Tracking of Time-varying Systems. Part I: Wiener Design of Algorithms with Time-invariant Gains

Mikael Sternad*, Lars Lindbom and Anders Ahlén*

October 3, 2000

Abstract- **We present a design method for adaptation laws that extend LMS by including general time-invariant filters. These adaptation algorithms are designed for tracking time-varying parameters of linear regression models, in situations where the regressors are stationary or have slowly time-varying properties. The structure and gain of the adaptation law are optimized for time-variations modeled as correlated ARIMA processes. The aim is to systematically use any available prior information to provide filtering, prediction or fixed lag smoothing estimates for arbitrary lags. We derive the optimal tracking algorithm with time-invariant adaptation gain. Compared to Kalman estimators, the tracking performance is nearly the same, while the complexity is much lower.**

The design method is based on a novel 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 is important for fast variations, where the filter design is performed iteratively. The most 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.

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 more 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.

*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.

We will here present a novel approach to the design of such adaptation laws that is based on Wiener filtering techniques. In Part II [28] we present results for the analysis of stability, performance and convergence in MSE of such algorithms.

The difficult problem of accurately tracking time-varying radio channels in D-AMPS 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 attempt to accurately track D-AMPS channels was reported in [26]. Subsequently, the algorithm proposed in [26] has been used in [17, 32] on D-AMPS 1900 channels. A thorough case study on this application was presented in [30].

A sequence of measurement vectors $\{y_t\}$ of dimension $n_y|1$ 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 $n_y|n_h$, is stationary with zero mean and covariance matrix

$$\mathbf{R} \triangleq \mathrm{E}\left[\varphi_t \varphi_t^*\right] \quad , \tag{2}$$

which is assumed to be nonsingular. In practice, $\mathbf{R}$ can be allowed to be slowly time-varying. The noise v_t is assumed to be a stationary stochastic process with zero mean and covariance matrix $\mathbf{R}_v$.

Our aim is to estimate the time-varying parameter vector $h_t = [h_{0,t} \dots h_{n_h-1,t}]^T$ in an environment with stationary (or slowly time-varying) statistics of the regressors and the noise. We here exclude AR- and ARX models, where the use of old measurements as regressors could make φ_t^* 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 $h_1, h_2 \dots$ with more elements than the available measurements $y_1, y_2 \dots$. To avoid this dilemma, models that represent assumptions on the relationship between h_t and h_τ for $\tau \neq t$ must be introduced.

Dynamic models of the time-varying parameters, sometimes denoted *hypermodels* [4, 5], may be deterministic [6, 7, 11, 25, 33] or stochastic [10, 20]. 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 $n_h|n_h$ 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

$$\tilde{h}_{t+k|t} \triangleq h_{t+k} - \hat{h}_{t+k|t} \tag{4}$$

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 ($k < 0$). We measure tracking performance by the steady state covariance matrix

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

where the expectation is taken with respect to e_t in (3) and v_t in (1) when $t \to \infty$ (after the initial transients).

Among all adaptation laws which perform linear operations on y_t, the Kalman filter will minimize (5) if $\varphi^*_t, \mathbf{R}_v, \boldsymbol{\mathcal{H}}(q^{-1})$ and $\mathbf{R}_e$ in (1),(3) are known [3, 18]. Kalman-based adaptive filters discussed in the literature are mostly based on first order models [12, 15, 37]

$$h_t = ah_{t-1} + e_t \;\Leftrightarrow\; h_t = \frac{1}{1-aq^{-1}}\mathbf{I}\; e_t \quad , \tag{6}$$

where a is a scalar, but Kalman estimators can of course be designed for more complicated linear models[1].

The computational complexity of Kalman estimators may not seldomly preclude their use, particularly in digital communications. A commonly used alternative of much lower complexity is the LMS law

$$\begin{aligned} \varepsilon_t &= y_t - \varphi^*_t \hat{h}_{t|t-1} && (7)\\ \hat{h}_{t+1|t} &= \hat{h}_{t|t-1} + \mu\varphi_t\varepsilon_t && (8) \end{aligned}$$

or equivalently

$$\hat{h}_{t+1|t} = \frac{\mu}{1-q^{-1}}\mathbf{I}\; \varphi_t\varepsilon_t \tag{9}$$

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 lags k.

[1]See e.g. [13, 12] for a discussion of Kalman estimators based on models $h_t = F_t h_{t-1} + e_t$ with a time-varying transition matrix F_t of dimension n_h.

The class of estimators generalize LMS by substituting a time-invariant matrix of linear transfer operators for the LMS operator $\mu/(1-q^{-1})\mathbf{I}$ in (9).

Based on (1),(3), the filter $\boldsymbol{\mathcal{M}}_k(q^{-1})$ in

$$\begin{aligned} \varepsilon_t &= y_t - \varphi^*_t \hat{h}_{t|t-1} && (10)\\ \hat{h}_{t+k|t} &= \boldsymbol{\mathcal{M}}_k(q^{-1})\varphi_t\varepsilon_t && (11) \end{aligned}$$

is to be designed so that (5) is minimized, under various structural constraints and assumptions. If in particular $\boldsymbol{\mathcal{M}}_k(q^{-1})$ is constrained to be diagonal, then the complexity of the algorithm grows linearly with n_h.

A related research program has been followed by Benveniste and co-workers [4, 5] who used state-space models and methods to perform an interesting analysis of multi-step adaptation laws with constant gains. However, this work, as well as practically all other analysis of LMS, RLS and Kalman-based tracking, has been focused exclusively on cases with slowly time-varying dynamics, since only then can tools of weak convergence theory and various methods of averaging [22, 31] be used.

By (1), (10) and (11), the tracking problem for $k = 1$ corresponds to the design of a time-invariant feedback controller $\boldsymbol{\mathcal{M}}_1(q^{-1})$ for a time-varying system $\varphi_t\varphi^*_t$, as illustrated by Figure 1.[2] Formulated in this way, the problem becomes tractable, and can be solved by an infinite horizon LQG or $\mathcal{H}_2$ feedback design, if $\varphi_t\varphi^*_t$ is approximated by its time-invariant average $\mathbf{R}$. However, such an approximation is valid only in the case of slow parameter variations, see Part II [28].

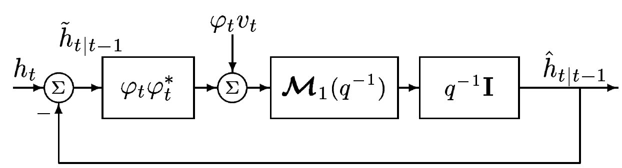

Figure 1: The one-step predictor ($k = 1$ in (10),(11)) could be seen as a linear time-invariant feedback regulator for a time-varying system containing a one-step delay $q^{-1}\mathbf{I}$ and a time-varying block $\varphi_t\varphi^*_t$.

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 fictitious measurement signal that can be constructed from y_t. In Section III, this filter is optimized using a polynomial approach [1, 2]. Section IV specializes to simplified solutions and Section V summarizes the proposed iterative design process and illustrates it with a design example.

Remarks on the 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

[2]See [23, 24] for an LMS analysis using this feedback structure.

operator q for q^{-1}. To simplify notation, the arguments q or q^{-1} are often omitted. 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 Measurements

Consider the signal prediction error (10) and insert (1) describing y_t, to obtain

$$\begin{aligned} \varepsilon_t &= \varphi_t^*(h_t - \hat{h}_{t|t-1}) + v_t \\ \varphi_t \varepsilon_t &= \varphi_t \varphi_t^* \tilde{h}_{t|t-1} + \varphi_t v_t \quad . \end{aligned} \tag{12}$$

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

$$Z_t = \varphi_t \varphi_t^* - \mathbf{R} \tag{13}$$

$$\eta_t = Z_t \tilde{h}_{t|t-1} + \varphi_t v_t \tag{14}$$

$$f_t = \mathbf{R} h_t + \eta_t \quad , \tag{15}$$

the vector $\varphi_t \varepsilon_t$ in (12) is now reformulated as

$$\begin{aligned} \varphi_t \varepsilon_t &= \mathbf{R}\tilde{h}_{t|t-1} + Z_t \tilde{h}_{t|t-1} + \varphi_t v_t \\ &= f_t - \mathbf{R}\hat{h}_{t|t-1} \quad . \end{aligned} \tag{16}$$

The feedback loop of Figure 1 has now been transformed into a time-invariant feedback $\mathbf{R}\tilde{h}_{t|t-1}$ plus a perturbation loop represented by $Z_t \tilde{h}_{t|t-1}$. The signal f_t defined in (15) can be regarded as a fictitious measurement, with $\mathbf{R}h_t$ and η_t being the signal and the noise, respectively. It can be constructed from known signals as depicted in the lower diagram of Figure 2. In the sequel, the noise terms η_t and $Z_t \tilde{h}_{t|t-1}$ will be referred to as the *gradient noise* and the *feedback noise*, respectively. The matrix Z_t, of dimension $n_h|n_h$, has zero mean by definition. This matrix was introduced by Gardner [9] and was referred to as the *autocorrelation matrix noise*.

B. Tracking Regarded as Time-invariant Filtering

Based on the relations (13)-(16), we may design a time-invariant stable rational matrix $\boldsymbol{\mathcal{L}}_k(q^{-1})$ that operates on f_t and provides an estimate of h_{t+k}

$$f_t = \mathbf{R}\hat{h}_{t|t-1} + \varphi_t \varepsilon_t = \mathbf{R}h_t + \eta_t \tag{17}$$

$$\hat{h}_{t+k|t} = \boldsymbol{\mathcal{L}}_k(q^{-1}) f_t \quad . \tag{18}$$

This filter $\boldsymbol{\mathcal{L}}_k(q^{-1})$ will be referred to as the *learning filter.* It will be shown that the design of a learning filter is equivalent to the design of $\boldsymbol{\mathcal{M}}_k(q^{-1})$ in (10)-(11).

As seen in the upper part of Figure 2, three terms influence the tracking performance via f_t: The scaled and

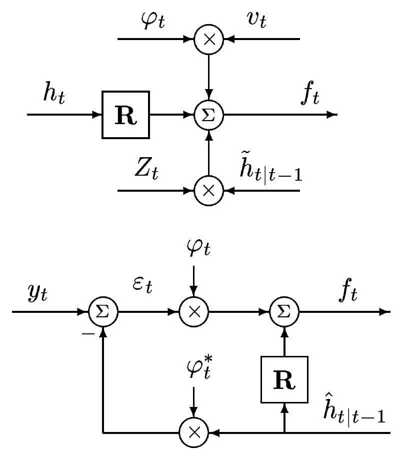

Figure 2: Two equivalent representations of the fictitious measurement f_t. The lower diagram constructs f_t via (16) from available signals when $\mathbf{R} = E[\varphi_t \varphi_t^*]$ is known.

rotated parameters $\mathbf{R}h_t$, representing the useful signal; the noise $\varphi_t v_t$; and old tracking errors via the feedback noise $Z_t \tilde{h}_{t|t-1}$.

The estimation error follows from (15) and (18) as

$$\tilde{h}_{t+k|t} = (q^k \mathbf{I} - \boldsymbol{\mathcal{L}}_k(q^{-1})\mathbf{R})h_t - \boldsymbol{\mathcal{L}}_k(q^{-1})\eta_t \quad , \tag{19}$$

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

If the innovations of η_t are uncorrelated with $\hat{h}_{t-i|t-i-1}, i \geq 0$, then an open-loop Wiener design of $\boldsymbol{\mathcal{L}}_k$ can be performed. Possible higher-order statistical dependencies 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. These conditions will not always be fulfilled but they hold approximately in many practically important circumstances, since the multiplication by Z_t in (14) acts as a scrambler.

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 an assumption, under which $\boldsymbol{\mathcal{L}}_k(q^{-1})$ will be optimized by just treating η_t in (19) as an additive noise, with known properties.

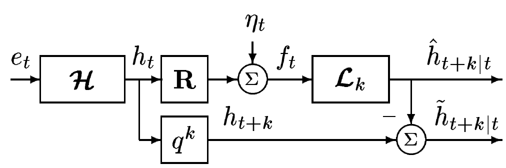

Figure 3: The filter design problem. The vector h_{t+k} is to be estimated from measurements f_t, such that the steady state tracking error covariance matrix is minimized.

C. Properties of the Gradient Noise

The feedback noise $Z_t\tilde{h}_{t|t-1}$ will not be *independent* of $\hat{h}_{t-i|t-i-1}$, due to the feedback loop in Figure 4 and the loop could become unstable. As discussed in Part II [28], the gain of $\mathcal{L}_1(q^{-1})$ cannot be allowed to be too large.

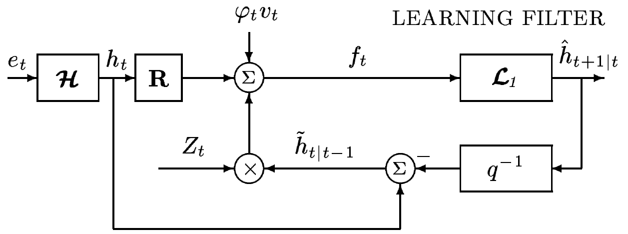

Figure 4: The feedback loop via the feedback noise $Z_t\tilde{h}_{t|t-1}$ may significantly affect the variance of the fictitious measurement f_t, and causes dependence with $\tilde{h}_{t|t-1}$.

Since the properties of η_t depend on $\mathcal{L}_1$, a tracking design will 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 Part II [28], 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. When the power of e_t becomes small relative to the power of $\varphi_t v_t$, then the impact of the feedback noise $Z_t\tilde{h}_{t|t-1}$ on our obtained tracking MSE vanishes. The feedback noise is negligible either when the parameters h_t vary slowly, or when the noise level is high[3]. Then, $\eta_t \approx \varphi_t v_t$, and η_t will be white whenever v_t or φ_t is a white sequence.

2. *Independent consecutive regression matrices.* If φ_t^* and φ_s^* are independent for $t \neq s$, then the feedback noise $Z_t\tilde{h}_{t|t-1}$ will be white with zero mean and its covariance can be derived exactly.

3. *FIR models with white zero mean regressors.* The performance can then be predicted exactly for models with two coefficients and with good accuracy for higher order models, 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

$$\operatorname{tr}\mathbf{P}_k = \lim_{t\to\infty} \mathrm{E} \sum_{i=0}^{n_h-1} |h_{i,t+k} - \hat{h}_{i,t+k|t}|^2 \ . \qquad (20)$$

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 and has a known rational spectral density $\phi_\eta(e^{\jmath\omega})$ modeled by a vector ARMA process

$$\eta_t = \frac{1}{N(q^{-1})}\boldsymbol{M}(q^{-1})\nu_t \ , \qquad (21)$$

where $\boldsymbol{M}$ is an $n_h|n_h$ 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 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

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

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

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

are time-invariant[4]. Above, $\boldsymbol{C}(q^{-1})$ is assumed stable, $D_u(q^{-1})$ is a polynomial with zeros on the unit circle and $\boldsymbol{D}_s(q^{-1})$ is a stable polynomial matrix □

Assumption 4 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 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

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

where the polynomial matrix

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

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

[4] 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.

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

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

The unique solution to the bilateral Diophantine equation

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

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 $n_h|n_h$, with generic degrees

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

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

Proof: See Appendix A.

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, 36, 39]. The Diophantine equation (26) is guaranteed to be solvable and 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) guarantees a stable spectral factor $\boldsymbol{\beta}$ with a leading matrix β_0 of full rank. Thus, $\boldsymbol{\beta}^{-1}$ is causal and stable.

Algorithms for solving polynomial matrix spectral factorizations and bilateral Diophantine equations are presented in [14, 19] and in [21].

The learning filters have real-valued coefficients when $\boldsymbol{C}, \boldsymbol{D}, \boldsymbol{M}, N$ 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. For predictors, $k > 0$, the complexity of the estimator (determined by the degree of $\boldsymbol{Q}_k$) is independent of the prediction horizon.

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.

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, then it is evident from (24) that all elements of $\boldsymbol{\mathcal{L}}_k^{opt}$ will have a notch at ω_1 since $N(e^{-j\omega_1}) \approx 0$, so

$$\boldsymbol{\mathcal{L}}_k^{opt}\Big|_{\omega=\omega_1} \approx 0 \ .$$

On the other hand, when the gradient noise is negligible, $\boldsymbol{M} \approx 0$, and (25), (26) are then for $k \leq 0$ solved by

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

The lag error in (19) then vanishes since

$$\boldsymbol{\mathcal{L}}_k^{opt} \approx q^k\mathbf{R}^{-1}$$

and this estimator attains $\mathbf{P}_k \approx 0$ for $k \leq 0$.

Weighted criteria and the excess MSE. In Appendix B, a generalization of Theorem 1 to dynamically weighted errors

$$\tilde{h}^w_{t+k|t} \triangleq \boldsymbol{\mathcal{W}}(q^{-1})\tilde{h}_{t+k|t}$$

is outlined. This result can be applied if we wish to minimize the excess mean square output error $\mathrm{E}\,||\varepsilon_t||_2^2 - \mathrm{tr}\,\mathbf{R}_v$ when regressors and tracking errors are assumed independent. Minimizing the excess MSE then corresponds to minimizing the trace of the covariance matrix of

$$\tilde{h}^w_{t+1|t} = \mathbf{R}^{1/2}\tilde{h}_{t+1|t} \ . \quad (28)$$

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 the equations in Theorem 1 for one value of k only. It is shown in Appendix C that the solutions for all k can then be computed 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 [27] for details, [46] for general methods and [30] for a specialization to fading mobile radio channels parametrized by uncertain Doppler frequencies.

C. Realizations and Interpretations

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

$$\begin{aligned}\phi_f &= \mathbf{R}\boldsymbol{D}^{-1}\boldsymbol{C}\,\mathbf{R}_e\boldsymbol{C}_*\boldsymbol{D}_*^{-1}\mathbf{R} + \frac{1}{NN_*}\boldsymbol{M}\boldsymbol{M}_* \\ &= \mathbf{R}\boldsymbol{D}^{-1}N^{-1}\boldsymbol{\beta}\boldsymbol{\beta}_*N_*^{-1}\boldsymbol{D}_*^{-1}\mathbf{R} \quad (29)\end{aligned}$$

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

$$f_t = \mathbf{R}\boldsymbol{D}^{-1}N^{-1}\boldsymbol{\beta}\epsilon_t \ \Leftrightarrow \ \epsilon_t = \boldsymbol{\beta}^{-1}N\boldsymbol{D}\,\mathbf{R}^{-1}f_t \quad (30)$$

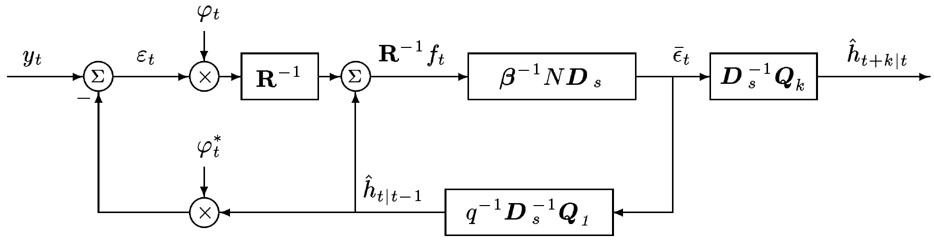

Figure 5: The optimized tracking algorithm with time-invariant gains.

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, (30) corresponds to a generalized innovation model [38]. By defining the signal

$$\bar{\epsilon}_t \triangleq \frac{1}{D_u(q^{-1})} \epsilon_t \;=\; \boldsymbol{\beta}^{-1} N \boldsymbol{D}_s \mathbf{R}^{-1} f_t \qquad (31)$$

we thus obtain a realization of (24) (Figure 5):

$$\varepsilon_t = y_t - \varphi_t^* \hat{h}_{t|t-1} \qquad (32)$$

$$\bar{\epsilon}_t = \boldsymbol{\beta}^{-1} N \boldsymbol{D}_s (\mathbf{R}^{-1} \varphi_t \varepsilon_t + \hat{h}_{t|t-1}) \qquad (33)$$

$$\hat{h}_{t+k|t} = \boldsymbol{D}_s^{-1} \boldsymbol{Q}_k \bar{\epsilon}_t \ . \qquad (34)$$

The realization (32)-(34) has good numerical properties since all involved filters are internally stable. 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 [8].

Corollary 1. The Wiener optimized filter $\boldsymbol{\mathcal{M}}_k$. The estimator (11) optimized by Theorem 1 is given by

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

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

$$\boldsymbol{\mathcal{R}} \triangleq [\boldsymbol{\beta} - q^{-1} N \boldsymbol{Q}_1]^{-1} N \boldsymbol{D}_s = \frac{1}{D_u} \boldsymbol{X}_1^{-1} N \boldsymbol{D}_s \qquad (36)$$

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

$$\boldsymbol{\beta} - q^{-1} \boldsymbol{Q}_1 N = D_u \boldsymbol{X}_1 \ . \qquad (37)$$

□

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 (33). 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,

$$\bar{\epsilon}_t = \boldsymbol{\mathcal{R}} \mathbf{R}^{-1} \varphi_t \varepsilon_t \ . \qquad (38)$$

The use of this expression in (34) gives (35)-(36), while (37) is verified in Appendix A. The leading coefficient of $\boldsymbol{X}_1$ is β_0 which has full rank, so $\boldsymbol{X}_1^{-1}(q^{-1})$ is causal □

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, $\boldsymbol{\mathcal{M}}_k(q^{-1})$ in (11) need not be stable, since it is a block in a feedback loop, see Figure 1. By (36), marginally stable model factors D_u^{-1} will be present in all elements of $\boldsymbol{\mathcal{M}}_k$.

If we define

$$\boldsymbol{D}(q^{-1}) \triangleq \mathbf{I} - q^{-1} \boldsymbol{F}(q^{-1}) \qquad (39)$$

so (3) could be expressed as

$$h_t = \boldsymbol{F}(q^{-1}) h_{t-1} + \boldsymbol{C}(q^{-1}) e_t \ , \qquad (40)$$

the realization (35) can then, after multiplying by D_u, be rewritten as

$$\hat{h}_{t+k|t} = \boldsymbol{F} \hat{h}_{t+k-1|t-1} + D_u \boldsymbol{Q}_k \boldsymbol{\mathcal{R}} \mathbf{R}^{-1} \varphi_t \varepsilon_t \ . \qquad (41)$$

The marginally stable factor D_u is a common factor in (41) and must be eliminated before implementation.

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 process and the implementation is simplified. It is shown in Part II [28] that the feedback noise is indeed negligible or white in several types of problems.

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 arbitrary lags k are presented in Corollary 2 in Appendix C.

Lemma 1. For white gradient noise η_t with covariance matrix $\mathbf{R}_\eta$, the solution to the Diophantine equation (26) for $k = 1$ is given by

$$\boldsymbol{Q}_1(q^{-1}) = q(\boldsymbol{\beta}(q^{-1}) - \boldsymbol{D}(q^{-1}) \beta_0) \qquad (42)$$
$$\boldsymbol{L}_{1*}(q) = \beta_0 \boldsymbol{\beta}_*(q) - \mathbf{R}^{-1} \mathbf{R}_\eta \mathbf{R}^{-1} \boldsymbol{D}_*(q) \qquad (43)$$

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

$$\boldsymbol{\beta}\boldsymbol{\beta}_* = \boldsymbol{C}\,\mathbf{R}_e \boldsymbol{C}_* + \boldsymbol{D}\,\mathbf{R}^{-1}\mathbf{R}_\eta \mathbf{R}^{-1}\boldsymbol{D}_* \quad (44)$$

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

$$q\boldsymbol{C}\;\mathbf{R}_e\boldsymbol{C}_* = \boldsymbol{Q}_1\boldsymbol{\beta}_* + q\boldsymbol{D}\,\boldsymbol{L}_{1*} \;\; . \quad (45)$$

By substituting the expressions (42) and (43) into the right hand side of (45), the lemma is verified □

Furthermore, the solution $\boldsymbol{\beta}(q^{-1})$ to the spectral factorization (25) can in the white noise case be obtained conveniently from the solution to an algebraic Riccati equation. See Result 3.3 in [27] for details.

The implementation of the tracker is also simplified since by (42) and $N = 1$ in (36),

$$\begin{aligned}\boldsymbol{\mathcal{R}} &= \left(\boldsymbol{\beta} - q^{-1}\boldsymbol{Q}_1\right)^{-1}\boldsymbol{D}_s \\ &= \left(\boldsymbol{\beta} - (\boldsymbol{\beta} - \boldsymbol{D}\,\beta_0)\right)^{-1}\boldsymbol{D}_s \;=\; \frac{1}{D_u}\beta_0^{-1} \quad (46)\end{aligned}$$

which simplifies the realizations of (35) and (41). By using (46) in (38), the innovation processes are

$$\bar{\epsilon}_t = \frac{1}{D_u}\beta_0^{-1}\mathbf{R}^{-1}\varphi_t\varepsilon_t \;\; ; \;\; \epsilon_t = \beta_0^{-1}\mathbf{R}^{-1}\varphi_t\varepsilon_t \;\; . \quad (47)$$

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 real scalar with $|a| \leq 1$ and $\mathbf{R}_e$ is given. The gradient noise is white, with $\mathbf{R}_\eta$ known. The spectral factorization (44) then becomes

$$(\beta_0+\beta_1 q^{-1})(\beta_0^*+\beta_1^* q) = \mathbf{R}_e+(1-aq^{-1})\mathbf{R}^{-1}\mathbf{R}_\eta\mathbf{R}^{-1}(1-aq) \quad (48)$$

Using $\mathbf{R}_\gamma \stackrel{\Delta}{=} \mathbf{R}^{-1}\mathbf{R}_\eta\mathbf{R}^{-1}$, we obtain

$$\begin{aligned}\beta_0\beta_0^* + \beta_1\beta_1^* &= \mathbf{R}_e + (1+a^2)\mathbf{R}_\gamma \\ \beta_1\beta_0^* = \beta_0\beta_1^* &= -a\mathbf{R}_\gamma\end{aligned}$$

These expressions represent a set of coupled second order equations in the elements of β_0 and β_1.

If we are interested in the one-step predictor, the solution (42) for $k = 1$ directly gives $\boldsymbol{Q}_1(q^{-1})$ as

$$\boldsymbol{Q}_1 = q(\beta_0 + \beta_1 q^{-1} - \beta_0 - a\beta_0 q^{-1}) \;=\; \beta_1 - a\beta_0 \;\; .$$

From (39), $\boldsymbol{F}\,(q^{-1}) = a\mathbf{I}$ and from (46), $\boldsymbol{\mathcal{R}}D_u = \beta_0^{-1}$. Thus, the realization (41) corresponds to a generalized LMS update equation

$$\hat{h}_{t+1|t} = a\hat{h}_{t|t-1} + (\beta_1 - a\beta_0)\beta_0^{-1}\mathbf{R}^{-1}\varphi_t\varepsilon_t \;. \quad (49)$$

The update (49) 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 of LMS (8). 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, β_0 and β_1 from (48) become diagonal and have all diagonal elements equal, so

$$(\beta_1 - a\beta_0)\beta_0^{-1}\mathbf{R}^{-1} = \mu\mathbf{I}$$

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 tracking MSE criterion (20) for possibly colored gradient noise (21) and for hypermodels in common denominator form

$$h_t = \frac{1}{D(q^{-1})}\boldsymbol{C}\,(q^{-1})e_t \;\; . \quad (50)$$

The structure of the learning filter is constrained to

$$\hat{h}_{t+k|t} = \boldsymbol{\mathcal{S}}_k(q^{-1})\mathbf{R}^{-1}f_t \;\; , \quad (51)$$

where $\boldsymbol{\mathcal{S}}_k(q^{-1})$ is a *diagonal* stable rational matrix. The design equations for this filter are derived and presented in [27]. They consist of n_h separate polynomial spectral factorizations and n_h scalar Diophantine equations.

When the regressors are white ($\mathbf{R}$ diagonal), or when the update suggested in [8] for FIR models with autoregressively generated inputs can be used, the computational complexity of this estimator grows linearly with the number of parameters n_h.

B. The Wiener LMS Algorithm

A further simplification is obtained by minimizing (20) for *white* gradient noise with covariance $\mathbf{R}_\eta$, and for diagonal hypermodels with equal elements

$$h_t = \frac{C(q^{-1})}{D(q^{-1})}\mathbf{I}\;e_t \;\; . \quad (52)$$

The learning filter is restricted to (51) but with all filters on the diagonal of $\boldsymbol{\mathcal{S}}_k(q^{-1})$ being equal:

$$\hat{h}_{t+k|t} = \frac{Q_k(q^{-1})}{\beta(q^{-1})}\mathbf{R}^{-1}f_t \;\; . \quad (53)$$

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 $\gamma \triangleq \mathrm{tr}\,\mathbf{R}_e/\mathrm{tr}\,\mathbf{R}^{-1}\mathbf{R}_\eta\mathbf{R}^{-1}$. The polynomials $Q_k(q^{-1})$ and $\beta(q^{-1})$ in (53) minimizing (20) are then obtained, together with a polynomial $L_{k*}(q)$, as the solution to one polynomial spectral factorization and one polynomial Diophantine equation

$$r\beta\beta_* = \gamma CC_* + DD_* \tag{54}$$
$$q^k\gamma CC_* = rQ_k\beta_* + qDL_{k*} \ . \tag{55}$$

As in Lemma 1, there exists closed-form solutions to (55). When the orders of the polynomials in (52) are no higher than 2, there furthermore exists a closed-form solution to the spectral factorization (54) [35].

An alternative realization of this estimator is

$$\hat{h}_{t|t} = \hat{h}_{t|t-1} + \mu\mathbf{R}^{-1}\varphi_t\varepsilon_t \tag{56}$$
$$\hat{h}_{t+k|t} = \frac{Q_k(q^{-1})}{Q_0(q^{-1})}\hat{h}_{t|t} \ , \tag{57}$$

where $Q_0(z^{-1})$ can be shown to be stable and where the scalar gain μ equals the leading coefficient of $Q_0(q^{-1})$

$$\mu = Q_0^0 = 1 - \frac{1}{r} \ , \tag{58}$$

with $r > 1$ obtained from (54). The WLMS algorithm is derived and discussed in [29] and is applied to the tracking of fading mobile radio channels in [30].

V. Iterative Design

For slow time-variations, the feedback noise is by definition negligible [28], 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

$$\mathbf{R}_{\eta\eta}(j) \triangleq \mathrm{E}\,[\eta_t\eta_{t-j}^*] \ . \tag{59}$$

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

$$\boldsymbol{M}(q^{-1})\boldsymbol{M}_*(q) = \sum_{j=-n_M}^{n_M} \mathbf{R}_{\eta\eta}(j)q^j \ . \tag{60}$$

Note that only the total covariance function (60), not the MA-model matrix $\boldsymbol{M}$, is needed in the design equations (26) and (25).

We 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 $\boldsymbol{\mathcal{L}}_1(q^{-1})$. Verify that the closed loop of Figure 4 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. 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 $\hat{h}_{t+1|t} = \boldsymbol{\mathcal{L}}_1(q^{-1})f_t$, obtain an estimated gradient noise time series from (17) as

$$\hat{\eta}_t = f_t - \mathbf{R}h_t = \varphi_t\varepsilon_t - \mathbf{R}(h_t - \hat{h}_{t|t-1}) \ . \tag{61}$$

Obtain an estimate $\hat{\mathbf{R}}_{\eta\eta}(j)$ of the covariance function (59),(60) 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 $\boldsymbol{\mathcal{L}}_1(\omega) \rightarrow 0 \ \forall \ \omega$ when the assumed noise power is increased. If Z_t has bounded elements, then the small gain theorem [43] will therefore imply that the closed loop of Figure 4 can be stabilized by assuming a sufficiently high noise power in the design of $\boldsymbol{\mathcal{L}}_1$.

The covariance function estimate provides additional information. If $\hat{\mathbf{R}}_{\eta\eta}(0)$ does not differ much from $\mathbf{R}\sigma_v^2$, then the time-variations can be regarded as slow, and step 1 above turns out to be sufficient. 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 $\mathbf{R}_\eta = \hat{\mathbf{R}}_{\eta\eta}(0)$.

It should be emphasized that the design methodology assumes a good hypermodel. With incorrect models, there is no reason to believe that the iterations minimize the tracking MSE.

Example 2. Iterative design and a comparison to Kalman, WLMS 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

$$\begin{pmatrix} y_t^1 \\ y_t^2 \end{pmatrix} = \begin{pmatrix} B_t^{11}(q^{-1}) & B_t^{12}(q^{-1}) \\ B_t^{21}(q^{-1}) & B_t^{22}(q^{-1}) \end{pmatrix} \begin{pmatrix} u_t^1 \\ u_t^2 \end{pmatrix} + \begin{pmatrix} v_t^1 \\ v_t^2 \end{pmatrix} , \tag{62}$$

where y_t^i is the sampled baseband signal at receiver i. Two-tap channels are assumed as in the IS-136 system [30], so

$$B_t^{ij}(q^{-1}) = b_{0,t}^{ij} + b_{1,t}^{ij}q^{-1} \ . \tag{63}$$

The model (62) can then 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

$$h_t = (\ b^{11}_{0,t}\ \ b^{11}_{1,t}\ \ b^{12}_{0,t}\ \ b^{12}_{1,t}\ \ b^{21}_{0,t}\ \ b^{21}_{1,t}\ \ b^{22}_{0,t}\ \ b^{22}_{1,t}\ )^T\ . \quad (64)$$

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

The messages are transmitted from moving mobile terminals, so the channel taps $b^{ij}_{\ell,t}$ 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.

$$\frac{1}{1 - 2\rho\cos(\omega_{\text{D},j}T/\sqrt{2})q^{-1} + \rho^2 q^{-2}} = \frac{1}{D(q^{-1}, \omega_{\text{D},j}T)}\ . \quad (65)$$

According to the discussion of [30], (65) provides a reasonable approximation to Rayleigh fading statistics [16][5] if

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

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

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

$$\boldsymbol{D}(q^{-1})h_t = e_t \quad (66)$$

with a diagonal auto-regression matrix

$$\boldsymbol{D}(q^{-1}) = \text{diag}[\ \boldsymbol{D}_{11}(q^{-1})\ \ \boldsymbol{D}_{12}(q^{-1})\ \ \boldsymbol{D}_{21}(q^{-1})\ \ \boldsymbol{D}_{22}(q^{-1})\ ] \quad (67)$$

$$\boldsymbol{D}_{ij}(q^{-1}) = D(q^{-1}, \omega_{\text{D},j}T)\mathbf{I}_2 \quad (68)$$

and a block-diagonal covariance matrix for e_t

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

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 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.

[5]Stable AR models of order 2-4 are actually often a better approximation of reality than the classical Jakes Rayleigh fading model of [16], which assumes an infinite number of local scatterers equally distributed on a circle, and has infinite peaks in its spectrum.

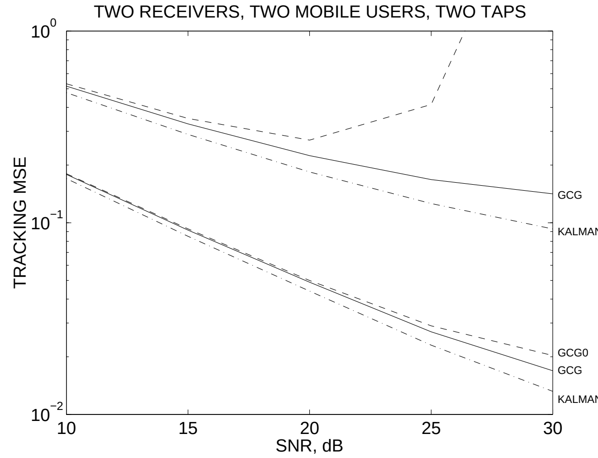

Figure 6: The sum of squared four-step channel tap prediction errors tr $\mathbf{P}_4$ in Example 2 when the first mobile moves at 45km/h while the second mobile has velocity 45km/h (lower curves) and 225km/h (upper curves). Results for one-shot designs assuming $\eta_t = \varphi_t v_t$ (dashed), full iterative design (solid) and the Kalman estimator (dash-dotted).

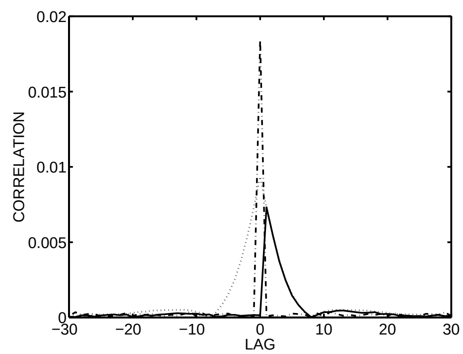

Figure 7: Absolute values of correlations in Example 2 at SNR 30dB, $\omega_{\text{D},2}T = 0.10$, estimated based on 10000 data. Correlation functions for element (3,3) of $E(\eta_t\eta^*_{t+\tau})$ (dash-dotted), $E(\eta_t\tilde{h}^*_{t+\tau})$ (solid) and $E(\tilde{h}_{t|t-1}\tilde{h}^*_{t+\tau|t+\tau-1})$ (dotted).

Prediction estimates of the channel taps are required in equalizers [30]. We here design four-step prediction estimators ($k = 4$) according to the iterative scheme outlined above for the two cases $\omega_{\text{D},2}T = 0.02$ and $\omega_{\text{D},2}T = 0.10$, and for an SNR per channel in the range 10dB-30dB. Figure 6 displays the tracking MSE tr $\mathbf{P}_4$ for two designs: a non-iterative design assuming slow time-variations (dashed curves) and the full iterative design (solid curves), using simulations of (66) 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 low SNR's: in the upper curve of Figure 6, the use of $\eta_t = \varphi_t v_t$ at SNR 30dB results in in-stability. 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 7, the gradient noise is white. This is predicted by the theory of [28] for two-tap FIR channels. 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 most difficult design point, SNR 30dB and $\omega_{D,2}T = 0.10$ (solid line for lags ≤ 0).

In Table 1, 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 [29, 30], exponentially windowed RLS and an LMS estimator. We also compare their computational complexity, as measured by the required number of real-valued multiplication-accumulation operations per sample[6].

The Kalman predictor is designed based on a state-space realization of (66) with 16 complex-valued states and with (1) as the measurement equation. The Wiener LMS algorithm (53) is not equipped to handle elements of h_t with differing dynamics. However, it was in [30] found to be robust against variations of the Doppler frequency of fading models, if $\omega_{D,}T$ is set at the high end of its uncertainty range, and if an integrator is included in the model (AR_2I-modelling). We thus design (53) for a model (52) with $C = 1$ and $D = D(q^{-1}, \omega_{D,2}T)(1-q^{-1})$, with $D(q^{-1}, \omega_{D,2}T)$ from (65).

From Table 1, 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 $\omega_{D,2}T = 0.10$, this is due to the better tuning of GCG to differing tap dynamics. At $\omega_{D,2}T = 0.02$, 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	$\omega_{D,2}T$	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.	5440	416	272	1564	132

Table 1: 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 the general linear time-invariant models (21), (22). 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 white or autoregressive regressors, their complexity grows linearly with the number of estimated parameters.

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 γ.

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. 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.

An 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],[39] is adopted. Consider the filtering problem depicted in Figure 3. 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, the optimal ξ_t must be zero and the covariance $\mathbf{P}_k$ obtained with the unperturbed estimator is minimal.

[6]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. We utilize the diagonal structure of $\boldsymbol{D}(q^{-1})$ and $\mathbf{R}_\eta$ and the block-diagonal structure of $\mathbf{R}_e$.

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

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

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 (19), 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. [34], to convert the orthogonality requirement of $\tilde{h}_{t+k|t}$ and ξ_t into the frequency domain relation

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

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

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

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

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

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

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

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],[36]. Let $\boldsymbol{Q}_k^o, \boldsymbol{L}_{k*}^o$ be one solution pair. Every solution to (26) can then be expressed as

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

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 (19) 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

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

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

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 in the transfer function matrix, i.e. if

$$\boldsymbol{\beta} - q^{-k}\boldsymbol{Q}_kN = D_u\boldsymbol{X}_k \qquad (A.7)$$

for some polynomial matrix $\boldsymbol{X}_k$. This condition is verified by right-multiplying the left-hand side of (A.7) by $\boldsymbol{\beta}_*$ (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,

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

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

Appendix B: Weighted Tracking MSE Criteria and the Excess Mean Square Error

Particular frequency ranges or linear combinations of the tracking error can be emphasized by introducing

$$\tilde{h}_{t+k|t}^w = \boldsymbol{\mathcal{W}}(q^{-1})\tilde{h}_{t+k|t} \ , \qquad (B.1)$$

where, in general, $\mathcal{W}(q^{-1})$ is a causal and stably invertible rational matrix. The criterion to be minimized is then given by

$$\lim_{t\to\infty} \mathrm{E}\, \tilde{h}^{w}_{t+k|t}\tilde{h}^{w*}_{t+k|t} \qquad \text{(B.2)}$$

with a minimizing solution readily obtained by slight modifications of Theorem 1. Let $\mathcal{W}(q^{-1})$ be parameterized in common denominator form

$$\mathcal{W}(q^{-1}) = \frac{1}{U(q^{-1})}\boldsymbol{V}(q^{-1}) \ , \qquad \text{(B.3)}$$

where $U(q^{-1})$ is a scalar polynomial and $\boldsymbol{V}(q^{-1})$ is a stably and causally invertible polynomial matrix. Introduce the stable polynomial matrices $\check{\boldsymbol{D}}_s$ and $\check{\boldsymbol{V}}_s$ via the coprime factorization

$$\boldsymbol{V}\boldsymbol{D}_s^{-1} = \check{\boldsymbol{D}}_s^{-1}\check{\boldsymbol{V}} \ . \qquad \text{(B.4)}$$

The solution for the weighted criterion is then obtained by substituting equations (24) and (26) of Theorem 1 by

$$\mathcal{L}_k^{opt} = \boldsymbol{V}^{-1}\check{\boldsymbol{D}}_s^{-1}\boldsymbol{Q}_k\boldsymbol{\beta}^{-1}N\boldsymbol{D}_s\mathbf{R}^{-1} \qquad \text{(B.5)}$$

and

$$q^k\check{\boldsymbol{V}}\boldsymbol{C}\,\mathbf{R}_e\boldsymbol{C}_*N_* = \boldsymbol{Q}_k\boldsymbol{\beta}_* + qUD_u\check{\boldsymbol{D}}_s\boldsymbol{L}_{k*} \ . \qquad \text{(B.6)}$$

When performing system identification, it is often of interest to minimize the mean squared prediction error

$$\begin{aligned}\mathrm{E}\,\|\varepsilon_t\|_2^2 &= \mathrm{E}\,\|y_t - \varphi_t^*\hat{h}_{t|t-1}\|_2^2 = \mathrm{E}\,\|\varphi_t^*\tilde{h}_{t|t-1} + v_t\|_2^2 \\ &= \mathrm{tr}\,\mathbf{R}_v + \mathrm{tr}\,\mathrm{E}\,\varphi_t\varphi_t^*\tilde{h}_{t|t-1}\tilde{h}^*_{t|t-1} \ , \qquad \text{(B.7)}\end{aligned}$$

where we assumed v_t and $\varphi_t^*\tilde{h}_{t|t-1}$ to be uncorrelated. The last term of (B.7) is known as the *excess mean square error.* Evidently, (B.7) is minimized by minimizing the excess MSE. However, it is very difficult to minimize this term as it stands, unless φ_t^* and $\tilde{h}_{t|t-1}$ are independent. Under that assumption, we should minimize

$$\mathrm{tr}\,\mathrm{E}\,(\varphi_t\varphi_t^*)\mathrm{E}\,(\tilde{h}_{t|t-1}\tilde{h}^*_{t|t-1}) = \mathrm{tr}\,\mathbf{R}\mathbf{P}_1 = \mathrm{tr}\,\mathbf{R}^{1/2}\mathbf{P}_1\mathbf{R}^{1/2} \ . \qquad \text{(B.8)}$$

By introducing (28), the criterion (B.8) is minimized by minimizing (B.2) □

Appendix C: 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,

$$\begin{aligned}\boldsymbol{Q}_{k+1}(q^{-1}) &= q\left(\boldsymbol{Q}_k(q^{-1}) - \boldsymbol{D}(q^{-1})\mathbf{Q}_0^k\right) \qquad \text{(C.1)} \\ \boldsymbol{L}_{k+1*}(q) &= q\boldsymbol{L}_{k*}(q) + \mathbf{Q}_0^k\boldsymbol{\beta}_*(q) \qquad \text{(C.2)}\end{aligned}$$

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

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

$$\boldsymbol{L}_{k-1*}(q) = q^{-1}\left(\boldsymbol{L}_{k*}(q) - \mathbf{L}_0^{k*}(\beta_0^*)^{-1}\boldsymbol{\beta}_*(q)\right) \qquad \text{(C.4)}$$

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

$$q^{k+1}\boldsymbol{C}\,\mathbf{R}_e\boldsymbol{C}_*N_* = \boldsymbol{Q}_{k+1}\boldsymbol{\beta}_* + q\boldsymbol{D}\,\boldsymbol{L}_{k+1*} \ . \qquad \text{(C.5)}$$

Multiplying both sides of (C.5) by q^{-1} and using the assumed relation (C.1) 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 (C.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 (C.3) and (C.4) are verified in the same way, with $k-1$ substituted for $k+1$ in (C.5), multiplying by q and inserting (C.3) and (C.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 (C.1) and of (C.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 $\boldsymbol{L}_{k-1*}(q)$ □

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

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

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

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

[7] 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.

[8] 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.

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

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

[11] 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.

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

[13] 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.

[14] D. Henrion *et. al.* "Polynomial toolbox, version 2," See http://www.math.utwente.nl/polbox.

[15] 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.

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

[20] 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.

[21] V. Kŭcera, "Factorization of rational spectral matrices: a survey of methods," IEE Control 91, Edinburgh, pp. 1074-1078, 1991.

[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] C. P. Kwong, "Control-theoretic design of the LMS and the sign algorithms in nonstationary environments," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, pp. 253-259, 1990.

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

[25] 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.

[26] 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.

[27] 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

[28] 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". Submitted.
www.signal.uu.se/Publications/abstracts/r002.html

[29] L. Lindbom, M. Sternad and A. Ahlén, "Tracking of time-varying mobile radio channels. Part I: The Wiener LMS algorithm". Submitted.
www.signal.uu.se/Publications/abstracts/r003.html

[30] L. Lindbom, A. Ahlén, M. Sternad and M. Falkenström, "Tracking of time-varying mobile radio channels. Part II: A case study on D-AMPS 1900 channels". Submitted.
www.signal.uu.se/Publications/abstracts/r004.html

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

[32] 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.

[33] 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.

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

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

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

[37] 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.

[38] 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.

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

[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 Processsing*, 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.