ITERATIVE WIENER DESIGN OF ADAPTATION LAWS WITH CONSTANT GAINS

Anders Ahlén, Mikael Sternad and Lars Lindbom

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

Abstract: We present a method for optimizing adaptation laws that are generalizations of the LMS algorithm. Time-varying parameters of linear regression models are estimated in situations where the regressors are stationary or have slowly time-varying properties. The parameter variations are modeled as ARIMA-processes and the aim is to use such prior information to provide high performance filtering, prediction or fixed lag smoothing estimates for arbitrary lags. The method is based on a novel signal transformation that recasts the algorithm design problem into a Wiener design.

1. INTRODUCTION

Adaptation algorithms that estimate time-varying parameters of linear regression models are fundamental tools in signal processing, control and digital communication [1]. When the statistics of the parameter variations are known, Kalman estimators are the optimal linear algorithms. However, their computational complexity is sometimes deemed unacceptable. Motivated by the required low complexity of channel estimators in mobile radio systems, we have proposed a class of adaptation laws that attain close to the optimal Kalman performance, at a computational complexity close to that of LMS [2, 3, 4]. An early version has successfully been used on D-AMPS 1900 channels [5, 6], and a case study on this application is presented in [7].

Design of a related class of constant-gain algorithms has been investigated by Benveniste and co-workers [8] for slowly varying regression parameters. We here outline a method that is effective also for tracking fast variations.

Notation: Here, $R(q^{-1})$, $\boldsymbol{R}(q^{-1})$ and $\mathcal{R}(q^{-1})$ denote polynomials, polynomial matrices and causal rational matrices, respectively, in the backward shift operator q^{-1}.

2. OUTLINE OF THE PROBLEM

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

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

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

$$\mathbf{R} \triangleq \mathrm{E}\,\varphi_t \varphi_t^* \tag{2}$$

is nonsingular. Furthermore, $\mathbf{R}$ is here assumed constant and known, while in practice it may be slowly time-varying. The aim is to estimate the time varying parameter vector

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

when the order n_h is known. Models describing the variation of h_t are sometimes called *hypermodels* [8]. We will here consider linear time-invariant stochastic models

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

where e_t is white noise with covariance matrix $\mathbf{R}_e$ and where $\mathcal{H}(q^{-1})$ is an $n_h|n_h$ matrix of stable or marginally stable transfer operators. Let the tracking error be denoted by

$$\tilde{h}_{t+k|t} \triangleq h_{t+k} - \hat{h}_{t+k|t} \tag{5}$$

where $\hat{h}_{t+k|t}$ is an estimate of h_{t+k} obtained at time t by filtering ($k = 0$), prediction ($k > 0$) or fixed lag smoothing ($k < 0$). Kalman estimators, based on (1) and on state-space realizations of (4), are the linear estimators that minimize the error covariance matrix

$$\mathbf{P}_{k,t} \triangleq \mathrm{E}\,\tilde{h}_{t+k|t}\tilde{h}^*_{t+k|t} \quad , \tag{6}$$

where the expectation is with respect to e_t and v_t. Since φ_t^* is time-varying, the Kalman gains will not converge to a steady state solution, so Riccati updates are required.

We here consider a class of adaptation laws obtained by using pre-designed linear time-invariant filters $\boldsymbol{\mathcal{M}}_k(q^{-1})$

$$\varepsilon_t = y_t - \varphi_t^* \hat{h}_{t|t-1} \tag{7}$$

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

that operate on $\varphi_t \varepsilon_t$, which is the negative instantaneous gradient of $|\varepsilon_t|^2$ with respect to $\hat{h}_{t|t-1}$. The LMS algorithm

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

where $\mu > 0$ is a scalar gain, constitutes a simple special case of the general structure (7),(8).

The rational matrix $\mathcal{M}_k$ can be selected to asymptotically minimize the tracking error covariance matrix (6) under various constraints and assumptions. Note that for any desired k, a one-step predictor $\mathcal{M}_1(q^{-1})$ must also be designed, due to the presence of $\hat{h}_{t|t-1}$ in (7).

3. THE LOOP TRANSFORMATION

The algorithm (7),(8) can be expressed as a stable and causal filter, denoted the *learning filter* $\mathcal{L}_k(q^{-1})$, that operates on a signal vector

$$f_t \triangleq \varphi_t\varepsilon_t + \mathbf{R}\hat{h}_{t|t-1} \quad , \tag{10}$$

since (7),(8) give

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

$$\hat{h}_{t+k|t} = \mathcal{M}_k(q^{-1})(\mathbf{I}+q^{-1}\mathbf{R}\mathcal{M}_1(q^{-1}))^{-1} f_t \triangleq \mathcal{L}_k(q^{-1})f_t \, . \tag{11}$$

Consider $\varphi_t\varepsilon_t$ and insert (1), describing y_t, into (7) to obtain

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

Adding and subtracting $\mathbf{R}\tilde{h}_{t|t-1}$ on the right-hand side of (12) gives

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

Define

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

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

which are called the *autocorrelation matrix noise* [9] and the *gradient noise*, respectively. The signal f_t, can then, from (10), (13), (14) and (15), be expressed as

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

see Fig. 2. The design of our adaptation law (7),(8) has now been transformed into a *Wiener filter design* for $\mathcal{L}_k(q^{-1})$, where η_t plays the role of noise, see Fig. 1.

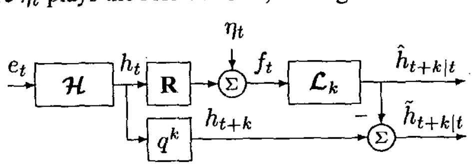

Fig. 1. The filter design problem. The vector h_{t+k} is to be estimated from f_t, such that the steady state tracking error covariance matrix of the parameter error $\tilde{h}_{t+k|t}$ is minimized.

The gradient noise η_t is affected by the term $Z_t\tilde{h}_{t|t-1}$, here called the *feedback noise*. It is shown in [4] that the feedback noise is negligible either when h_t has small increments or when the noise v_t has high variance. Such situations are denoted "slow variations" [4],[10]. The optimal learning filter will then operate in open loop, with $\eta_t \approx \varphi_t v_t$. Stability and convergence in MSE is then guaranteed by stability of the learning filter, which follows directly from a Wiener design. (While the learning filter $\mathcal{L}_k(q^{-1})$ must be stable, the filter $\mathcal{M}_1(q^{-1})$ in (11) need not be stable, since it works within the feedback loop of Fig. 2.) An iterative design must bee performed when $Z_t\tilde{h}_{t|t-1}$ cannot be neglected, see Section 5.

4. LEARNING FILTER OPTIMIZATION

The transfer operator $\mathcal{L}_k(q^{-1})$ can be adjusted to minimize (6) for $t \to \infty$ when $\mathcal{H}(q^{-1})$ in (4) and the properties of η_t are given. The learning filter is here designed under the constraint of stability, and under the following assumptions.

Assumption A1: The sequence $\{\varphi_t^*\}$ is stationary and known up to time t, with $\mathbf{R}$ known and nonsingular □

Assumption A2: The gradient noise η_t is white and stationary with zero mean and covariance matrix $\mathbf{R}_\eta$. It is uncorrelated with h_{t-i} and with $\hat{h}_{t-i|t-i-1}, i \geq 0$ □

Assumption A3: The linear regression coefficients are described by a stochastic vector ARIMA process

$$\boldsymbol{D}(q^{-1})h_t = \boldsymbol{C}(q^{-1})e_t \quad , \tag{17}$$

with $\mathbf{R}_e = \mathrm{E}\, e_t e_t^*$ nonsingular, where $\boldsymbol{D} = D_u\boldsymbol{D}_s$. Here $\boldsymbol{C}$ and $\boldsymbol{D}_s$ are monic and stably invertible, while the polynomial D_u has zeros on the stability limit (unit circle) □

Under Assumptions A1-A3, the (generalized) innovations model of $f_t = \mathbf{R}h_t + \eta_t$ can be expressed as

$$f_t = \mathbf{R}\boldsymbol{D}^{-1}\beta\epsilon_t \;\Leftrightarrow\; \epsilon_t = \beta^{-1}\boldsymbol{D}\mathbf{R}^{-1}f_t \tag{18}$$

where the polynomial matrix $\beta(q^{-1})$ is the stably invertible spectral factor and ϵ_t is the white zero mean innovation sequence with unit covariance matrix. By defining the signal

$$\bar{\epsilon}_t \triangleq \frac{1}{D_u(q^{-1})}\epsilon_t = \beta^{-1}\boldsymbol{D}_s\mathbf{R}^{-1}f_t \quad , \tag{19}$$

a Wiener designed adaptation law can be realized as in Fig. 3, in which $\boldsymbol{D}_s^{-1}\boldsymbol{Q}_k$ represents the causal factor of the realizable MIMO Wiener solution. By comparing Fig. 1 and Fig. 3, it follows that the optimized causal learning filter is

$$\mathcal{L}_k^{opt} = \boldsymbol{D}_s^{-1}\boldsymbol{Q}_k\beta^{-1}\boldsymbol{D}_s\mathbf{R}^{-1} \, . \tag{20}$$

The polynomial matrix $\boldsymbol{Q}_k(q^{-1})$ can be obtained from closed-form expressions [3]. In particular, $\boldsymbol{Q}_1 = q(\beta - \boldsymbol{D}\beta_0)$, with β_0 being the leading coefficient matrix of $\beta(q^{-1})$. With this expression and (11),(20), the Wiener optimized filter matrix $\mathcal{M}_k(q^{-1})$ in (8) can be shown [3] to be given by

$$\mathcal{M}_k^{opt} = \boldsymbol{D}^{-1}\boldsymbol{Q}_k\beta_0^{-1}\mathbf{R}^{-1} \, . \tag{21}$$

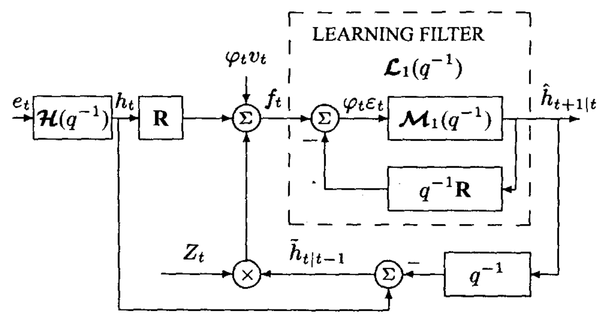

Fig. 2. The prediction learning filter operates in open loop for slow variations, when $Z_t\tilde{h}_{t|t-1}$ can be neglected. For fast variations, the feedback noise $Z_t\tilde{h}_{t|t-1}$ has to be taken into account.

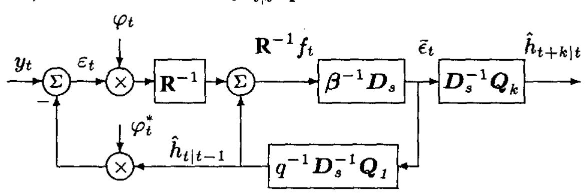

Fig. 3. The Wiener optimized tracking algorithm.

5. ITERATIVE WIENER DESIGN

For slow time-variations, the feedback noise is by definition negligible [4], so we may perform a one-shot design using $\eta_t = \varphi_t v_t$. Otherwise, the properties of η_t depend on $\mathcal{L}_1(q^{-1})$ via (15). The multiplication by Z_t in (15) acts as a scrambler, which for FIR models with white inputs will reduce the correlation between the feedback noise and $\tilde{h}_{t|t-1}$, so Assumption A2 still holds approximately for white noise v_t. The open-loop design can then be performed iteratively. We proceed as follows:

1. Perform a one-step predictor design for slow time-variations, i.e. use $\mathbf{R}_\eta = \mathrm{E}\,\varphi_t v_t v_t^* \varphi_t^*$ to design $\mathcal{L}_1(q^{-1})$. Verify that the closed loop around $\mathcal{L}_1(q^{-1})$ of Fig. 2 is stable. If not, scale up $\mathbf{R}_\eta$ to decrease the gain of $\mathcal{L}_1(q^{-1})$.

2. Based on a simulation of φ_t, v_t, h_t and of $\hat{h}_{t|t-1}$, estimate $\mathbf{R}_\eta$ from $\hat{\eta}_t = \varphi_t\varepsilon_t - \mathbf{R}(h_t - \hat{h}_{t|t-1})$ (see (10),(16)), by using sample averages over $\hat{\eta}_t$.

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

Repeat steps 2. and 3. until the difference in $\hat{h}_{t+1|t}$ becomes small. Then, design $\mathcal{L}_k(q^{-1})$ for the desired k.

Generalizations to colored gradient noise and uncertain hypermodels exist, see [2, 3].

It will be possible to find an initial stable solution under mild conditions. If $\mathcal{H}$ is stable, then $\mathcal{L}_1(\omega) \to 0\ \forall\,\omega$ when the assumed noise power is increased. If Z_t has bounded elements, then the small gain theorem [11] implies that the closed loop of Fig. 2 can be stabilized by assuming a sufficiently high noise power in the design of $\mathcal{L}_1(q^{-1})$.

Example. 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. A receiver with two diversity branches detects both users u_t^1 and u_t^2 simultaneously. Two-tap fading channels are assumed so the model can then be expressed by (1) with

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

and

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

where the complex channel taps $b_{0,t}^{ij}$ and $b_{1,t}^{ij}$ associate with the mobile user j and the receiver branch i. The transmitted symbols $\{u_t^j\}$, here assumed to be known by the receiver, are white QPSK symbols with zero mean and $\mathbf{R} = \mathbf{I}_8$. The zero mean noise v_t is white with variance $\sigma_v^2\mathbf{I}_2$.

Second order statistics of fading radio channel taps can be well approximated by autoregressive models. These are here assumed to be of second order and described by

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

where $\omega_{\mathrm{D},j}$ is the maximum Doppler angular frequency for mobile j and T is the sampling time. (This model provides a reasonable approximation to classical Jakes Rayleigh fading statistics if the pole radius is selected as $\rho = 0.999 - 0.1\omega_{\mathrm{D},j}T$ for $\omega_{\mathrm{D},j}T \le 0.10$.)

We investigate $\omega_{\mathrm{D},j} \in [0.02\ 0.10]$, which approximately correspond to vehicle speeds from 45km/h to 225km/h in symbol spaced sampled ANSI-136 1900MHz systems.

If the two vehicles have different velocities, yielding $\omega_{\mathrm{D},1}$ and $\omega_{\mathrm{D},2}$ respectively, and if the channels to different receivers are assumed uncorrelated, then $\boldsymbol{C} = \mathbf{I}_8$ and $\boldsymbol{D} = \mathrm{diag}[\boldsymbol{D}_{11}\boldsymbol{D}_{12}\boldsymbol{D}_{21}\boldsymbol{D}_{22}]$ in the hypermodel (17), with diagonal blocks $\boldsymbol{D}_{ij}(q^{-1}) = D(q^{-1}, \omega_{\mathrm{D},j}T)\mathbf{I}_2$.

The receiver is assumed to be synchronized to mobile 1, resulting in zero correlation between taps from mobile 1. We assume correlation 0.8 in the taps from mobile 2 and set the SNR equal for both users. This determines $\mathbf{R}_e$ which becomes 2×2-block diagonal. The velocity of mobile 1 is fixed to 45km/h, while the velocity of mobile 2 is varied.

Four-step prediction estimators ($k = 4$) are appropriate in Viterbi detectors [7]. They are designed 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. Fig. 4 displays the tracking MSE $\mathrm{tr}\,\mathbf{P}_4$ for designs assuming slow time-variations (dashed curves) and full iterative designs (solid curves), measured from simulations of length 10000. A single iteration was sufficient at all design points except at 30dB in the upper curves.

The performance of the constant-gain tracker is close to

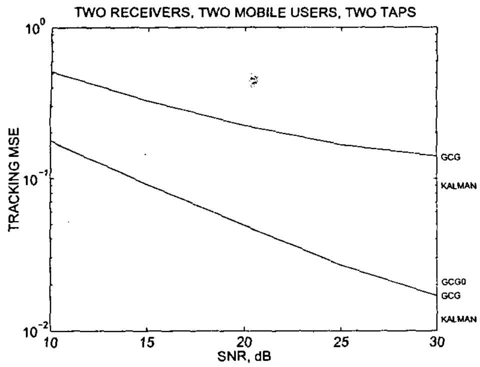

Fig. 4. The sum of squared four-step channel tap prediction errors $\mathrm{tr}\,\mathbf{P}_4$ when mobile 1 moves at 45km/h while mobile 2 has velocity 45km/h (lower curves) and 225km/h (upper curves). Results for one-shot designs assuming $\eta_t = \varphi_t v_t$ (dashed), full iterative design (solid) and the Kalman 4-step predictor (dash-dotted).

that of the Kalman estimator[1] 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 Fig. 4, the use of $\eta_t = \varphi_t v_t$ at SNR 30dB results in instability. A design theory based on slow time-variations [8] 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.

In Table 1, we compare the tracking MSE and the computational complexity[2] for Kalman predictors, for the Wiener design, denoted the general constant gain algorithm (GCG), for exponentially windowed RLS and for LMS estimators. The GCG Wiener design attains nearly the same performance as the Kalman estimator, at much lower complexity. Note that the use of RLS would in this example result in *both* bad performance and a high computational load.

SNR	$\omega_{D,2}T$	Kalman	GCG	RLS	LMS
10	0.10	0.477	0.516	1.43	1.58
30	0.10	0.093	0.142	0.82	1.00
10	0.02	0.170	0.179	0.33	0.413
30	0.02	0.013	0.017	0.077	0.115
	#mult.	5440	416	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, obtained by optimized Kalman tracking, the general constant gain algorithm (GCG), RLS and LMS adaptation algorithms.

[1] The Kalman predictor is designed based on a state-space realization of (17) with 16 complex-valued states with (1) as the measurement equation.

[2] Measured as the required number of real-valued multiplication-accumulation operations per sample. We utilize the diagonal structure of $\boldsymbol{D}(q^{-1})$ and $\mathbf{R}_\eta$ and the block-diagonal structure of $\mathbf{R}_e$.

6. CONCLUSIONS

We have outlined the design of a class of adaptation laws which are generalizations of LMS. For details, see [2, 3, 4]. Compared to Kalman tracking of linear regression parameters, a main advantage with the proposed class of algoritms is their lower computational complexity. Another advantage is that it becomes more straightforward to design fixed-lag smoothing estimators. A disadvantage is that our Wiener design is a steady-state solution, which could lead to worse transient properties than for a Kalman estimator.

7. REFERENCES

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

[2] L.Lindbom, *A Wiener filtering approach to the design of tracking algorithms.* PhD Thesis, Dept. of Technology, Uppsala University, Sweden, 1995.

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

[4] 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". Subm. www.signal.uu.se/Publications/abstracts/r002.html

[5] 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.

[6] 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.

[7] M.Sternad, L.Lindbom and A.Ahlén "Tracking of time-varying mobile radio channels with WLMS Algorithms: A case study on D-AMPS 1900 channels," *IEEE Vehicular Technology Conf. VTC2000-Spring*, Tokyo, Japan, May 15-18, 2000, pp. 2507-2511. To appear in *IEEE Trans. Com.* www.signal.uu.se/Publications/abstracts/r004.html

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

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

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

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