ADAPTATION WITH CONSTANT GAINS: ANALYSIS FOR SLOW VARIATIONS

Lars Lindbom, Mikael Sternad and Anders Ahlén

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

Abstract: Adaptation laws with constant gains, that adjust parameters of linear regression models, are investigated. The class of algorithms includes LMS as its simplest member. Closed-form expressions for the tracking MSE are obtained for parameters described by ARIMA processes. A key element of the analysis is that adaptation algorithms are expressed as linear time-invariant filters, here called learning filters, that work in open loop for slow parameter variations. Performance analysis can then easily be performed for slow variations, and stability is assured by stability of these learning filters.

1. INTRODUCTION

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

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

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. The parameter vector

$$h_t = (h_{0,t} \ \dots \ h_{n_h-1,t})^T \ , \tag{2}$$

with n_h known, is to be estimated.

We shall here investigate a class of linear time-invariant estimators

$$\hat{h}_{t+k|t} = \boldsymbol{\mathcal{M}}_k(q^{-1})\varphi_t\varepsilon_t \ , \tag{3}$$

operating on

$$\varphi_t\varepsilon_t = \varphi_t(y_t - \varphi_t^*\hat{h}_{t|t-1}) \ , \tag{4}$$

the negative gradient of $|\varepsilon_t|^2$. Above, $\hat{h}_{t+k|t}$ is an estimate of h_{t+k} obtained at time t by filtering ($k=0$), prediction ($k>0$) or fixed lag smoothing ($k<0$). Most known linear algorithms with constant gains fit into this structure, e.g. the LMS algorithm with $\boldsymbol{\mathcal{M}}_1(q^{-1}) = (\mu/(1-q^{-1}))\,\mathbf{I}$.

Wiener methods for adjusting $\boldsymbol{\mathcal{M}}_k(q^{-1})$ have been developed in [1],[2]. In this presentation, we outline a performance analysis of adaptation laws (3) that track slowly varying parameter vectors h_{t+k}. The resulting estimation error $\tilde{h}_{t+k|t} = h_{t+k} - \hat{h}_{t+k|t}$ is investigated and closed-form expressions for the steady-state parameter tracking error covariance matrix

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

that hold under the following assumption, will be presented.

Assumption 1: The parameter vector h_t is well described by a linear time-invariant vector ARIMA process and the noise v_t is stationary and zero mean, while φ_t^*, with known dimension, is stationary with zero mean and nonsingular covariance matrix $\mathbf{R} = \mathrm{E}\,\varphi_t\varphi_t^*$. Moreover, h_t, v_t, and φ_t^* are mutually independent □

We do not assume consecutive φ_t^* to be independent. However, the assumption that the regressors φ_t^* are independent of the parameters h_t and of the noise v_t excludes the analysis of AR and ARX models.

Notation: Here, $R(q^{-1})$, $\boldsymbol{R}(q^{-1})$ and $\boldsymbol{\mathcal{R}}(q^{-1})$, denote polynomials, polynomial matrices and causal rational matrices, respectively in the backward shift operator q^{-1} ($q^{-1}y_t = y_{t-1}$), while $\boldsymbol{R}_*(q)$ is the conjugate-transpose of the polynomial matrix $\boldsymbol{R}(q^{-1})$.

2. THE LEARNING FILTER

The basis for the analysis is that all linear time-invariant estimators (3) can alternatively be expressed as

$$\hat{h}_{t+k|t} = \boldsymbol{\mathcal{L}}_k(q^{-1})f_t = \sum_{i=0}^{\infty} \mathbf{L}_i^k f_{t-i} \ , \tag{6}$$

$$\boldsymbol{\mathcal{L}}_k(q^{-1}) \triangleq \boldsymbol{\mathcal{M}}_k(q^{-1})(\mathbf{I} + q^{-1}\mathbf{R}\boldsymbol{\mathcal{M}}_1(q^{-1}))^{-1} \ . \tag{7}$$

Here $\boldsymbol{\mathcal{L}}_k(q^{-1})$, called the *learning filter*, must be stable and causal. It operates on a signal vector f_t, obtained by inserting (1) into $\varphi_t\varepsilon_t$ and then adding and subtracting $\mathbf{R}\hat{h}_{t|t-1}$ on the right-hand side of (4), giving

$$\begin{aligned} \varphi_t\varepsilon_t &= \mathbf{R}h_t - \mathbf{R}\hat{h}_{t|t-1} + (\varphi_t\varphi_t^* - \mathbf{R})\tilde{h}_{t|t-1} + \varphi_t v_t \\ &\triangleq f_t - \mathbf{R}\hat{h}_{t|t-1} \ . \end{aligned} \tag{8}$$

Thus,

$$f_t = \mathbf{R}\hat{h}_{t|t-1} + \varphi_t\varepsilon_t \triangleq \mathbf{R}h_t + \eta_t \ . \tag{9}$$

0-7803-7041-4/01/$10.00 ©2001 IEEE

The use of $\varphi_t\varepsilon_t = f_t - \mathbf{R}q^{-1}\mathcal{M}_1\varphi_t\varepsilon_t$ from (9) in (3) gives (6),(7). The signal f_t can be regarded as a measurement consisting of a rotated parameter vector $\mathbf{R}h_t$ disturbed by an additive noise η_t, called the *gradient noise*. If the term η_t is regarded as a zero mean stationary additive noise, then stable and causal filters (6) can be designed analytically to minimize the tracking error covariance matrix $\mathbf{P}_k$.

Optimal Wiener design, presented in [1], results in stable learning filters $\mathcal{L}_k(q^{-1})$. For adaptation laws obtained by other means, stability must be verified separately. For the LMS algorithm,

$$(\mathbf{I} - q^{-1}\mathbf{I})\hat{h}_{t+1|t} = \mu\varphi_t\varepsilon_t \ , \tag{10}$$

the learning filter

$$\mathcal{L}_1(q^{-1}) = (\mathbf{I} - (\mathbf{I} - \mu\mathbf{R})q^{-1})^{-1}\mu \tag{11}$$

is obtained by inserting $\varphi_t\varepsilon_t = f_t - q^{-1}\mathbf{R}\hat{h}_{t+1|t}$ from (9) in (10), or directly by inserting $\mathcal{M}_1(q^{-1}) = (\mu/(1-q^{-1}))\,\mathbf{I}$ in (7). The stability requirement on $\mathcal{L}_1(q^{-1})$ in (11) corresponds to the classical condition for convergence in the mean [3]

$$0 < \mu < \frac{2}{\lambda_{\max}} \ , \tag{12}$$

where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{R}$. This readily follows by an eigenvalue decomposition of $\mathbf{R}$ in (11).

In open loop, stability and bounded estimation errors are guaranteed by stability of the learning filter. However, since (9) and (8) imply that

$$\eta_t = Z_t\tilde{h}_{t|t-1} + \varphi_t v_t \ , \tag{13}$$

where Z_t is the zero mean matrix

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

the gradient noise η_t will contain a time-varying feedback term $Z_t\tilde{h}_{t|t-1}$ from old parameter errors, here called the *feedback noise*. This feedback involves the one-step prediction learning filter $\mathcal{L}_1(q^{-1})$, see Fig. 1. The gradient noise η_t will therefore be influenced by the design of the estimator and the feedback may in a general case cause instability.

A sufficient but conservative condition for stability of the feedback is provided by the small gain theorem [4]: If $\mathcal{L}_1(q^{-1})Z_t$ is causal and L_p-stable, stability is preserved if

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

In [5], less conservative conditions are derived for FIR systems with white input data.

3. PERFORMANCE ANALYSIS

By (6), (9) and (13), the tracking error can be expressed as

$$\tilde{h}_{t+k|t} = (\mathbf{I} - q^{-k}\mathcal{L}_k\mathbf{R})h_{t+k} - \mathcal{L}_k\varphi_t v_t - \mathcal{L}_k Z_t\tilde{h}_{t|t-1} \ . \tag{15}$$

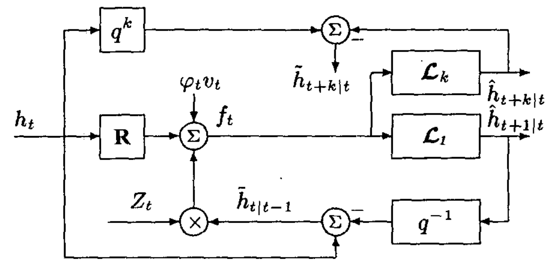

Fig. 1. The feedback loop around $\mathcal{L}_1(q^{-1})$ via the feedback noise $Z_t\tilde{h}_{t|t-1}$ may significantly affect the fictitious measurement f_t and may also cause instability. The variations of h_t will be regarded as slow when this feedback can be neglected.

Thus, three terms affect the tracking error: The *lag error* $(\mathbf{I} - q^{-k}\mathcal{L}_k\mathbf{R})h_{t+k}$, an error term $\mathcal{L}_k\varphi_t v_t$ caused by *measurement noise*, and a feedback noise term $\mathcal{L}_k Z_t\tilde{h}_{t|t-1}$ influenced by *old parameter tracking errors*.

If $\tilde{h}_{t+k|t}$ is stationary, then the error covariance matrix (5) will under Assumption 1 be given by

$$\mathbf{P}_k = \lim_{t\to\infty}\left(\mathbf{V}^k_{h,t} + \mathbf{V}^k_{\varphi v,t} + \mathbf{V}^k_{Z\tilde{h},t} + \mathbf{V}^k_{hZ\tilde{h},t} + \mathbf{V}^k_{\varphi vZ\tilde{h},t}\right) \tag{16}$$

where

$$\mathbf{V}^k_{h,t} = \mathrm{E}\,(\mathbf{I} - q^{-k}\mathcal{L}_k\mathbf{R})h_{t+k}\left((\mathbf{I} - q^{-k}\mathcal{L}_k\mathbf{R})h_{t+k}\right)^* \tag{17}$$

$$\mathbf{V}^k_{\varphi v,t} = \mathrm{E}\,(\mathcal{L}_k\varphi_t v_t)(\mathcal{L}_k\varphi_t v_t)^* \tag{18}$$

$$\mathbf{V}^k_{Z\tilde{h},t} = \mathrm{E}\,(\mathcal{L}_k Z_t\tilde{h}_{t|t-1})(\mathcal{L}_k Z_t\tilde{h}_{t|t-1})^* \ . \tag{19}$$

The two last terms in (16) are due to correlations between the feedback noise, the lag error and $\varphi_t v_t$, respectively.

Slow variations and slow adaptation are a common assumption in analysis of adaptation algorithms [6, 7, 9, 10]. This concept can be quantified in various ways. Sometimes, indicators referred to as "the degree of nonstationarity" (DNS) [8],[11] are introduced. In the book [12] by Macchi, the degree of nonstationarity is defined as

$$\sqrt{\frac{\mathrm{E}\,||\varphi_t^*(h_t - h_{t-1})||_2^2}{\mathrm{E}\,|v_t|^2}} \ . \tag{20}$$

Parameter variations are considered slow if this quantity is always small.

We introduce and motivate a definition that is related to (20) but is more useful in the present formalism:

Definition 1: Regression parameters are regarded as slowly time-varying when the feedback noise $Z_t\tilde{h}_{t|t-1}$ can be neglected in an optimal MSE design, without affecting the tracking error covariances significantly □

For slow parameter variations, we can thus neglect the three last terms in (16) when evaluating the performance.

Let the feedback noise contribution to η_t be neglected[1] and assume that the remaining gradient noise can be represented by a stably invertible vector-valued ARMA process with common stable denominator

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

with ν_t being a zero mean white noise with a nonsingular covariance matrix

$$\mathrm{E}\,\nu_t\nu_t^* = \mathbf{R}_\nu \triangleq \lambda_\nu \bar{\mathbf{R}}_\nu \ . \tag{22}$$

Furthermore, let the assumed vector-valued ARIMA model of h_t be represented in polynomial form with a common stable or marginally stable denominator $D(q^{-1})$

$$h_t = \frac{1}{D(q^{-1})}\boldsymbol{C}(q^{-1})e_t \ , \tag{23}$$

with a nonsingular covariance matrix

$$\mathrm{E}\,e_te_t^* = \mathbf{R}_e \triangleq \lambda_e \bar{\mathbf{R}}_e \ . \tag{24}$$

The design of $\mathcal{L}_k(q^{-1})$ can now be based on Theorem 1 in [1] (with $\boldsymbol{D}(q^{-1}) = D(q^{-1})\mathbf{I}$). By invoking (1),(6),(9), (21)-(24), the MSE-optimal learning filter is then given by

$$\mathcal{L}_k^{opt}(q^{-1}) = \boldsymbol{Q}_k(q^{-1})\beta(q^{-1})^{-1}N(q^{-1})\mathbf{R}^{-1} \ . \tag{25}$$

Here $\beta(q^{-1})$ is the stable solution to the left polynomial matrix spectral factorization

$$\beta\beta_* = \lambda_e \boldsymbol{C}\bar{\mathbf{R}}_e\boldsymbol{C}_* NN_* + \lambda_\nu D\mathbf{R}^{-1}\boldsymbol{M}\bar{\mathbf{R}}_\nu\boldsymbol{M}_*\mathbf{R}^{-1}D_* \tag{26}$$

while $\boldsymbol{Q}_k(q^{-1})$, together with a polynomial matrix $\boldsymbol{L}_{k*}(q)$, is the unique solution to the Diophantine equation

$$\lambda_e q^k \boldsymbol{C}\bar{\mathbf{R}}_e\boldsymbol{C}_* N_* = \boldsymbol{Q}_k\beta_* + q\mathbf{I}D\boldsymbol{L}_{k*} \ . \tag{27}$$

Let $\bar{\mathbf{R}}_e$ and $\bar{\mathbf{R}}_\nu$ be nonsingular and fixed, while the scalars λ_e and λ_ν may vary. For a vanishing parameter-drift-to-noise ratio $\lambda_e/\lambda_\nu \to 0$, the variations will be slow according to (20). We now state that the feedback noise can in fact be neglected in this situation, which occurs either when the parameters h_t vary slowly, or when the noise level is high.

Lemma 1: Let the learning filter $\mathcal{L}_k(q^{-1})$ be obtained by (25)–(27), assuming $\eta_t = \varphi_t v_t$. Under Assumption 1, the relative impact of the feedback noise $Z_t\tilde{h}_{t|t-1}$ on the true error (15) will then tend to zero as $\lambda_e/\lambda_\nu \to 0$.

Proof: See Lemma 1 and Lemma 2 of [5] □

With a negligible feedback noise, the feedback around $\mathcal{L}_I(q^{-1})$ in Figure 1 can be neglected. Stability will for $||Z_t\tilde{h}_{t|t-1}|| \to 0$ be assured by stability of the learning filter.

[1] It can be noted from (13) and (14) that the feedback noise vanishes when $\varphi_t\varphi_t^*$ can be substituted by its average $\mathbf{R}$.

The Wiener design presented in [1] can for slow variations be performed directly, without iterations. Furthermore, whenever the feedback noise can be neglected, steady state tracking error covariance matrices can be calculated for a given (not necessarily optimally designed) adaptation law with constant gain. More specifically, for slowly varying parameters and for stationary tracking errors, the covariance matrix $\mathbf{P}_k$ of the steady-state k-step tracking error, defined by (5), is under Assumption 1 given by

$$\mathbf{P}_k = \lim_{t\to\infty}(\mathbf{V}_{h,t}^k + \mathbf{V}_{\varphi v,t}^k) \ . \tag{28}$$

For systems described by (1),(6),(9), (21)-(24), this equals

$$\mathbf{P}_k = \frac{\lambda_e}{2\pi j}\oint(\mathbf{I} - z^{-k}\mathcal{L}_k\mathbf{R})\frac{\boldsymbol{C}\bar{\mathbf{R}}_e\boldsymbol{C}_*}{DD_*}(\mathbf{I} - z^k\mathcal{L}_{k*}\mathbf{R})\frac{dz}{z} + \frac{\lambda_v}{2\pi j}\oint\mathcal{L}_k\frac{\boldsymbol{M}\bar{\mathbf{R}}_\nu\boldsymbol{M}_*}{NN_*}\mathcal{L}_{k*}\frac{dz}{z} \ . \tag{29}$$

The integrands of (29) provide the distribution in the frequency domain of the parameter lag error and the noise-induced error, respectively.

The expression (29) holds for arbitrary parameter dynamics (23) as well as for colored measurement noise and correlated regressors, as long as the feedback noise can be neglected. Note, however, that $\tilde{h}_{t+k|t}$ must be assumed stationary, implying bounded variance, even for parameters generated by marginally stable systems (23). Stationarity is guaranteed if and only if $\mathcal{L}_k(q^{-1})$ is stable and all elements of the lag error matrix in (15), $\mathbf{I}-q^{-k}\mathcal{L}_k(q^{-1})\mathbf{R}$, are assured to contain all marginally stable factors of $D(q^{-1})$ in their numerators. This property will be guaranteed in the MSE-optimal design [1]. In the case of integration, $D = 1 - q^{-1}$, the low-frequency gain (for $z = e^{j\omega} = 1$) of $\mathcal{L}_k(z^{-1})$ must be $\mathbf{R}^{-1}$. This is e.g. true for the LMS filter (11).

4. EXAMPLE

The validity of expression (29) will be investigated for a scalar FIR system with two parameters

$$y_t = h_{0,t}u_t + h_{1,t}u_{t-1} + v_t \ , \tag{30}$$

with white zero mean noise and with white zero mean regressors, $u_t \in \{1, -1, i, -i\}$, yielding $\mathbf{R} = \mathbf{I}_2$. The parameter dynamics is governed by the second order AR process

$$h_t = 2p\cos\omega_o h_{t-1} - p^2h_{t-2} + e_t \ , \tag{31}$$

with $p = 0.999$ and where ω_o is here varied to provide different magnitudes of the DNS expressed by (20). The variance of v_t is set to 0.01 and $\mathbf{R}_e = \lambda_e\mathbf{I}$, where λ_e is selected to give an output SNR of 20 dB, with $\mathrm{E}\,||h_t||_2^2 = 1$.

One-step prediction estimates are obtained by LMS and by a Wiener-designed adaptation law (WLMS), tuned to the dynamics of the FIR system by using (25)-(27). (These equations become simple and scalar, with closed-form solutions, in this case [2].) The LMS step-size μ is tuned to minimize the simulated performance, whereas the Wiener design is here based on the assumption of slow parameter variations according to Definition 1.

Table 1 compares $\mathrm{tr}\,\mathbf{P}_1$ with corresponding estimates, obtained by simulation over 100000 data (italic figures). Note the much lower tracking error variance for the Wiener design as compared to LMS.

The term $\mathrm{tr}\,\mathbf{V}^1_{Z\tilde{h}} = \lim_{t\to\infty} \mathbf{V}^1_{Z\tilde{h},t}$, which is the largest term due to the feedback noise in (16), is also measured. This term essentially explains the difference between the expression (29) that neglects the feedback noise, and the true performance. A more accurate performance analysis that works for fast variations in FIR parameters by taking also the term $\mathrm{tr}\,\mathbf{V}^1_{Z\tilde{h}}$ into account is presented in [5].

For LMS, the expression (29) predicts the performance reasonably well for ω_o below 0.005, while the Wiener design, performance is well predicted by the theory up to $\omega_o \approx 0.02$ if the limit for significant deviations is set to 10%.

Figure 2 displays the spectra of the two error term integrands in (29) for $\omega_o = 0.01$. Note the peak of the LMS lag error around ω_o and the contribution of high-frequency noise to the LMS error spectrum.

The bottom line of this example is that it does indeed pay off to use more sophisticated algorithms than LMS even in cases of slow parameter variations. One reason is that LMS implicitly assumes random walk parameter variations. Furthermore, the degree of nonstationarity (20) is designed for LMS but gives a somewhat crude indication of when a simplified analysis is adequate in general. The concept of negligible feedback noise gives better guidelines.

Table 1. The asymptotic tracking error $\mathrm{tr}\,\mathbf{P}_1$ when second order FIR models (30),(31) are tracked by LMS and WLMS. Theoretical predictions from (29) (bold) and simulation results (in italics).

	ω_o	0.001	0.005	0.01	0.02	0.10
DNS:	(20)	.0141	.0510	.1005	.2002	.9996
LMS:	$\mathrm{tr}\,\mathbf{P}_1$	**.0011**	**.0027**	**.0045**	**.0075**	**.0360**
		.0012	*.0030*	*.0052*	*.0099*	*.0650*
	$\mathrm{tr}\,\mathbf{V}^1_{Z\tilde{h}}$	*.0001*	*.0003*	*.0007*	*.0020*	*.0278*
WIENER	$\mathrm{tr}\,\mathbf{P}_1$	**.0007**	**.0013**	**.0019**	**.0028**	**.0061**
DESIGN:		*.0007*	*.0014*	*.0021*	*.0031*	*.0076*
	$\mathrm{tr}\,\mathbf{V}^1_{Z\tilde{h}}$	*.0000*	*.0001*	*.0002*	*.0003*	*.0015*

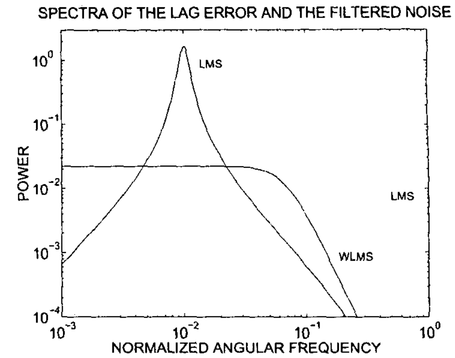

Fig. 2. The lag error (solid) and the filtered noise (dashed), which equals $0.01|\mathcal{L}_1(\omega)|^2$, for Wiener estimators (WLMS) and for LMS, with $\omega_o = 0.01$ in (31).

5. REFERENCES

[1] 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

[2] L.Lindbom, M.Sternad and A.Ahlén, "Tracking of time-varying mobile radio channels. Part I: The Wiener LMS algorithm," To appear in *IEEE Trans. on Communications.* www.signal.uu.se/Publications/abstracts/r003.html

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

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

[5] 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

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

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

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

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

[10] L.Guo and L.Ljung, "Performance analysis of general tracking algorithms," *IEEE Trans. on Automatic Control,* vol. 40, pp. 1388-1402, 1995.

[11] O.Macchi, "Optimization of adaptive identification for time-varying filters," *IEEE Trans. on Automatic Control,* vol. 31, pp. 283-287, 1986.

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