods for Control Systems Design,* Springer, London, 1996, pp. 171-222.

[39] 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.

[40] C. Tidestav, M. Sternad and A. Ahlén, "Reuse within a cell - interference rejection or multiuser detection?" *IEEE Transactions on Communications*, vol. 47, pp. 1511-1522, 1999.

[41] C. Tidestav, *The Multivariable Decision Feedback Equalizer: Multiuser Detection and Interference Rejection.* PhD Thesis, Signals and Systems, Uppsala University, Sweden, 1999.
www.signal.uu.se/Publications/abstracts/a993.html

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

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

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

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

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