Postprint of: Niedźwiecki M., On "cheap smoothing" opportunities in identification of time-varying systems, Automatica, Vol. 44, Iss. 5 (2008), pp. 1191-1200, DOI: 10.1016/j.automatica.2007.09.007



# On 'cheap smoothing' opportunities in identification of time-varying systems

Maciej Niedźwiecki [a]

[a] *Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control, Gdańsk University of Technology, Narutowicza 11/12, Gdańsk , Poland, Tel: + 48 58 3472519; fax: +48 58 3415821*

**Abstract**

In certain applications of nonstationary system identification the model-based decisions can be postponed, i.e. executed with a delay. This allows one to incorporate into the identification process not only the currently available information, but also a number of "future" data points. The resulting estimation schemes, which involve smoothing, are not causal. Despite the possible performance improvements, the existing smoothing algorithms are seldom used in practice, mainly because of their high computational requirements. We show that the computationally attractive smoothing procedures can be obtained by means of compensating estimation delays that arise in the standard exponentially weighted least squares, least mean squares and Kalman filter based parameter trackers.

*Key words:* system identification, time-varying processes, adaptive filtering

## 1 Introduction

Consider the problem of tracking of a nonstationary discrete-time stochastic system governed by

$$y(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t) + v(t) \tag{1}$$

where $t = \ldots, -1, 0, 1, \ldots$ denotes normalized time, $y(t)$ is the system output, $\boldsymbol{\varphi}(t) = [\varphi_1(t), \ldots, \varphi_n(t)]^{\mathrm{T}}$ is the regression vector (e.g. made up of the past input measurements), $\boldsymbol{\theta}(t) = [\theta_1(t), \ldots, \theta_n(t)]^{\mathrm{T}}$ is the vector of unknown and time-varying system coefficients, and $v(t)$ denotes measurement noise.
The problem of estimation of the parameter vector $\boldsymbol{\theta}(t)$, based on the available data, can be solved in many different ways. When all that is known about system parameters is that they vary slowly with time, the most frequently used

*Email address:* maciekn@eti.pg.gda.pl (Maciej Niedźwiecki).

identification algorithms are those based on the exponentially weighted least squares (EWLS) approach, the least mean squares (LMS) approach and the Kalman filter (KF) approach – see e.g. (Haykin, 1996) and (Niedźwiecki, 2000), among many others.

Denote by $\widehat{\boldsymbol{\theta}}(t)$ the estimate of $\boldsymbol{\theta}(t)$ and by $\varepsilon(t)$ – the one-step-ahead prediction error evaluated at instant $t$

$$\varepsilon(t) = y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}(t-1) \tag{2}$$

The EWLS parameter update can be summarized as follows

$$\begin{aligned} \widehat{\boldsymbol{\theta}}(t) &= \widehat{\boldsymbol{\theta}}(t-1) + \mathbf{R}^{-1}(t)\boldsymbol{\varphi}(t)\varepsilon(t) \\ \mathbf{R}(t) &= \eta\mathbf{R}(t-1) + \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) \end{aligned} \tag{3}$$

where $\mathbf{R}(t)$ denotes the exponentially weighted regression matrix and $\eta$, $0 < \eta < 1$, is the so-called forgetting constant, which decides upon the estimation memory of the EWLS algorithm. Using the well-known matrix inversion lemma it is possible to derive recursive formula for direct updating of $\mathbf{R}^{-1}(t)$, which allows one to avoid inverting the regression matrix at each step of the EWLS algorithm (Niedźwiecki, 2000).

The LMS algorithm

$$\widehat{\boldsymbol{\theta}}(t) = \widehat{\boldsymbol{\theta}}(t-1) + \mu\boldsymbol{\varphi}(t)\varepsilon(t) \tag{4}$$

can be thought of as a simplified EWLS algorithm, obtained by replacing the time-varying and data-dependent weighting matrix $\mathbf{R}^{-1}(t)$ with a constant, scalar stepsize parameter $\mu > 0$.

Finally, justification of the KF algorithm

$$\begin{aligned} \widehat{\boldsymbol{\theta}}(t) &= \widehat{\boldsymbol{\theta}}(t-1) + \mathbf{S}(t)\boldsymbol{\varphi}(t)\varepsilon(t) \\ \mathbf{S}(t) &= \frac{\mathbf{P}(t-1)}{1 + \boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{P}(t-1)\boldsymbol{\varphi}(t)} \\ \mathbf{P}(t) &= (\mathbf{I}_n - \mathbf{S}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t))\mathbf{P}(t-1) + \kappa^2\mathbf{I}_n \end{aligned} \tag{5}$$

where $\mathbf{I}_n$ denotes the $n \times n$ identity matrix, comes from the area of statistical filtering theory. When

**(i)** the (white) measurement noise $v(t)$ is Gaussian: $v(t) \sim \mathcal{N}(0, \sigma_v^2)$

**(ii)** the sequence of one-step parameter changes $\mathbf{w}(t) = \boldsymbol{\theta}(t) - \boldsymbol{\theta}(t-1)$, independent of $\{v(t)\}$, is made up of uncorrelated random variables with Gaussian distribution: $\mathbf{w}(t) \sim \mathcal{N}(\mathbf{0}, \sigma_w^2\mathbf{I}_n)$ (random walk model)

**(iii)** $\kappa^2 = \sigma_w^2/\sigma_v^2$

the KF algorithm is an optimal estimation procedure in the sense that it provides the estimates of $\boldsymbol{\theta}(t)$ with the smallest possible mean-square errors (Lewis, 1986). One should be careful, though, not to overemphasize this feature of the Kalman filter approach. When conditions listed above (rather naive, from the practical viewpoint) are not met, the KF algorithm can no longer be claimed optimal. It provides yet another way for recursive estimation of time-varying system coefficients, neither more nor less appropriate than the two approaches mentioned earlier. In typical applications the scalar coefficient $\kappa$ is treated instrumentally, as a user-dependent tuning "knob", deciding upon the estimation memory of the KF algorithm, i.e. it plays a similar role as $\eta$ and $\mu$ in the EWLS and LMS algorithms, respectively – see next section for more details.

All three estimation schemes presented above are causal, i.e. at each time instant $t$ they provide parameter estimates that are functions of the current and past measurements: $y(s), \boldsymbol{\varphi}(s), s \leq t$. While in the adaptive prediction and adaptive control applications causality is an obvious requirement, there are some other practical problems, such as adaptive noise canceling or adaptive channel equalization, where the causality constraint can be relaxed by means of incorporating into the estimation process a certain number of "future" measurements. Such noncausal estimation schemes are feasible whenever the model-based decisions can be postponed, i.e. executed with a delay. For example, when a decision delay of $\tau$ sampling intervals is tolerable, the data set $Z(t) = \{y(1), \boldsymbol{\varphi}(1), \ldots, y(t), \boldsymbol{\varphi}(t)\}$ can be used to estimate $\boldsymbol{\theta}(t-\tau)$, rather than to estimate $\boldsymbol{\theta}(t)$. When appropriately designed, such estimator of $\boldsymbol{\theta}(t-\tau)$, based on all past measurements and $\tau$ "future" measurements, will yield better results than its causal counterpart.

The design platform which in a straightforward way leads to such solutions is Kalman smoothing (KS) – an extension of the Kalman filtering approach. Despite the potential performance improvements, the KS algorithms are seldom used in practice, mainly because of their high computational complexity. We will show that computationally attractive parameter smoothing procedures can be obtained by means of compensating estimation delays which arise in the standard EWLS, LMS and KF algorithms.

The paper is organized as follows. Section 2 presents analysis of the estimation delay effects occurring in the classical EWLS/LMS/KF tracking algorithms. The 'cheap smoothing' algorithms are described in Section 3 and their estimation properties are studied in Section 4. The simplified smoothing rules, yielding further computational savings, are described in Section 5. Section 6 shows the results of simulation experiments and Section 7 presents some practical recommendations. Finally, Section 8 concludes.

## 2 Estimation delay effects in parameter tracking algorithms

Suppose that

**(A1)** The measurement noise sequence $\{v(t)\}$ is zero-mean and white with variance $\sigma_v^2$.

**(A2)** The sequence of regression vectors sequence $\{\boldsymbol{\varphi}(t)\}$ is zero-mean, wide-sense stationary and ergodic with the covariance matrix $\mathrm{E}[\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)] = \boldsymbol{\Phi} > 0$.

Then, when the forgetting constant $\eta$ in the EWLS algorithm is sufficiently close to one, the following steady-state approximation can be used (Ljung & Gunnarsson, 1985), (Niedźwiecki, 2000)

$$\mathbf{R}^{-1}(t) \cong (1-\eta)\mathbf{\Phi}^{-1}$$

Similarly, when the coefficient $\kappa$ in the KF algorithm is sufficiently close to zero, one gets (Ljung & Gunnarsson, 1985), (Niedźwiecki, 2000)

$$\mathbf{S}(t) \cong \kappa\mathbf{\Phi}^{-1/2}$$

where $\mathbf{\Phi}^{-1/2} = (\mathbf{\Phi}^{1/2})^{-1}$, and $\mathbf{\Phi}^{1/2} > 0$ denotes the (unique) square root of the covariance matrix $\mathbf{\Phi} : \mathbf{\Phi}^{1/2}\mathbf{\Phi}^{1/2} = \mathbf{\Phi}$. Using these approximations the three estimation algorithms described above can be written down in the following "standardized" form

$$\widehat{\boldsymbol{\theta}}(t) = \widehat{\boldsymbol{\theta}}(t-1) + \gamma\mathbf{A}\boldsymbol{\varphi}(t)\varepsilon(t) \tag{6}$$

where the small adaptation gain $\gamma$ and the constant matrix $\mathbf{A}$ are given by

$$\begin{aligned} \text{EWLS} &: \gamma = 1-\eta \ , \quad \mathbf{A} = \mathbf{\Phi}^{-1} \\ \text{LMS} &: \quad \gamma = \mu \quad , \quad \mathbf{A} = \mathbf{I}_n \\ \text{KF} &: \quad \gamma = \kappa \quad , \mathbf{A} = \mathbf{\Phi}^{-1/2} \end{aligned}$$

Even though derivation of the "standardized" algorithm is based on heuristic arguments, the results of a more rigorous statistical analysis of tracking performance of the EWLS, LMS and KF schemes, presented in (Guo & Ljung, 1995), are consistent with the analogous results based on (6) (Ljung & Gunnarsson, 1985), (Niedźwiecki, 2000). For this reason we will adopt (6) as the starting point for our study of the estimation delay effects that occur in parameter tracking algorithms. Later on, in Section 6, we will show that experimental results fully confirm conclusions drawn from such approximate analysis.
Combining (6) with (1) and (2) one obtains

$$\widehat{\boldsymbol{\theta}}(t) = (\mathbf{I}_n - \gamma\mathbf{A}\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t))\widehat{\boldsymbol{\theta}}(t-1) + \gamma\mathbf{A}\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t) + \gamma\mathbf{A}\boldsymbol{\varphi}(t)v(t) \tag{7}$$

For sufficiently small values of the adaptation gain $\gamma$ and for sufficiently slow changes in $\boldsymbol{\theta}(t)$ (compared to the changes in $\boldsymbol{\varphi}(t)$), the analysis of (7) can be carried using the averaging technique (Bai, Fu & Sastry, 1988), leading to the following approximation

$$\widehat{\boldsymbol{\theta}}(t) = (\mathbf{I}_n - \gamma\mathbf{A}\mathbf{\Phi})\widehat{\boldsymbol{\theta}}(t-1) + \gamma\mathbf{A}\mathbf{\Phi}\boldsymbol{\theta}(t) + \gamma\mathbf{A}\boldsymbol{\varphi}(t)v(t) \tag{8}$$

Denote by $\bar{\boldsymbol{\theta}}(t) = \mathrm{E}[\widehat{\boldsymbol{\theta}}(t)|\boldsymbol{\theta}(s), s \leq t]$ the mean path of parameter estimates. Using (7) one arrives at

$$\begin{aligned}\bar{\boldsymbol{\theta}}(t) &\cong (\mathbf{I}_n - \gamma \mathbf{A}\boldsymbol{\Phi})\bar{\boldsymbol{\theta}}(t-1) + \gamma \mathbf{A}\boldsymbol{\Phi}\boldsymbol{\theta}(t) \\ &\cong \gamma \left[\mathbf{I}_n - (\mathbf{I}_n - \gamma \mathbf{A}\boldsymbol{\Phi})q^{-1}\right]^{-1} \mathbf{A}\boldsymbol{\Phi}\boldsymbol{\theta}(t) \end{aligned} \tag{9}$$

where $q^{-1}$ denotes the backward shift operator.

### *2.1 EWLS algorithm*

Since in the EWLS case it holds that $\mathbf{A}\boldsymbol{\Phi} = \mathbf{I}_n$, the relationship (9) can be rewritten in the form

$$\bar{\boldsymbol{\theta}}(t) \cong \mathbf{F}_{\mathrm{EWLS}}(q^{-1})\boldsymbol{\theta}(t) \tag{10}$$

where

$$\mathbf{F}_{\mathrm{EWLS}}(q^{-1}) = \mathrm{diag}\{F(q^{-1}), \ldots, F(q^{-1})\}$$

and

$$F(q^{-1}) = \frac{1-\eta}{1-\eta q^{-1}}$$

According to (10) the mean path of the EWLS estimates $\{\bar{\boldsymbol{\theta}}(t)\}$ can be regarded a result of passing $\{\boldsymbol{\theta}(t)\}$ through a linear lowpass filter $F(q^{-1})$. This means that for slow parameter changes the main contribution to the bias error $\boldsymbol{\theta}(t) - \bar{\boldsymbol{\theta}}(t)$ is due to the lag distortions – the mean trajectory $\{\bar{\boldsymbol{\theta}}(t)\}$ can be approximately viewed as a *delayed* version of the true trajectory $\{\boldsymbol{\theta}(t)\}$. The dominant time delay $\tau_o$ introduced by the filter $F(q^{-1})$ – called the estimation delay in (Niedźwiecki, 2000) – can be defined in different ways.
One possibility, based on the frequency-domain concepts, is to set

$$\tau_o = \mathrm{int}[t_o], \quad t_o = \frac{\eta}{1-\eta} \tag{11}$$

where $\mathrm{int}[x]$ denotes the integer number that is closest to $x$ and

$$t_o = -\lim_{\omega \mapsto 0} \frac{\phi_F(\omega)}{\omega} = -\lim_{\omega \mapsto 0} \frac{d\phi_F(\omega)}{d\omega}$$

is the nominal (low-frequency) delay of the filter $F(e^{-j\omega}) = A_F(\omega)e^{j\phi_F(\omega)}$; $\omega \in (-\pi, \pi]$ denotes the normalized angular frequency.
Another solution is based on the time domain arguments. One can define $\tau_o$ as the average delay of $\bar{\boldsymbol{\theta}}(t)$ with respect to $\boldsymbol{\theta}(t)$ for a given class of parameter variations

$$\tau_o = \arg\inf_{\tau} \mathrm{E}[||\bar{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t-\tau)||^2]$$

where averaging is carried over different realizations of $\{\boldsymbol{\theta}(t)\}$. As shown in (Niedźwiecki, 2000), when system parameters evolve according to the random walk model, the average delay can be obtained by means of solving the following equation

$$\sum_{t=0}^{\tau_o-1} f(t) \cong \sum_{t=\tau_o}^{\infty} f(t) \tag{12}$$

where $f(t) = \mathcal{Z}^{-1}[F(z^{-1})] = (1-\eta)\eta^t$ denotes the impulse response of the filter $F(q^{-1})$. Slightly abusing the term, which was coined in statistics to characterize distributions of random variables (note that $f(t) > 0$, $\forall t \geq 0$ and $\sum_{t=0}^{\infty} f(t) = 1$, i.e. $f(t)$ can be regarded a discrete probability function of a fictitious random variable), $\tau_o$ can be called the "median" of the impulse response $f(t)$.
Observe that $(1-\eta)\sum_{t=0}^{\tau_o-1} \eta^t = 1-\eta^{\tau_o}$ and $(1-\eta)\sum_{t=\tau_o}^{\infty} \eta^t = \eta^{\tau_o}$. Therefore (12) is equivalent to $\eta^{\tau_o} \cong 0.5$ or, after transformation, to $\tau_0 \ln \eta \cong \ln 0.5 \cong -0.7$. Since for the values of $\eta$ close to one it holds that $\ln \eta \cong \eta - 1$, one finally arrives at

$$\tau_o = \mathrm{int}[t_o], \quad t_o = \frac{0.7}{1-\eta} \tag{13}$$

Using the approximation $F(q^{-1}) \cong q^{-\tau_o}$, i.e.

$$\mathbf{F}_{\mathrm{EWLS}}(q^{-1}) \cong \mathbf{G}_{\mathrm{EWLS}}(q^{-1}) = q^{-\tau_o}\mathbf{I}_n$$

one can rewrite (10) in the following form

$$\bar{\boldsymbol{\theta}}(t) \cong \mathbf{G}_{\mathrm{EWLS}}(q^{-1})\boldsymbol{\theta}(t) = \boldsymbol{\theta}(t-\tau_o) \tag{14}$$

which will be a convenient starting point for our further considerations. The "pure delay" approximation of $F(q^{-1})$ is of course very crude, but it will lead us to computationally attractive solutions.

### *2.2 LMS and KF algorithms*

Let $\mathbf{Q}$ be a unitary matrix, made up of the eigenvectors of $\boldsymbol{\Phi}$, converting $\boldsymbol{\Phi}$ into a diagonal form

$$\mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{Q}\mathbf{Q}^{\mathrm{T}} = \mathbf{I}_n, \quad \mathbf{Q}^{\mathrm{T}}\boldsymbol{\Phi}\mathbf{Q} = \boldsymbol{\Lambda}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix made up of the eigenvalues of $\boldsymbol{\Phi}$: $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}$. Note that (9) can be rewritten in the form

$$\bar{\boldsymbol{\theta}}(t) \cong \mathbf{Q}\mathbf{F}(q^{-1})\mathbf{Q}^{\mathrm{T}}\boldsymbol{\theta}(t) \tag{15}$$

Depending on the algorithm used, $\mathbf{F}(q^{-1})$ is equal to $\mathbf{F}_{\mathrm{LMS}}(q^{-1}) = \mathrm{diag}\{F_1(q^{-1}), \ldots, F_n(q^{-1})\}$ where

$$F_i(q^{-1}) = \frac{\mu\lambda_i}{1-(1-\mu\lambda_i)q^{-1}}\ , \quad i = 1, \ldots, n$$

or to $\mathbf{F}_{\mathrm{KF}}(q^{-1}) = \mathrm{diag}\{F_1^*(q^{-1}), \ldots, F_n^*(q^{-1})\}$ where

$$F_i^*(q^{-1}) = \frac{\kappa\sqrt{\lambda_i}}{1-(1-\kappa\sqrt{\lambda_i})q^{-1}}\ , \quad i = 1, \ldots, n$$

Similarly as in the EWLS case, one can use the approximations $F_i(q^{-1}) \cong q^{-\tau_i}$, $F_i^*(q^{-1}) \cong q^{-\tau_i^*}$, where $\tau_i = \mathrm{int}[t_i]$, $\tau_i^* = \mathrm{int}[t_i^*]$ and

$$t_i = \frac{1-\mu\lambda_i}{\mu\lambda_i}\ ,\ t_i^* = \frac{1-\kappa\sqrt{\lambda_i}}{\kappa\sqrt{\lambda_i}} \quad \text{or} \quad t_i = \frac{0.7}{\mu\lambda_i}\ ,\ t_i^* = \frac{0.7}{\kappa\sqrt{\lambda_i}}$$

leading to

$$\bar{\boldsymbol{\theta}}(t) \cong \mathbf{Q}\mathbf{G}(q^{-1})\mathbf{Q}^{\mathrm{T}}\boldsymbol{\theta}(t) \tag{16}$$

where $\mathbf{G}(q^{-1})$ is equal to $\mathbf{G}_{\mathrm{LMS}}(q^{-1}) = \mathrm{diag}\{q^{-\tau_1}, \ldots, q^{-\tau_n}\}$, or to $\mathbf{G}_{\mathrm{KF}}(q^{-1}) = \mathrm{diag}\{q^{-\tau_1^*}, \ldots, q^{-\tau_n^*}\}$.

Note that, unlike the EWLS case, different components of the vector $\mathbf{Q}^{\mathrm{T}}\boldsymbol{\theta}(t)$ are delayed by different amounts of time. The corresponding time constants, $\tau_1, \ldots, \tau_n$ and $\tau_1^*, \ldots, \tau_n^*$, respectively, depend both on the adaptation gains $\mu$, $\kappa$ and – via the eigenvalues $\lambda_1, \ldots, \lambda_n$ – on the covariance structure of the excitation. In the special case where $\boldsymbol{\Phi} = \sigma_\varphi^2\mathbf{I}_n$, all eigenvalues of $\boldsymbol{\Phi}$ (and hence also all time delays) are identical, leading to $\bar{\boldsymbol{\theta}}(t) \cong \boldsymbol{\theta}(t-\tau_o)$, where $\tau_o = \tau_1 = \ldots = \tau_n$ or $\tau_o = \tau_1^* = \ldots = \tau_n^*$ are the common delays. From the qualitative viewpoint the latter result is identical with (14).

**Remark**

Denote by $\lambda_{\max}$ and $\lambda_{\min}$ the maximum and minimum eigenvalues of $\boldsymbol{\Phi}$, respectively. Similarly, denote by $t_{\max}$, $t^*_{\max}$ and $t_{\min}$, $t^*_{\min}$ the corresponding maximum/minimum delays. For small adaptation gains it holds that

$$\frac{t_{\max}}{t_{\min}} \cong \frac{\lambda_{\max}}{\lambda_{\min}}\ , \quad \frac{t^*_{\max}}{t^*_{\min}} \cong \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}$$

This means that the estimation delay spread will be always larger for the LMS estimator than for the KF estimator.

## 3 'Cheap smoothing'

### *3.1 Algorithms*

Based on (14), the approximately debiased estimate of $\boldsymbol{\theta}(t)$ can be obtained from

$$\widetilde{\boldsymbol{\theta}}(t) = \mathbf{G}_{\mathrm{EWLS}}(q)\widehat{\boldsymbol{\theta}}(t) = \widehat{\boldsymbol{\theta}}(t+\tau_o) \tag{17}$$

which means that the smoothing effect is achieved simply by delaying the estimates provided by the EWLS tracker by $\tau_o$ sampling intervals. According

to (17), $\widehat{\boldsymbol{\theta}}(t)$ should be regarded as an estimate of $\boldsymbol{\theta}(t-\tau_o)$, rather than as an estimate of $\boldsymbol{\theta}(t)$. To use such a fixed-lag smoother one should incorporate into the adaptive loop a decision delay equal to $\tau_d = \tau_o$ sampling intervals.
In the similar way one can obtain smoothing rules for the LMS and KF estimators. Let $\boldsymbol{\beta}(t) = \mathbf{Q}^{\mathrm{T}}\boldsymbol{\theta}(t)$ and $\widehat{\boldsymbol{\beta}}(t) = \mathbf{Q}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}(t)$. Consider the LMS algorithm first. According to (15), for slow parameter changes it holds that

$$\begin{aligned}\bar{\boldsymbol{\beta}}(t) &= \mathrm{E}[\widehat{\boldsymbol{\beta}}(t)|\boldsymbol{\theta}(s), s \le t] \cong \mathbf{F}_{\mathrm{LMS}}(q^{-1})\mathbf{Q}^{\mathrm{T}}\boldsymbol{\theta}(t)\\ &= \mathbf{F}_{\mathrm{LMS}}(q^{-1})\boldsymbol{\beta}(t) \cong \mathbf{G}_{\mathrm{LMS}}(q^{-1})\boldsymbol{\beta}(t)\end{aligned}$$

which suggests the following three-step procedure for computing "debiased" LMS estimates

$$\begin{aligned}\widehat{\boldsymbol{\beta}}(t) &= \mathbf{Q}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}(t)\\ \widetilde{\boldsymbol{\beta}}(t) &= \mathbf{G}_{\mathrm{LMS}}(q)\widehat{\boldsymbol{\beta}}(t) = [\widehat{\beta}_1(t+\tau_1), \ldots, \widehat{\beta}_n(t+\tau_n)]^{\mathrm{T}}\\ \widetilde{\boldsymbol{\theta}}(t) &= \mathbf{Q}\widetilde{\boldsymbol{\beta}}(t)\end{aligned} \tag{18}$$

The decision delay associated with (18) is equal to $\tau_d = \max\{\tau_1, \ldots, \tau_n\}$ sampling intervals.
Similarly, for the KF estimator one obtains

$$\begin{aligned}\widehat{\boldsymbol{\beta}}(t) &= \mathbf{Q}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}(t)\\ \widetilde{\boldsymbol{\beta}}(t) &= \mathbf{G}_{\mathrm{KF}}(q)\widehat{\boldsymbol{\beta}}(t) = [\widehat{\beta}_1(t+\tau_1^*), \ldots, \widehat{\beta}_n(t+\tau_n^*)]^{\mathrm{T}}\\ \widetilde{\boldsymbol{\theta}}(t) &= \mathbf{Q}\widetilde{\boldsymbol{\beta}}(t)\end{aligned} \tag{19}$$

and $\tau_d = \max\{\tau_1^*, \ldots, \tau_n^*\}$.
When the admissible delay $\tau$ is smaller than $\tau_d$, one can use the following modified versions of (17) - (19):

$$\begin{aligned}\text{EWLS}: &\quad \widetilde{\boldsymbol{\theta}}(t) = \widehat{\boldsymbol{\theta}}(t-\tau)\\ \text{LMS}: &\quad \widetilde{\boldsymbol{\theta}}(t) = \mathbf{Q}\mathbf{G}_{\mathrm{LMS}}(q,\tau)\mathbf{Q}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}(t)\\ \text{KF}: &\quad \widetilde{\boldsymbol{\theta}}(t) = \mathbf{Q}\mathbf{G}_{\mathrm{KF}}(q,\tau)\mathbf{Q}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}(t)\end{aligned} \tag{20}$$

where

$$\begin{aligned}\mathbf{G}_{\mathrm{LMS}}(q,\tau) &= \mathrm{diag}\{q^{\min\{\tau,\tau_1\}}, \ldots, q^{\min\{\tau,\tau_n\}}\}\\ \mathbf{G}_{\mathrm{KF}}(q,\tau) &= \mathrm{diag}\{q^{\min\{\tau,\tau_1^*\}}, \ldots, q^{\min\{\tau,\tau_n^*\}}\}\end{aligned}$$

**Remark**

The idea of reducing estimation bias by means of incorporating in the processing loop an appropriately chosen decision delay can be traced back to Hedelin (Hedelin, 1977). Hedelin demonstrated that delaying state estimates provided by Kalman filter can be regarded an efficient form of smoothing. However,

| Algorithm | Computational complexity | Smoothing overhead |
|---|---|---|
| EWLS | $2n^2 + 5n$ | $+\,0$ |
| LMS | $2n + 1$ | $+\,2n^2$ |
| KF | $1.5n^2 + 5.5n$ | $+\,2n^2$ |

**Table I** Computational complexity and smoothing overheads of the three parameter estimation algorithms analyzed in the paper.

Hedelin did not show how the optimal delay(s) can be determined. The results presented above seem to be interesting from at least two reasons. First, we have revealed the hidden 'delay structure' (depending on the eigendecomposition of $\mathbf{\Phi}$) of the LMS and KF estimators. Second, we have shown how the corresponding delays can be computed and compensated.

### *3.2 Computational complexity*

Table I shows comparison of the computational complexity of the basic algorithms (the number of multiply/add operations needed to complete one cycle of computations) and the corresponding smoothing overheads (the number of additional operations required to perform smoothing). The count was made for the computationally efficient mechanizations of the EWLS/KF algorithms (which differ from (3) and (5)) and takes into consideration symmetry of the matrices $\mathbf{R}(t)$ and $\mathbf{P}(t)$. The cost of performing eigendecomposition of $\mathbf{\Phi}$ was not included since, assuming that the process $\{\boldsymbol{\varphi}(t)\}$ is stationary, such operation is performed only once. Note that there is no smoothing overhead in the case of the EWLS estimator, and that for the LMS/KF estimators the overhead does not depend on the decision delay $\tau_d$.
This should be confronted with complexity of the fixed-lag Kalman smoothing (KS) algorithm designed for the system (1) – see Chapter 7 in (Niedźwiecki, 2000) – which is equal to $(0.5\tau^2 + 2\tau + 1.5)n^2 + (2.5\tau + 5.5)n$ multiply/add operations per time update. Note that in this case the smoothing overhead, i.e. the difference between the computational complexity of the KS algorithm and complexity of the KF algorithm (i.e. the zero-lag smoother) is equal to

$$(0.5\tau^2 + 2\tau)n^2 + 2.5\tau n$$

and rapidly grows with the smoothing lag $\tau$.

## 4 Mean-square error analysis

Observe that $\mathbf{G}_{\mathrm{EWLS}}(q) = \mathbf{Q}\mathbf{G}_{\mathrm{EWLS}}(q)\mathbf{Q}^{\mathrm{T}}$ and hence the relationships (17), (18) and (19) can be written down in the following unified form

$$\widetilde{\boldsymbol{\theta}}(t) = \mathbf{Q}\mathbf{G}(q)\mathbf{Q}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}(t) \tag{21}$$

where, depending on the algorithm used, $\mathbf{G}(q)$ is equal to $\mathbf{G}_{\mathrm{EWLS}}(q)$, $\mathbf{G}_{\mathrm{LMS}}(q)$ or $\mathbf{G}_{\mathrm{KF}}(q)$.
To compare estimation accuracy of the EWLS/LMS/KF estimators $\widehat{\boldsymbol{\theta}}(t)$ and their modified versions $\widetilde{\boldsymbol{\theta}}(t)$, we will assume that

(**A3**) $\{\boldsymbol{\theta}(t)\}$ is a zero-mean wide-sense stationary process with a spectral density (matrix) function $\mathbf{S}_{\theta}(\omega)$.

The mean-square parameter estimation error yielded by the basic algorithm can be expressed in the form

$$\mathrm{E}[||\widehat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t)||^2] = \mathrm{E}[||\boldsymbol{\delta}(t)||^2] + \mathrm{E}[||\boldsymbol{\eta}(t)||^2] \tag{22}$$

where $\boldsymbol{\delta}(t) = \widehat{\boldsymbol{\theta}}(t) - \bar{\boldsymbol{\theta}}(t)$, $\boldsymbol{\eta}(t) = \bar{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t)$ and averaging is carried over different realizations of $\{v(t)\}$ and $\{\boldsymbol{\theta}(t)\}$. The first term on the right-hand side of (22) constitutes the variance component of mean-square error (MSE), whereas the second term can be recognized as its bias component. The analogous expression for the modified algorithm reads

$$\mathrm{E}[||\widetilde{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t)||^2] = \mathrm{E}[||\boldsymbol{\sigma}(t)||^2] + \mathrm{E}[||\boldsymbol{\xi}(t)||^2] \tag{23}$$

where $\boldsymbol{\sigma}(t) = \widetilde{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}'(t)$, $\boldsymbol{\xi}(t) = \boldsymbol{\theta}'(t) - \boldsymbol{\theta}(t)$ and $\boldsymbol{\theta}'(t) = \mathrm{E}[\widetilde{\boldsymbol{\theta}}(t)|\boldsymbol{\theta}(s), s \leq t+\tau_d]$.

### *4.1 Variance*

We will prove that the variance components of the compared mean-square errors are approximately the same:

**Theorem**

Under (A1) - (A3) it holds that

$$\mathrm{E}[||\boldsymbol{\sigma}(t)||^2] \cong \mathrm{E}[||\boldsymbol{\delta}(t)||^2] \tag{24}$$

**Proof**

First, after combining (8) with (9) one arrives at

$$\boldsymbol{\delta}(t) \cong (\mathbf{I}_n - \gamma\mathbf{A}\boldsymbol{\Phi})\boldsymbol{\delta}(t-1) + \gamma\mathbf{A}\boldsymbol{\varphi}(t)v(t)$$

which shows that, under the small adaptation gain conditions, the sequence $\{\boldsymbol{\delta}(t)\}$ can be regarded as asymptotically zero-mean and asymptotically wide-sense stationary. Denote by $\mathbf{S}_\delta(\omega)$ the spectral density function of $\boldsymbol{\delta}(t)$

$$\mathbf{S}_\delta(\omega) = \gamma^2\sigma_v^2 \left(\mathbf{I}_n - (\mathbf{I}_n - \gamma\mathbf{A}\boldsymbol{\Phi})e^{-j\omega}\right)^{-1} \mathbf{A}\boldsymbol{\Phi}\mathbf{A}^{\mathrm{T}} \left(\mathbf{I}_n - (\mathbf{I}_n - \gamma\boldsymbol{\Phi}\mathbf{A}^{\mathrm{T}})e^{j\omega}\right)^{-1}$$

and let $\boldsymbol{\Delta} = \mathrm{E}[\boldsymbol{\delta}(t)\boldsymbol{\delta}^{\mathrm{T}}(t)]$. Observe that

$$\mathrm{E}[||\boldsymbol{\delta}(t)||^2] = \mathrm{tr}\{\boldsymbol{\Delta}\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\{\mathbf{S}_\delta(\omega)\}d\omega \tag{25}$$

Since $\boldsymbol{\sigma}(t) = \mathbf{Q}\mathbf{G}(q)\mathbf{Q}^{\mathrm{T}}\boldsymbol{\delta}(t)$ one obtains

$$\mathbf{S}_\sigma(\omega) = \mathbf{Q}\mathbf{G}(e^{j\omega})\mathbf{Q}^{\mathrm{T}}\mathbf{S}_\delta(\omega)\mathbf{Q}\mathbf{G}(e^{-j\omega})\mathbf{Q}^{\mathrm{T}}$$

Note that

$$\mathrm{E}[||\boldsymbol{\sigma}(t)||^2] = \mathrm{tr}\{\boldsymbol{\Sigma}\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\{\mathbf{S}_\sigma(\omega)\}d\omega \tag{26}$$

where $\boldsymbol{\Sigma} = \mathrm{E}[\boldsymbol{\sigma}(t)\boldsymbol{\sigma}^{\mathrm{T}}(t)]$. The relationship (24) follows directly from (25) - (26) and from the identity

$$\mathrm{tr}\{\mathbf{S}_\sigma(\omega)\} = \mathrm{tr}\{\mathbf{S}_\delta(\omega)\mathbf{Q}\mathbf{G}(e^{-j\omega})\mathbf{Q}^{\mathrm{T}}\mathbf{Q}\mathbf{G}(e^{j\omega})\mathbf{Q}^{\mathrm{T}}\} = \mathrm{tr}\{\mathbf{S}_\delta(\omega)\mathbf{I}_n\} = \mathrm{tr}\{\mathbf{S}_\delta(\omega)\}$$

### *4.2 Bias*

In order to compare the bias terms in (14) and (15) note that

$$\boldsymbol{\eta}(t) \cong \mathbf{Q}[\mathbf{F}(q^{-1}) - \mathbf{I}_n]\mathbf{Q}^{\mathrm{T}}\boldsymbol{\theta}(t)$$

$$\mathbf{S}_\eta(\omega) \cong \mathbf{Q}[\mathbf{F}(e^{-j\omega}) - \mathbf{I}_n]\mathbf{Q}^{\mathrm{T}}\mathbf{S}_\theta(\omega)\mathbf{Q}[\mathbf{F}(e^{j\omega}) - \mathbf{I}_n]\mathbf{Q}^{\mathrm{T}}$$

where, depending on the algorithm used, $\mathbf{F}(q^{-1})$ is equal to $\mathbf{F}_{\mathrm{EWLS}}(q^{-1})$, $\mathbf{F}_{\mathrm{LMS}}(q^{-1})$ or $\mathbf{F}_{\mathrm{KF}}(q^{-1})$. This leads to

$$\mathrm{E}[||\boldsymbol{\eta}(t)||^2] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\{\mathbf{S}_\eta(\omega)\}d\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\{\mathbf{B}(\omega)\mathbf{Q}^{\mathrm{T}}\mathbf{S}_\theta(\omega)\mathbf{Q}\}d\omega$$

where $\mathbf{B}_{\mathrm{EWLS}}(\omega) = \mathrm{diag}\{B(\omega), \ldots, B(\omega)\}$, $\mathbf{B}_{\mathrm{LMS}}(\omega) = \mathrm{diag}\{B_1(\omega), \ldots, B_n(\omega)\}$, $\mathbf{B}_{\mathrm{KF}}(\omega) = \mathrm{diag}\{B_1^*(\omega), \ldots, B_n^*(\omega)\}$ and

$$\begin{aligned} B(\omega) &= \left|1 - F(e^{-j\omega})\right|^2 \\ B_i(\omega) &= \left|1 - F_i(e^{-j\omega})\right|^2, \quad i = 1, \ldots, n \\ B_i^*(\omega) &= \left|1 - F_i^*(e^{-j\omega})\right|^2, \quad i = 1, \ldots, n \end{aligned}$$

For the modified EWLS/LMS/KF estimators the analogous expression is

$$\mathrm{E}[||\boldsymbol{\xi}(t)||^2] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\{\mathbf{S}_\xi(\omega)\}d\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\{\widetilde{\mathbf{B}}(\omega)\mathbf{Q}^{\mathrm{T}}\mathbf{S}_\theta(\omega)\mathbf{Q}\}d\omega$$

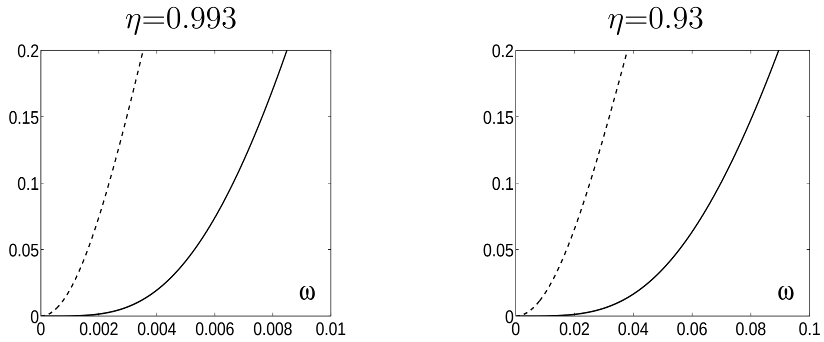


Fig. 1. Bias characteristics $B(\omega)$ (broken line) and $\widetilde{B}(\omega)$ (solid line) for two different values of $\eta$. Note the horizontal scale difference between the left figure and the right figure.

where $\widetilde{\mathbf{B}}_{\mathrm{EWLS}}(\omega) = \mathrm{diag}\{\widetilde{B}(\omega), \ldots, \widetilde{B}(\omega)\}$, $\widetilde{\mathbf{B}}_{\mathrm{LMS}}(\omega) = \mathrm{diag}\{\widetilde{B}_1(\omega), \ldots, \widetilde{B}_n(\omega)\}$, $\widetilde{\mathbf{B}}_{\mathrm{KF}}(\omega) = \mathrm{diag}\{\widetilde{B}_1^*(\omega), \ldots, \widetilde{B}_n^*(\omega)\}$ and

$$\begin{aligned}
\widetilde{B}(\omega) &= \left|1 - e^{j\omega\tau_o} F(e^{-j\omega})\right|^2 \\
\widetilde{B}_i(\omega) &= \left|1 - e^{j\omega\tau_i} F_i(e^{-j\omega})\right|^2, \quad i = 1, \ldots, n \\
\widetilde{B}_i^*(\omega) &= \left|1 - e^{j\omega\tau_i^*} F_i^*(e^{-j\omega})\right|^2, \quad i = 1, \ldots, n
\end{aligned}$$

Figure 1 shows the plots of the bias characteristics $B(\omega)$ and $\widetilde{B}(\omega)$ for two values of $\eta$ (0.993 and 0.93). Note that application of the delay compensation technique allows one to widen the stopband area of the bias characteristic $\widetilde{B}(\omega)$ compared to $B(\omega)$. The same remark applies to bias characteristics associated with the LMS and KF algorithms, irrespective of the values of $\mu$, $\kappa$ and $\lambda_1, \ldots, \lambda_n$. Therefore for slow parameter changes the bias component of the mean-square estimation error will be always smaller for the modified EWLS/LMS/KF algorithms than for the original algorithms

$$\mathrm{E}[||\boldsymbol{\xi}(t)||^2] < \mathrm{E}[||\boldsymbol{\eta}(t)||^2] \tag{27}$$

It should be stressed that, unlike causal estimation schemes, bias reduction is *not* achieved at the cost of increasing the variance component of MSE, which is approximately the same for the compared algorithms.

### *4.3 Mean-square error*

Combining (22) and (23) with (24) and (27) one arrives at

$$\mathrm{E}[||\widetilde{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t)||^2] < \mathrm{E}[||\widehat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t)||^2] \tag{28}$$

which means that the modified EWLS/LMS/KF algorithms will be always more accurate than the original ones. It should be stressed that improvement can be expected for *all* values of the adaptation gain $\gamma$ (although for “large” gains, which entail small delays, it may be marginal).

## 5 Simplified smoothing rules

In order to use (18) or (19), one should either know or perform eigendecomposition of the covariance matrix $\mathbf{\Phi}$. We will derive simplified smoothing rules which allow one to avoid this step – at the cost of decreasing estimation accuracy of the corresponding algorithms.

### *5.1 KF algorithm*

When the eigenvalues $\lambda_1, \ldots, \lambda_n$ are not identical, one may attempt to replace different delays in (19) with the same average delay $\tau^*_{\text{av}} = \text{int}[t^*_{\text{av}}]$ where

$$t^*_{\text{av}} = \frac{\sum_{i=1}^n t^*_i}{n}$$

After replacing $\tau^*_1, \ldots, \tau^*_n$ with $\tau^*_{\text{av}}$ in (19), one arrives at the following simplified smoothing formula

$$\widetilde{\boldsymbol{\theta}}_{\text{av}}(t) = \widehat{\boldsymbol{\theta}}(t + \tau^*_{\text{av}}) \tag{29}$$

which resembles (17).

It turns out that for small adaptation gains the average delay can be evaluated without performing eigendecomposition of $\mathbf{\Phi}$. Actually, observe that for small values of $\kappa$ it holds that $t^*_i \cong 1/(\kappa\sqrt{\lambda_i})$ or $t^*_i \cong 0.7/(\kappa\sqrt{\lambda_i})$. Since $\sum_{i=1}^n (1/\sqrt{\lambda_i}) = \text{tr}\{\mathbf{\Phi}^{-1/2}\}$ and $\mathbf{S}(t) \cong \kappa\mathbf{\Phi}^{-1/2}$, the local estimate of the average delay can be obtained from

$$t^*_{\text{av}} \cong \frac{\text{tr}\{\mathbf{S}(t)\}}{n\kappa^2} \quad \text{or} \quad t^*_{\text{av}} \cong \frac{0.7\text{tr}\{\mathbf{S}(t)\}}{n\kappa^2}$$

Since the matrix $\mathbf{S}(t)$, needed to evaluate $t^*_{\text{av}}$, is recursively updated by the KF algorithm, the corresponding smoothing overlay, equal to 1 multiplication and $n$ additions, is negligible.

### *5.2 LMS algorithm*

To preserve low complexity of the LMS algorithm a simpler estimation scheme is needed than the one described above. To fulfill this requirement one can define the average delay $t_{\text{av}}$ in terms of the average eigenvalue of $\mathbf{\Phi}$, namely

$$t_{\text{av}} = \frac{1 - \mu\lambda_{\text{av}}}{\mu\lambda_{\text{av}}} \quad \text{or} \quad t_{\text{av}} = \frac{0.7}{\mu\lambda_{\text{av}}}$$

where

$$\lambda_{\text{av}} = \frac{\sum_{i=1}^n \lambda_i}{n}$$

Since $\lambda_{\text{av}} = \rho/n$ where $\rho = \sum_{i=1}^n \lambda_i = \text{tr}\{\mathbf{\Phi}\} = \text{E}[||\boldsymbol{\varphi}(t)||^2]$, the local estimate of $\lambda_{\text{av}}$ can be computed as $\widehat{\rho}(t)/n$, where $\widehat{\rho}(t)$ denotes the local, exponentially weighted estimate of $\text{E}[||\boldsymbol{\varphi}(t)||^2]$

$$\widehat{\rho}(t) = \eta\widehat{\rho}(t-1) + (1-\eta)\boldsymbol{\varphi}^{\text{T}}(t)\boldsymbol{\varphi}(t) \tag{30}$$

Similarly as before $\eta$, $0 < \eta < 1$, denotes the forgetting constant.
After replacing in (18) $\tau_1, \ldots, \tau_n$ with $\tau_{\text{av}} = \text{int}[t_{\text{av}}]$ one obtains

$$\widetilde{\boldsymbol{\theta}}_{\text{av}}(t) = \widehat{\boldsymbol{\theta}}(t + \tau_{\text{av}}) \tag{31}$$

When the average delay is computed in the way described above the smoothing overhead is equal to $n+5$ operations only. Hence, the total cost of realizing the simplified LMS-based smoother is equal to $3n+6$ operations per time update.

## 6 Computer simulations

Three simulation experiments were arranged to check properties of the analyzed algorithms.

### *6.1 Example 1*

The simulated two-tap finite impulse response (FIR) system was governed by

$$y(t) = \theta_1(t)u(t) + \theta_2(t)u(t-1) + v(t), \quad v(t) \sim N(0, \sigma_v^2)$$

$$u(t) = au(t-1) + e(t), \quad e(t) \sim N(0, \sigma_e^2), \quad |a| < 1$$

where $\{e(t)\}$ denotes an i.i.d. sequence, independent of $\{v(t)\}$. System parameters were generated using the random walk model

$$\boldsymbol{\theta}(t) = \boldsymbol{\theta}(t-1) + \mathbf{w}(t), \quad \mathbf{w}(t) \sim N(0, \sigma_w^2 \mathbf{I}_2)$$

Note that in the case considered $\boldsymbol{\theta}(t) = [\theta_1(t), \theta_2(t)]^{\mathrm{T}}$, $\boldsymbol{\varphi}(t) = [u(t), u(t-1)]^{\mathrm{T}}$ and (in steady state)

$$\boldsymbol{\Phi} = \sigma_u^2 \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}, \quad \mathbf{Q} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \quad \boldsymbol{\Lambda} = \sigma_u^2 \begin{bmatrix} 1-a & 0 \\ 0 & 1+a \end{bmatrix}$$

where $\sigma_u^2 = \sigma_e^2/(1-a^2)$.
The advantage of this example is that it is fully analytical and hence it allows one to check how well experimental results fit theoretical evaluations.
For any causal estimator $\widehat{\boldsymbol{\theta}}(t)$, designed for the system described above, it holds that (Ravikanth & Meyn, 1999)

$$\mathrm{E}[||\widehat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t)||^2] \geq c_{\text{LTB}} \cong \sigma_w \sigma_v \ \mathrm{tr}\{\boldsymbol{\Phi}^{-1/2}\}$$

where $c_{\text{LTB}}$ denotes the lower tracking bound – a variant of the so-called posterior (or Bayesian) Cramér-Rao bound (PCRB), applicable to systems with random coefficients.

Similarly, for any noncausal estimator $\widetilde{\boldsymbol{\theta}}(t)$ one arrives at the following fundamental limitation (Niedźwiecki, 2007)

$$\mathrm{E}[||\widetilde{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t)||^2] \geq c_{\mathrm{LSB}} \cong \frac{1}{2}\, c_{\mathrm{LTB}}$$

where $c_{\mathrm{LSB}}$ is the lower smoothing bound. The difference $c_{\mathrm{LTB}} - c_{\mathrm{LSB}} \cong 0.5 c_{\mathrm{LTB}}$ specifies the possible margin of improvement achievable by means of smoothing.

Based on the approximation (6), one can derive the following unified expression for the mean-square parameter tracking error

$$\mathrm{E}[||\widehat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t)||^2] \cong \frac{\gamma\sigma_v^2}{2}\, \mathrm{tr}\{\mathbf{A}\} + \frac{\sigma_w^2}{2\gamma}\, \mathrm{tr}\{(\mathbf{A}\boldsymbol{\Phi})^{-1}\} \tag{32}$$

which holds for all three estimation algorithms, provided that system parameters obey the random walk model – see (Guo & Ljung, 1995), (Niedźwiecki, 2000). According to (32), the optimally tuned KF algorithm ($\gamma_{\mathrm{opt}} = \kappa_{\mathrm{opt}} = \sigma_w/\sigma_v$) should yield

$$\mathrm{E}[||\widehat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t)||^2]_{\kappa=\kappa_{\mathrm{opt}}} = \sigma_w\sigma_v\; \mathrm{tr}\{\boldsymbol{\Phi}^{-1/2}\} = c_{\mathrm{LTB}}$$

This is an expected result since, in the case considered, the KF algorithm is the optimal estimation procedure – see comment in Section 1.

The analogous expression for the EWLS and LMS algorithms ($\eta_{\mathrm{opt}} = 0.97$, $\mu_{\mathrm{opt}} = 0.0825$) is

$$\mathrm{E}[||\widehat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t)||^2]_{\eta=\eta_{\mathrm{opt}}} = \mathrm{E}[||\widehat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t)||^2]_{\mu=\mu_{\mathrm{opt}}} = \sigma_w\sigma_v\sqrt{n\mathrm{tr}\{\boldsymbol{\Phi}^{-1}\}}$$

and gives the values larger than $c_{\mathrm{LTB}}$ unless all eigenvalues of $\boldsymbol{\Phi}$ are identical. The following values were adopted: $\sigma_v^2 = 1$, $\sigma_w^2 = 0.0001$, $\sigma_e^2 = 1$ and $a = 0.8$, resulting in $\sigma_u^2 \cong 2.78$, $\lambda_1 = 0.56$, $\lambda_2 = 5.0$, $c_{\mathrm{LTB}} \cong 0.18$ and $c_{\mathrm{LSB}} \cong 0.09$.

Figure 2 shows comparison of the results yielded by different families of algorithms discussed in the paper: EWLS, LMS and KF. Performance of all estimators was quantified in terms of the associated mean-square errors. The MSE of an estimator $\widehat{\boldsymbol{\theta}}(t)$ was evaluated by means of combined time and ensemble averaging. First, for each realization of $\{\boldsymbol{\theta}(t)\}$, $\{u(t)\}$ and $\{v(t)\}$, the following steady state performance index was computed

$$I = \frac{1}{2000}\sum_{t=2001}^{4000} ||\boldsymbol{\theta}(t) - \widehat{\boldsymbol{\theta}}(t)||^2$$

The obtained results were next averaged over 200 realizations of $\{\boldsymbol{\theta}(t)\}$ and 200 realizations of $\{u(t), v(t)\}$ (i.e. over 200×200 realizations altogether). The same set of realizations was used for different algorithms and different values of $\gamma$.

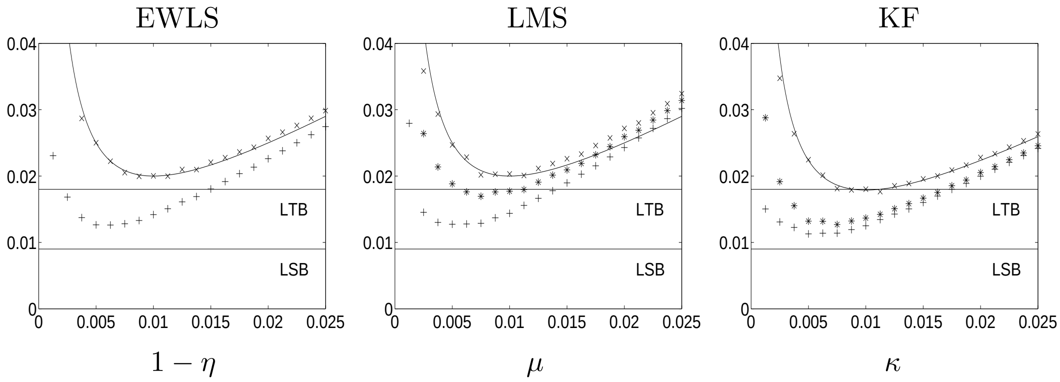


Fig. 2. Dependence of the mean-square parameter estimation errors on the adaptation gains $1-\eta$, $\mu$ and $\kappa$ for the original EWLS, LMS and KF algorithms ($\times$), for the modified algorithms ($+$), and (in the latter two cases) for the simplified versions of the modified algorithms ($*$). The lower tracking bound (LTB) and the lower smoothing bound (LSB) are indicated by horizontal lines. Solid lines show theoretical dependence of MSE on adaptation gains for the original algorithms.

Figure 2 shows the mean-square errors yielded by the original (causal) estimators $\widehat{\theta}(t)$, the modified estimators $\widetilde{\theta}(t)$ and (where applicable) the simplified versions of the modified estimators $\widetilde{\theta}_{\mathrm{av}}(t)$, for 20 equidistant values of the adaptation gain $\gamma$ (i.e. $1-\eta$, $\mu$ or $\kappa$) picked from the interval $[0.00125, 0.025]$. Additionally, it depicts theoretical dependence of MSE on $\gamma$, and the lower estimation bounds for tracking (LTB) and smoothing (LSB). To optimize performance of the modified algorithms, estimation delays were computed using the median-like measure, which is known to be the best choice for random walk parameter variations (when the nominal delays were used instead, the performance was only slightly worse).
The obtained results pretty well illustrate our main points:

- Despite the obvious differences in design principles the EWLS, LMS and KF algoritms perform comparably.
- There is good agreement between the theoretical MSE curves and the results of computer simulations. For the EWLS and KF estimators the fit is satisfactory in the entire range of adaptation gains. For the LMS estimator, discrepancies between theory and experiment occur for "large" values of $\mu$ (which is understandable, as all theoretical results were obtained under the small gain hypothesis).
- All smoothing algorithms perform *uniformly* better than their filtering counterparts, and all offer performance that exceeds the lower tracking bound for a certain range of adaptation gains. The simplified LMS and KF smoothers are less efficient than their "exact" versions.
- The optimally tuned "cheap smoothers" yield mean-square errors that are pretty close (regarding simplicity of the solution) to the lower smoothing bound, achievable by means of using the infinite-lag, genuine Kalman smoother.

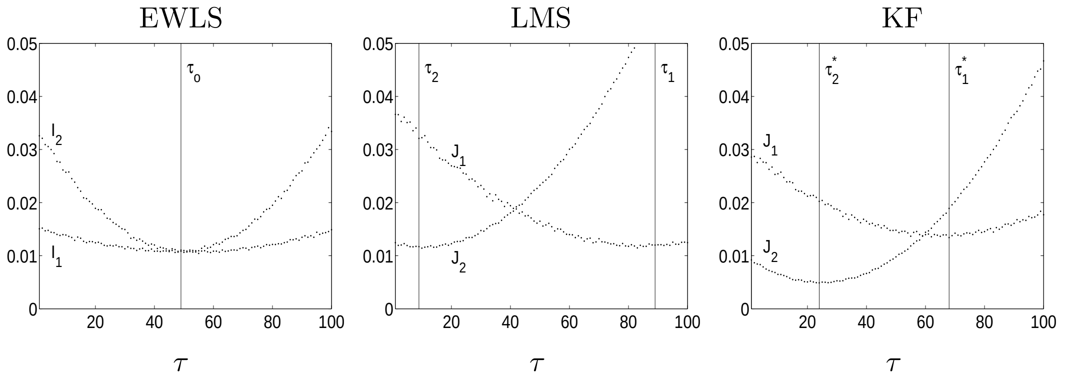


Fig. 3. Dependence of the mean-square parameter matching errors, yielded by the EWLS, LMS and KF algorithms, on the delay $\tau$ for $\eta = 0.98$, $\mu = 0.02$ and $\kappa = 0.02$, respectively; the nominal delays $\tau_o$, $\tau_1$, $\tau_2$ and $\tau_1^*$, $\tau_2^*$ are marked with vertical lines.

### *6.2 Example 2*

In our second simulation experiment sinusoidal parameter changes were enforced

$$\theta_1(t) = 1.5 + \sin(2\pi t/3000), \quad \theta_2(t) = 0.5 + \sin(2\pi t/1500)$$

while the remaining simulation details (input, noise) were kept unchanged. This experiment was certainly less biased as none of the estimation approaches was "handicapped" (for random walk parameter variation the KF algorithms are known to yield the best results). Mean-square errors were computed in the same way as before (200 different realizations of $\{u(t), v(t)\}$ were used to compute ensemble averages).
Figure 3 illustrates the estimation delay structure of the EWLS, LMS and KF estimators. The plots show how the ensemble averages of the parameter matching errors

$$I_i(\tau) = \frac{1}{2000} \sum_{t=2001}^{4000} (\theta_i(t) - \widehat{\theta}_i(t+\tau))^2, \quad i = 1, 2$$

$$J_i(\tau) = \frac{1}{2000} \sum_{t=2001}^{4000} (\beta_i(t) - \widehat{\beta}_i(t+\tau))^2, \quad i = 1, 2$$

depend on the delay $\tau$ for a fixed value of $\gamma$ ($\gamma = 0.02$). In agreement with theory, the smallest matching errors are obtained when the corresponding values of $\tau$ are close to the nominal delays. Note that, exactly as predicted, the two estimation delays are identical for the EWLS estimator and that they differ from each other for the LMS and KF estimators.

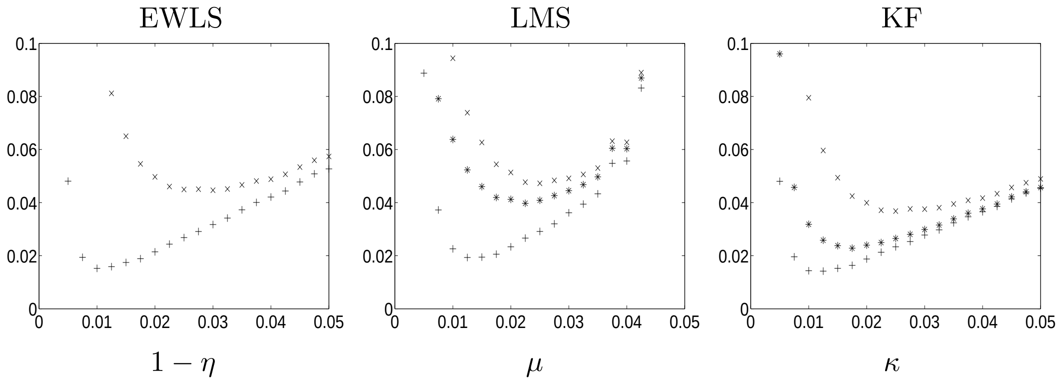


Fig. 4. Dependence of the mean-square parameter estimation errors on the adaptation gains $1-\eta$, $\mu$ and $\kappa$ for the original EWLS, LMS and KF algorithms ($\times$), for the modified algorithms ($+$), and (in the latter two cases) for the simplified versions of the modified algorithms ($*$).

Figure 4 shows the plots of the mean-square estimation errors obtained for different estimation algorithms (no theoretical curves are shown as in this case they are not available). Nominal delays were used for the purpose of smoothing; when the median-like delays were used instead, the performance was only slightly worse. From the qualitative viewpoint the obtained results are similar to those presented earlier. Note that the potential rates of the MSE reduction, achievable by means of smoothing, are higher for the deterministically (slowly) time-varying system than for the system with randomly drifting coefficients.

### *6.3 Example 3*

Assumption (A2) restricts analysis carried out in Sections 3, 4 and 5 to dynamic models with a finite impulse response structure, where the regression vector has the form $\boldsymbol{\varphi}(t) = [u(t-1), \ldots, u(t-n)]^{\mathrm{T}}$. When the class of considered parametrizations is extended to nonstationary ARX (autoregressive with exogenous input) models, where $\boldsymbol{\varphi}(t) = [y(t-1), \ldots, y(t-r), u(t-1), \ldots, u(t-p)]^{\mathrm{T}}$, $r+p=n$, the covariance matrix of $\boldsymbol{\varphi}(t)$ is not time-invariant any more (unless all autoregressive coefficients are constant).
Since EWLS is a local, i.e. finite-memory estimation procedure, one may argue (although it would be difficult to prove this in a formal way) that for the proposed debiasing technique to be effective, only a *local* stationarity of $\{\boldsymbol{\varphi}(t)\}$ is needed. Of course, this remark applies also to the LMS/KF-based smoothers, but unlike the EWLS case, where the smoothing formula does not depend on the eigenstructure of the covariance matrix of $\{\boldsymbol{\varphi}(t)\}$, in order to use the proposed LMS and KF algorithms in a covariance-nonstationary environment, one would need to perform on-line estimation and eigendecomposition of the matrix $\boldsymbol{\Phi}(t)$, which is not a computationally "cheap" solution any more. Even though the simplified LMS/KF-based smoothers are free of the drawback mentioned above, they can be expected to be less efficient than the EWLS-based smoother.

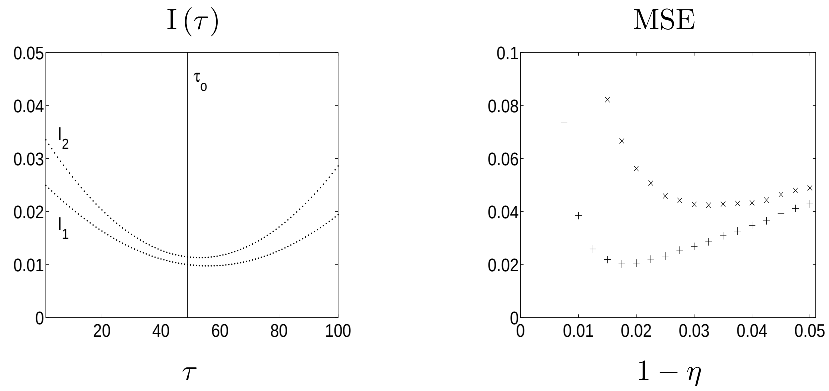


Fig. 5. Left plot shows dependence of the mean-square parameter matching errors, yielded by the EWLS algorithm (applied to identification of an AR signal) on the delay $\tau$ for $\eta = 0.98$; the nominal delay $\tau_o$ is marked with a vertical line. Right plot shows dependence of the mean-square parameter estimation errors on the adaptation gain $1 - \eta$ for the original EWLS algorithm ($\times$) and for its modified version ($+$).

The purpose of our third example was to show that even when the assumption (A2) is violated, the EWLS cheap smoother works pretty well. The identification task was to estimate time-varying autoregressive (AR) coefficients of a nonstationary AR signal governed by

$$y(t) = \theta_1(t)y(t-1) + \theta_2(t)y(t-2) + v(t)$$

where

$$\theta_1(t) = 0.5\sin(2\pi t/1000), \quad \theta_2(t) = 0.5\sin(2\pi t/750)$$

In this case $\boldsymbol{\varphi}(t) = [y(t-1), y(t-2)]^{\mathrm{T}}$.

Figure 5 shows how the ensemble averages of the parameter matching errors $I_1(\tau)$ and $I_2(\tau)$, computed for 200 different realizations of $\{v(t)\}$, depend on the delay $\tau$ for $\eta = 0.98$. Note good agreement between the nominal delay (derived under the assumption that the covariance matrix of $\boldsymbol{\varphi}(t)$ is time-invariant) and the true delay.

The plots of the mean-square estimation errors, displayed in Figure 5, confirm usefulness of the proposed smoothing technique. Application of the simplified versions of the LMS-based and KF-based smoothers also yielded satisfactory results, but the obtained improvements were less significant than those offered by the EWLS-based scheme.

## 7 Practical recommendations

From the three estimation schemes, discussed in the paper, the EWLS approach seems to be the most appealing one. Since the EWLS-based cheap smoother simply delays the estimates provided by the EWLS tracker, it does not incur additional computational costs and can be easily extended to a more general class of system models.

The main advantage of the LMS algorithm is its low computational complexity, linear in the number of estimated coefficients. Note that only the simplified smoothing formula preserves this feature of LMS. Additionally, one should remember that when the eigenvalue disparity index of $\mathbf{\Phi}$ is large, LMS algorithms may suffer from very slow initial convergence, which is a serious drawback in many applications; in contrast with this the EWLS/KF algorithms are known of fast initial convergence.
Even though when identification is carried for a system with randomly drifting coefficients the KF-based smoother outperforms the EWLS-based smoother, this fact is of little practical significance. This is because the RW model usually provides just a crude local approximation to a true parameter variation, i.e. in majority of cases RW is an instrumental model only.
From the two ways of determining the estimation delay we recommend the frequency-domain approach, as it is not focused on any particular class of parameter variations. The median-like delay measure was optimized for random walk parameter trajectories and it usually slightly underestimates the true delay in other cases.

## 8 Conclusion

We have shown that accuracy of the classical EWLS, LMS and KF parameter tracking algorithms can be considerably improved by means of compensating estimation delays. The resulting parameter smoothing algorithms are computationally inexpensive. They can be used in all applications of adaptive filtering which allow one to postpone the model-based decisions, i.e. execute them with a delay.

## References

Bai, E.-W., L.-C. Fu & S. Sastry (1988). Averaging analysis for discrete time and sampled data adaptive systems. *IEEE Trans. Circuits and Systems*, **vol. 35**, 137–148.

Guo, L. & L. Ljung (1995). Performance analysis of general tracking algorithms. *IEEE Trans. Automat. Contr.*, **vol. 40**, 1388–1402.

Haykin, S. (1996). *Adaptive Filter Theory.* Prentice Hall, Englewood Cliffs NJ.

Hedelin, P. (1977). Can the zero-lag filter be a good smoother? *IEEE Trans. on Information Theory*, **vol. 23**, 490–499.

Lewis, F.L. (1986). *Optimal Estimation.* Wiley. New York.

Ljung, L. & & Gunnarsson S. (1990). Adaptive tracking in system identification - a survey. *Automatica*, **vol. 26**, 7–22.

Niedźwiecki, M. (2000). *Identification of Time-varying Processes.* Wiley. New York.

Niedźwiecki, M. (2007). On the lower smoothing bound in identification of time-varying systems. Gdańsk University of Technology, Technical Report # ETI-1/2007, http://www.eti.pg.gda.pl/katedry/ksa/pracownicy/Maciej.Niedzwiecki, (to appear in *Automatica*).

Ravikanth, R. & S.P. Meyn (1999). Bounds on achievable performance in the identification and adaptive control of time-varying systems. *IEEE Trans. Automat. Contr.*, **vol. 44**, 670–682.