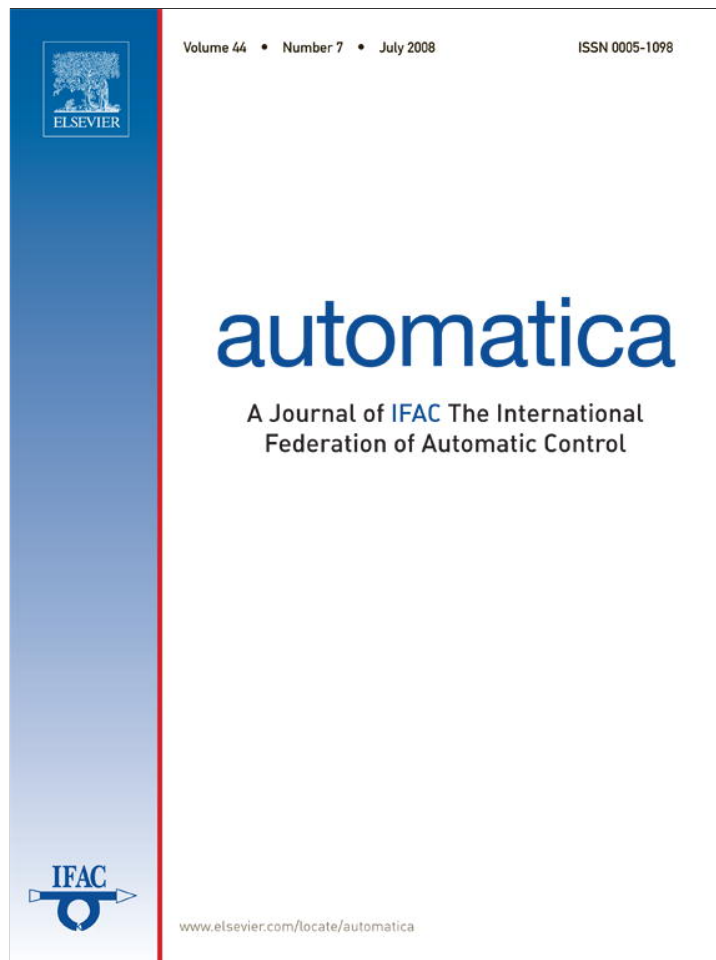


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Optimal and suboptimal smoothing algorithms for identification of time-varying systems with randomly drifting parameters[☆]

Maciej Niedźwiecki^{*}

Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control, Gdańsk University of Technology, Narutowicza 11/12, Gdańsk, Poland

Received 16 May 2007; received in revised form 13 September 2007; accepted 31 October 2007

Available online 19 March 2008

Abstract

Noncausal estimation algorithms, which involve smoothing, can be used for off-line identification of nonstationary systems. Since smoothing is based on both past and future data, it offers increased accuracy compared to causal (tracking) estimation schemes, incorporating past data only. It is shown that efficient smoothing variants of the popular exponentially weighted least squares and Kalman filter-based parameter trackers can be obtained by means of backward-time filtering of the estimates yielded by both algorithms. When system parameters drift according to the random walk model and the adaptation gain is sufficiently small, the properly tuned two-stage Kalman filtering/smoothing algorithm, derived in the paper, achieves the Cramér–Rao type lower smoothing bound, i.e. it is the optimal noncausal estimation scheme. Under the same circumstances performance of the modified exponentially weighted least-squares algorithm is often only slightly inferior to that of the Kalman filter-based smoother.

© 2008 Elsevier Ltd. All rights reserved.

Keywords: System identification; Time-varying processes; Noncausal estimation

1. Introduction

Consider the problem of identification of a linear time-varying discrete-time system governed by

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

$$\boldsymbol{\theta}(t) = \boldsymbol{\theta}(t-1) + \mathbf{w}(t) \quad (2)$$

where $t = \dots, -1, 0, 1, \dots$ denotes normalized (dimensionless) time, $y(t)$ denotes the system output, $\boldsymbol{\varphi}(t) = [\varphi_1(t), \dots, \varphi_n(t)]^T$ is a known regression vector, $v(t)$ denotes white measurement noise, $\boldsymbol{\theta}(t) = [\theta_1(t), \dots, \theta_n(t)]^T$ is the vector of unknown and time-varying system coefficients and, finally, $\mathbf{w}(t)$ denotes the one-step parameter change. We will be further assuming that

(A1) $\{v(t)\}$ is a sequence of zero-mean independent and identically distributed (i.i.d.) random variables with variance $\text{var}[v(t)] = \sigma_v^2$.

(A2) The sequence of regression vectors $\{\boldsymbol{\varphi}(t)\}$, independent of $\{v(t)\}$, is stationary and ergodic with covariance matrix $E[\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^T(t)] = \boldsymbol{\Phi} > 0$.

In this paper we will restrict our attention to two least-squares type parameter estimation frameworks known as exponentially weighted least-squares (EWLS) approach and the Kalman filter (KF) approach — see e.g. Haykin (1996) and Niedźwiecki (2000), among many others.

The EWLS estimates can be obtained by solving the following minimization problem

$$\hat{\boldsymbol{\theta}}(t) = \arg \min_{\boldsymbol{\theta}} \sum_{i=1}^t \eta^{t-i} \left(y(i) - \boldsymbol{\varphi}^T(i)\boldsymbol{\theta} \right)^2$$

where $\eta, 0 < \eta < 1$, denotes the so-called forgetting constant. The resulting recursive algorithm has the following

[☆] A shortened version of this paper will appear in the preprints of the 17th IFAC Congress, 2008, Seoul, Korea. This paper was recommended for publication in revised form by Associate Editor Johan Schoukens under the direction of Editor Torsten Söderström.

^{*} Tel.: +48 58 3472519; fax: +48 58 3415821.

E-mail address: maciekn@eti.pg.gda.pl.

(well-known) form

$$\begin{aligned} \hat{\boldsymbol{\theta}}(t) &= \hat{\boldsymbol{\theta}}(t-1) + \mathbf{R}(t)\boldsymbol{\varphi}(t)\varepsilon(t) \\ \mathbf{R}(t) &= \frac{\boldsymbol{\Sigma}(t-1)}{\eta + \boldsymbol{\varphi}^T(t)\boldsymbol{\Sigma}(t-1)\boldsymbol{\varphi}(t)} \\ \boldsymbol{\Sigma}(t) &= \frac{1}{\eta} \left[\mathbf{I}_n - \mathbf{R}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^T(t) \right] \boldsymbol{\Sigma}(t-1) \end{aligned} \quad (3)$$

where $\varepsilon(t)$ denotes the one-step-ahead prediction error evaluated at instant t , $\boldsymbol{\Sigma}(t) = \left[\sum_{i=1}^t \eta^{t-i} \boldsymbol{\varphi}(i)\boldsymbol{\varphi}^T(i) \right]^{-1}$ is the inverse of the exponentially weighted regression matrix and \mathbf{I}_n denotes the $n \times n$ identity matrix.

All that one needs to assume when deriving the EWLS estimator is that system parameters vary slowly with time — no specific model of parameter variation is used. In contrast with this, the KF approach is based on an explicit (hypothetical) model of parameter changes, namely within this framework one assumes that the estimated coefficients evolve according to the random walk (RW) model, i.e.

(A3) $\{\mathbf{w}(t)\}$ is the sequence of zero-mean i.i.d. random variables, independent of $\{v(t)\}$ and $\{\boldsymbol{\varphi}(t)\}$, with covariance matrix $\mathbf{W} = \text{cov}[\mathbf{w}(t)] = \sigma_w^2 \mathbf{I}_n$.

The optimal, in the mean-square sense, estimator of $\boldsymbol{\theta}(t)$ has the form Lewis (1986)

$$\hat{\boldsymbol{\theta}}(t) = E[\boldsymbol{\theta}(t) | \mathcal{Z}_-(t)]$$

where $\mathcal{Z}_-(t) = \{y(1), \boldsymbol{\varphi}(1), \dots, y(t), \boldsymbol{\varphi}(t)\}$ denotes the observation history available at instant t .

Under (A1)–(A3) and under Gaussian assumptions imposed on $\{v(t)\}$ and $\{\mathbf{w}(t)\}$:

(A4) The sequences $\{v(t)\}$ and $\{\mathbf{w}(t)\}$ are normally distributed.

the conditional mean estimates can be computed recursively using the celebrated Kalman filtering algorithm

$$\begin{aligned} \hat{\boldsymbol{\theta}}(t) &= \hat{\boldsymbol{\theta}}(t-1) + \mathbf{S}(t)\boldsymbol{\varphi}(t)\varepsilon(t) \\ \mathbf{S}(t) &= \frac{\mathbf{P}(t-1)}{1 + \boldsymbol{\varphi}^T(t)\mathbf{P}(t-1)\boldsymbol{\varphi}(t)} \\ \mathbf{P}(t) &= \left[\mathbf{I}_n - \mathbf{S}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^T(t) \right] \mathbf{P}(t-1) + \kappa^2 \mathbf{I}_n \end{aligned} \quad (4)$$

where $\kappa^2 = \sigma_w^2 / \sigma_v^2$. In practical applications, where the RW model can be regarded only as a crude approximation of a true description of parameter changes, the coefficient $\kappa > 0$, similarly as the forgetting constant η in the EWLS algorithm, is treated as a user-dependent “knob”, allowing one to tune the parameter tracking algorithm to the degree of nonstationarity of the identified process. Both algorithms, described above, have the finite-memory property — the influence of past measurements on parameter estimates diminishes with the age of the samples. In the case of the EWLS algorithm the rate at which past data are “forgot” is exponential, while the KF algorithm is sometimes referred to as an adaptive filter with linear forgetting — see Niedźwiecki (2000, Chapter 8.1) for explanation of this terminology.

When $\eta = 1$ and $\kappa = 0$, i.e. when the data forgetting mechanisms are switched off, both algorithms reduce to the classical recursive least-squares (RLS) algorithm and they work identically.

The EWLS and KF algorithms are causal estimation schemes, which means that at each time instant t they provide parameter estimates that are functions of the current and past measurements only. While in most real-time (e.g. control) applications causality is an obvious requirement, there are some important practical problems that can be solved without imposing this constraint. Consider, for example, the problem of adaptive noise cancelling, where an unmeasurable disturbance $d(t)$ is removed from $y(t)$ by exploiting its correlation with an auxiliary, measurable reference signal $r(t)$. Such reference signal is usually recorded using an additional microphone placed in the close vicinity of the noise source (engine, fan etc.). The approach makes use of the fact that the disturbance $d(t)$ can be approximately expressed in the form $\boldsymbol{\varphi}^T(t)\boldsymbol{\theta}$, where $\boldsymbol{\varphi}(t)$ denotes the regression vector made up of samples drawn from the reference signal $r(t)$ and $\boldsymbol{\theta}$ denotes the vector of unknown weights, which should be adjusted adaptively so as to minimize the mean-square cancellation error $E[(y(t) - \boldsymbol{\varphi}^T(t)\boldsymbol{\theta})^2]$. While in the static acoustical environments the optimal weight vector $\boldsymbol{\theta}$ is unknown but constant, when the pickup and/or auxiliary microphones change their position over time, $\boldsymbol{\theta}$ should be regarded as a time-varying quantity. In cases like this, adaptive noise cancellation in a natural way falls into the analysis framework considered in this paper. When cancellation is performed on-line, disturbance estimates are obtained from $\hat{d}(t) = \boldsymbol{\varphi}^T(t)\hat{\boldsymbol{\theta}}(t)$ where $\hat{\boldsymbol{\theta}}(t) = f[\mathcal{Z}_-(t)]$ is the (causal) estimate yielded by the parameter tracking algorithm, such as EWLS or KF. However, when the signals $y(t)$ and $r(t)$ are prerecorded and then processed in an off-line mode (which is typical in the case of surveillance data, “black-box” data etc.), the situation is different. Suppose that the available data record is of the form $\mathcal{Z} = \{y(1), \boldsymbol{\varphi}(1), \dots, y(N), \boldsymbol{\varphi}(N)\}$. Then a more accurate estimate of $d(t)$ can be obtained from $\tilde{d}(t) = \boldsymbol{\varphi}^T(t)\tilde{\boldsymbol{\theta}}(t)$ where the quantity $\tilde{\boldsymbol{\theta}}(t) = g[\mathcal{Z}]$, incorporating all past data points $\mathcal{Z}_-(t)$ and $N - t$ “future” data points $\mathcal{Z}_+(t+1) = \{y(t+1), \boldsymbol{\varphi}(t+1), \dots, y(N), \boldsymbol{\varphi}(N)\}$, is the smoothed (noncausal) estimate of $\boldsymbol{\theta}(t)$.

Smoothing opportunities are seldom taken advantage in system identification. Basically, there are two reasons why this happens. First, the currently available noncausal identification procedures, such as the Rauch–Tung–Striebel smoothing algorithm described e.g. in Meditch (1973), are computationally very expensive. Second, many practitioners seem to be simply unaware of the fact that such noncausal solutions (often deceptively called unrealizable), are perfectly applicable in all off-line adaptive signal processing situations, such as the one discussed above.

In this paper we will demonstrate that very good smoothed estimates of time-varying system parameters can be obtained by means of backward-time filtering of the results yielded by the classical tracking algorithms, such as EWLS and KF. We will show that not only is such approach computationally attractive, but it also leads to estimation algorithms with

very good properties. In particular, when system parameters evolve according to the RW model and the adaptation gain is sufficiently small, the appropriately filtered KF estimates are statistically efficient, i.e. their accuracy reaches the Cramér–Rao type lower smoothing bound established recently (Niedźwiecki, 2007a).

2. Unified analysis framework

When the forgetting constant in the EWLS algorithm is sufficiently close to 1 and when the coefficient κ in the KF algorithm is sufficiently close to 0, both algorithms can be – under (A1) and (A2) – approximately written down in the following standardized form (Guo & Ljung, 1995; Ljung & Gunnarsson, 1990; Niedźwiecki, 2007b)

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \gamma \mathbf{A} \boldsymbol{\varphi}(t) \varepsilon(t) \quad (5)$$

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

$$\text{EWLS : } \gamma = 1 - \eta, \quad \mathbf{A} = \boldsymbol{\Phi}^{-1}$$

$$\text{KF : } \gamma = \kappa, \quad \mathbf{A} = \boldsymbol{\Phi}^{-1/2}$$

and $\boldsymbol{\Phi}^{-1/2} = (\boldsymbol{\Phi}^{1/2})^{-1}$. The matrix $\boldsymbol{\Phi}^{1/2} > 0$ is the (unique) square root of the covariance matrix $\boldsymbol{\Phi}$: $\boldsymbol{\Phi}^{1/2} \boldsymbol{\Phi}^{1/2} = \boldsymbol{\Phi}$. Furthermore, for sufficiently small values of the adaptation gain γ and for sufficiently slow changes in $\boldsymbol{\theta}(t)$ (compared to the changes in $\boldsymbol{\varphi}(t)$), the analysis of (5) can be carried out using the averaging technique (Bai, Fu, & Sastry, 1988), leading to the following approximation

$$\hat{\boldsymbol{\theta}}(t) = (\mathbf{I}_n - \gamma \mathbf{A} \boldsymbol{\Phi}) \hat{\boldsymbol{\theta}}(t-1) + \gamma \mathbf{A} \boldsymbol{\Phi} \boldsymbol{\theta}(t) + \gamma \mathbf{A} \boldsymbol{\varphi}(t) v(t) \quad (6)$$

which will be the basis of our further investigations.

We will analyze and optimize the performance of the EWLS and KF algorithms assuming that system parameters evolve according to the RW model, as specified in (A3) and (A4). Even though often criticized as “unrealistic”, the random walk case is an important benchmark problem in identification of nonstationary systems, since it allows one to derive closed-form expressions which explicitly relate the mean-square estimation errors to the adaptation gain γ and second-order system statistics σ_v^2 , σ_w^2 and $\boldsymbol{\Phi}$. Using such expressions one can compare estimation properties of the analyzed algorithms. Additionally, since for the system with randomly drifting parameters the Cramér–Rao type lower smoothing bound is known, one can also check statistical efficiency of different solutions.

3. EWLS algorithm

3.1. Causal estimation

For the EWLS algorithm it holds that $\gamma = 1 - \eta$, $\mathbf{A} = \boldsymbol{\Phi}^{-1}$ and hence the recursive relationship (6) can be rewritten in an explicit form as

$$\hat{\boldsymbol{\theta}}(t) = F(q^{-1})\boldsymbol{\theta}(t) + F(q^{-1})\mathbf{z}(t) \quad (7)$$

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

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

and $\mathbf{z}(t) = \boldsymbol{\Phi}^{-1} \boldsymbol{\varphi}(t) v(t)$. Note that $\{\mathbf{z}(t)\}$ is a white noise sequence with covariance matrix $\mathbf{Z} = \text{cov}[\mathbf{z}(t)] = \sigma_v^2 \boldsymbol{\Phi}^{-1}$.

According to (7), the parameter estimation error $\Delta \hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t)$ can be expressed in the form

$$\begin{aligned} \Delta \hat{\boldsymbol{\theta}}(t) &= [F(q^{-1}) - 1]\boldsymbol{\theta}(t) + F(q^{-1})\mathbf{z}(t) \\ &= \frac{F(q^{-1}) - 1}{1 - q^{-1}} \mathbf{w}(t) + F(q^{-1})\mathbf{z}(t) \end{aligned}$$

leading to the following expression for the steady-state mean-square error (MSE)

$$\begin{aligned} \mathcal{T}_{\text{EWLS}} &= \text{E}[\|\Delta \hat{\boldsymbol{\theta}}(t)\|^2] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr} \{ \mathbf{S}_{\Delta \hat{\boldsymbol{\theta}}}(\omega) \} d\omega \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{F(e^{-j\omega}) - 1}{1 - e^{-j\omega}} \right|^2 \text{tr} \{ \mathbf{S}_w(\omega) \} d\omega \\ &\quad + \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| F(e^{-j\omega}) \right|^2 \text{tr} \{ \mathbf{S}_z(\omega) \} d\omega. \end{aligned}$$

Since under (A1)–(A3) it holds that $\mathbf{S}_w(\omega) = \mathbf{W} = \sigma_w^2 \mathbf{I}_n$, $\mathbf{S}_z(\omega) = \mathbf{Z} = \sigma_v^2 \boldsymbol{\Phi}^{-1}$, $\forall \omega \in [-\pi, \pi]$, one finally arrives at

$$\mathcal{T}_{\text{EWLS}} = c_1 I_1 [F(e^{-j\omega})] + c_2 I_2 [F(e^{-j\omega})] \quad (8)$$

where

$$\begin{aligned} I_1 [F(e^{-j\omega})] &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{F(e^{-j\omega}) - 1}{1 - e^{-j\omega}} \right|^2 d\omega \\ I_2 [F(e^{-j\omega})] &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| F(e^{-j\omega}) \right|^2 d\omega \end{aligned}$$

and $c_1 = \text{tr}\{\mathbf{W}\} = n\sigma_w^2$, $c_2 = \text{tr}\{\mathbf{Z}\} = \sigma_v^2 \text{tr}\{\boldsymbol{\Phi}^{-1}\}$. The first term on the right-hand side of (8) constitutes the bias component of the mean-square error and the second term – its variance component.

By means of residue calculus (Jury, 1964) one obtains

$$\begin{aligned} I_1 [F(e^{-j\omega})] &= \frac{(1 - \gamma)^2}{\gamma(2 - \gamma)} \cong \frac{1}{2\gamma} \\ I_2 [F(e^{-j\omega})] &= \frac{\gamma}{2 - \gamma} \cong \frac{\gamma}{2} \end{aligned}$$

$$\begin{aligned} \mathcal{T}_{\text{EWLS}} &= \frac{n(1 - \gamma)^2 \sigma_w^2}{\gamma(2 - \gamma)} + \frac{\gamma \sigma_v^2 \text{tr}\{\boldsymbol{\Phi}^{-1}\}}{2 - \gamma} \\ &\cong \frac{n\sigma_w^2}{2\gamma} + \frac{\gamma \sigma_v^2 \text{tr}\{\boldsymbol{\Phi}^{-1}\}}{2} \end{aligned}$$

where approximations are tight for sufficiently small values of γ .

Since the bias component of MSE is inversely proportional to the adaptation gain γ , whereas its variance component is proportional to γ , to obtain good tracking results one should trade off both error terms. The optimal value of γ , i.e. the one

that minimizes $\mathcal{T}_{\text{EWLS}}$, can be obtained by means of solving

$$\frac{\gamma_{\text{opt}}^2}{1 - \gamma_{\text{opt}}} = \frac{c_1}{c_2}. \quad (9)$$

Using the small gain approximations ($\gamma \ll 1$) one arrives at $\gamma_{\text{opt}} \cong \sqrt{c_1/c_2}$ and

$$(\mathcal{T}_{\text{EWLS}})_{\text{min}} = \mathcal{T}_{\text{EWLS}}|_{\gamma=\gamma_{\text{opt}}} \cong \sigma_v \sigma_w \sqrt{n \text{tr}\{\Phi^{-1}\}}. \quad (10)$$

3.2. Noncausal estimation

To obtain the smoothed estimate of $\theta(t)$, further denoted by $\tilde{\theta}(t)$, we will pass the estimates yielded by the EWLS algorithm through an appropriately designed noncausal filter $G(q^{-1}) = \dots + g_{-1}q^{-1} + g_0 + g_1q^1 + \dots$

$$\tilde{\theta}(t) = G(q^{-1})\hat{\theta}(t). \quad (11)$$

For causal estimators such two-stage scheme, combining *explicit* filtering of parameter estimates with *implicit* filtering (7) imposed by the EWLS approach, was proposed and analyzed in Niedźwiecki (1990).

As shown in Niedźwiecki (2007b), a very simple form of smoothing can be obtained by setting $G(q^{-1}) = q^{\tau_o}$, where τ_o is the nominal (low frequency) delay of the filter $F(q^{-1})$. In this case $\tilde{\theta}(t)$ is simply regarded as an estimate of $\theta(t - \tau_o)$, instead of $\theta(t)$. The approach described below is more sophisticated. To make a judicious choice of $G(q^{-1})$ we will examine the effect it has on estimation errors $\Delta\tilde{\theta}(t) = \tilde{\theta}(t) - \theta(t)$. Note that the steady-state mean-square parameter estimation error can be in this case written down in the form

$$\begin{aligned} \mathcal{T}'_{\text{EWLS}} = E[\|\Delta\tilde{\theta}(t)\|^2] &= c_1 I_1 \left[X(e^{-j\omega}) \right] \\ &+ c_2 I_2 \left[X(e^{-j\omega}) \right] \end{aligned} \quad (12)$$

where

$$X(e^{-j\omega}) = F(e^{-j\omega})G(e^{-j\omega}). \quad (13)$$

We will look for a transfer function $X(e^{-j\omega})$ which minimizes $\mathcal{T}'_{\text{EWLS}}$. Observe that (12) can be expressed in the form

$$\mathcal{T}'_{\text{EWLS}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f \left[X(e^{-j\omega}) \right] d\omega \quad (14)$$

$$f[X] = c_1(X - 1)(X^* - 1)HH^* + c_2XX^*$$

where $H(e^{-j\omega}) = 1/(1 - e^{-j\omega})$ and $*$ denotes the complex conjugation. Note that $f[\cdot]$ is a real-valued function.

Even though formally minimization of (14) is a variational calculus problem (since X is a function of ω), when no causality constraints are imposed on $X(q^{-1})$, it can be solved in a pretty straightforward way by minimizing $f[X(e^{-j\omega})]$ for every value of $\omega \in [-\pi, \pi]$. Observe that for a fixed value of ω the quantity $X = X(e^{-j\omega})$ is a complex variable. Hence, minimization of $f[X]$ becomes a standard problem

of optimization of a quadratic cost function (rather than functional). To solve this problem, we will require that

$$\left. \frac{\partial f}{\partial X^*} \right|_{X=X_{\text{opt}}} = 0. \quad (15)$$

The left-hand side of (15) is a symbolic differentiation, where X and X^* are regarded as independent variables (see e.g. Appendix B in Haykin (1996)). Using standard rules of differentiation with respect to complex-valued variables one obtains $X_{\text{opt}} = c_1HH^*/(c_2 + c_1HH^*)$ or equivalently

$$X_{\text{opt}}(q^{-1}) = \frac{c_1/c_2}{c_1/c_2 + (1 - q^{-1})(1 - q)}. \quad (16)$$

Since the right-hand side of (16) can be rewritten in the form

$$\frac{\gamma^2}{[1 - (1 - \gamma)q^{-1}][1 - (1 - \gamma)q]}$$

where γ is the solution of $\gamma^2/(1 - \gamma) = c_1/c_2$, one finally arrives at (cf. (9))

$$X_{\text{opt}}(q^{-1}) = F_{\text{opt}}(q^{-1})F_{\text{opt}}(q) \quad (17)$$

$$F_{\text{opt}}(q^{-1}) = \frac{\gamma_{\text{opt}}}{1 - (1 - \gamma_{\text{opt}})q^{-1}}.$$

According to (13) and (17), when the EWLS algorithm is optimally tuned, the best smoothing results can be obtained by choosing

$$G_{\text{opt}}(q^{-1}) = \frac{X_{\text{opt}}(q^{-1})}{F_{\text{opt}}(q^{-1})} = F_{\text{opt}}(q) = \frac{\gamma_{\text{opt}}}{1 - (1 - \gamma_{\text{opt}})q}. \quad (18)$$

Note that the filter $G_{\text{opt}}(q^{-1})$ is *anticausal*, which means that the smoothed estimates $\tilde{\theta}(t)$ can be obtained by backward-time filtering of the estimates yielded by the EWLS algorithm. This can be done recursively using the following simple formula

$$\begin{aligned} \tilde{\theta}(N) &= \hat{\theta}(N) \\ \tilde{\theta}(t) &= (1 - \gamma)\tilde{\theta}(t + 1) + \gamma\hat{\theta}(t) \\ t &= N - 1, \dots, 1 \end{aligned} \quad (19)$$

where the optimal gain γ_{opt} , usually not known *a priori*, was replaced with γ – the gain used in the tracking algorithm. Making such a choice is equivalent to adopting $G(q^{-1}) = F(q) = \gamma/(1 - (1 - \gamma)q)$. Since in the case considered $X(e^{-j\omega}) = |F(e^{-j\omega})|^2$ and

$$\begin{aligned} I_1 \left[X(e^{-j\omega}) \right] &= \frac{2(1 - \gamma)^2}{\gamma(2 - \gamma)^3} \cong \frac{1}{4\gamma} \\ I_2 \left[X(e^{-j\omega}) \right] &= \frac{\gamma(2 - 2\gamma + \gamma^2)}{(2 - \gamma)^3} \cong \frac{\gamma}{4} \end{aligned}$$

one arrives at

$$\mathcal{T}'_{\text{EWLS}} \cong \frac{n\sigma_w^2}{4\gamma} + \frac{\gamma\sigma_v^2 \text{tr}\{\Phi^{-1}\}}{4} \cong \frac{1}{2} \mathcal{T}_{\text{EWLS}} \quad (20)$$

which means that, irrespective of the choice of γ , the proposed smoothing procedure allows one to reduce the mean-square parameter estimation errors by the factor of 2. Of course, the same applies to the best achievable performance

$$\begin{aligned} (\mathcal{T}'_{\text{EWLS}})_{\min} &= \mathcal{T}'_{\text{EWLS}}|_{\gamma=\gamma_{\text{opt}}} \cong \frac{1}{2}\sigma_v\sigma_w\sqrt{n\text{tr}\{\Phi^{-1}\}} \\ &\cong \frac{1}{2}(\mathcal{T}_{\text{EWLS}})_{\min}. \end{aligned} \quad (21)$$

Remark. Note that the optimal smoothing (noncausal) gain is identical with the optimal tracking (causal) gain. This has important practical implications. When γ_{opt} is not known, which is a typical situation in practice, adaptive optimization of the tracking algorithm will also guarantee performance optimization of the two-step smoothing procedure. Optimization of the adaptation gain is possible using sequential or parallel estimation techniques (Niedźwiecki, 2000). The first case uses a single tracking algorithm, equipped with an adjustable gain. The second case takes several algorithms, with different gains, runs them in parallel and compares them according to their predictive abilities.

4. KF algorithm

4.1. Causal estimation

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

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

where $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_n\}$ is a diagonal matrix made up of the eigenvalues of Φ .

For the KF algorithm it holds that $\gamma = \kappa$ and $\mathbf{A} = \Phi^{-1/2}$. In this case the relationship (6) can be rewritten in a closed form as (Niedźwiecki, 2007b)

$$\hat{\boldsymbol{\theta}}(t) \cong \mathbf{Q}\mathbf{F}(q^{-1})\mathbf{Q}^T\boldsymbol{\theta}(t) + \mathbf{Q}\Lambda^{-1/2}\mathbf{F}(q^{-1})\mathbf{Q}^T\mathbf{z}(t) \quad (22)$$

where

$$\mathbf{F}(q^{-1}) = \text{diag}\{F_1(q^{-1}), \dots, F_n(q^{-1})\}$$

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

and $\mathbf{z}(t) = \Phi^{-1/2}\boldsymbol{\varphi}(t)v(t)$, $\mathbf{Z} = \text{cov}[\mathbf{z}(t)] = \sigma_v^2\mathbf{I}_n$.

Based on (22) the following error model can be derived

$$\Delta\hat{\boldsymbol{\theta}}(t) \cong \mathbf{Q}\frac{\mathbf{F}(q^{-1}) - \mathbf{I}_n}{1 - q^{-1}}\mathbf{Q}^T\mathbf{w}(t) + \mathbf{Q}\Lambda^{-1/2}\mathbf{F}(q^{-1})\mathbf{Q}^T\mathbf{z}(t)$$

leading to

$$\begin{aligned} \mathcal{T}_{\text{KF}} &= E[\|\Delta\hat{\boldsymbol{\theta}}(t)\|^2] = \frac{1}{2\pi} \\ &\times \int_{-\pi}^{\pi} \text{tr} \left\{ \mathbf{Q} \frac{\mathbf{F}(e^{-j\omega}) - \mathbf{I}_n}{1 - e^{-j\omega}} \mathbf{Q}^T \mathbf{S}_w(\omega) \mathbf{Q} \frac{\mathbf{F}(e^{j\omega}) - \mathbf{I}_n}{1 - e^{j\omega}} \mathbf{Q}^T \right\} d\omega \\ &+ \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{tr} \left\{ \mathbf{Q} \Lambda^{-1/2} \mathbf{F}(e^{-j\omega}) \mathbf{Q}^T \mathbf{S}_z(\omega) \mathbf{Q} \mathbf{F}(e^{j\omega}) \Lambda^{-1/2} \mathbf{Q}^T \right\} d\omega. \end{aligned}$$

Since in the case considered $\mathbf{S}_w(\omega) = \mathbf{W} = \sigma_w^2\mathbf{I}_n$, $\mathbf{S}_z(\omega) = \mathbf{Z} = \sigma_v^2\mathbf{I}_n$, $\forall \omega \in [-\pi, \pi]$, one arrives at

$$\begin{aligned} \mathcal{T}_{\text{KF}} &\cong \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{i=1}^n \left[\sigma_w^2 \left| \frac{F_i(e^{-j\omega}) - 1}{1 - e^{-j\omega}} \right|^2 \right. \\ &\quad \left. + \frac{\sigma_v^2}{\lambda_i} |F_i(e^{-j\omega})|^2 \right] d\omega \\ &= c_1 \sum_{i=1}^n I_1[F_i(e^{-j\omega})] + \sum_{i=1}^n c_{2i} I_2[F_i(e^{-j\omega})] \end{aligned} \quad (23)$$

where $c_1 = \text{tr}\{\mathbf{W}\} = n\sigma_w^2$ and $c_{2i} = \sigma_v^2/\lambda_i$, $i = 1, \dots, n$.

For small adaptation gains one obtains

$$I_1[F_i(e^{-j\omega})] \cong \frac{\kappa\sqrt{\lambda_i}}{2}, \quad I_2[F_i(e^{-j\omega})] \cong \frac{1}{2\kappa\sqrt{\lambda_i}}$$

and therefore (23) can be expressed in the form

$$\mathcal{T}_{\text{KF}} = \sum_{i=1}^n \left[\frac{\sigma_w^2}{2\kappa\sqrt{\lambda_i}} + \frac{\kappa\sigma_v^2}{2\sqrt{\lambda_i}} \right] = \frac{\text{tr}\{\Phi^{-1/2}\}}{2} \left(\frac{\sigma_w^2}{\kappa} + \kappa\sigma_v^2 \right)$$

where the last transition stems from the fact that $\sum_{i=1}^n 1/\sqrt{\lambda_i} = \text{tr}\{\Lambda^{-1/2}\} = \text{tr}\{\Phi^{-1/2}\}$.

Optimization of \mathcal{T}_{KF} is straightforward — the minimum value of the mean-square parameter estimation error is achieved for

$$\kappa = \kappa_{\text{opt}} = \frac{\sigma_w}{\sigma_v} \quad (24)$$

which is an expected result since, under (24), the algorithm (4) is a “true” Kalman filter, i.e. the optimal parameter tracking procedure. Note that

$$(\mathcal{T}_{\text{KF}})_{\min} = \mathcal{T}_{\text{KF}}|_{\kappa=\kappa_{\text{opt}}} \cong \sigma_v\sigma_w\text{tr}\{\Phi^{-1/2}\}. \quad (25)$$

4.2. Noncausal estimation

In agreement with Niedźwiecki (2007b), the following scheme will be used to obtain smoothed KF estimates

$$\begin{aligned} \hat{\boldsymbol{\beta}}(t) &= \mathbf{Q}^T\hat{\boldsymbol{\theta}}(t), \quad \tilde{\boldsymbol{\beta}}(t) = \mathbf{G}(q^{-1})\hat{\boldsymbol{\beta}}(t), \\ \tilde{\boldsymbol{\theta}}(t) &= \mathbf{Q}\tilde{\boldsymbol{\beta}}(t) \end{aligned} \quad (26)$$

where $\mathbf{G}(q^{-1}) = \text{diag}\{G_1(q^{-1}), \dots, G_n(q^{-1})\}$ and $G_i(q^{-1})$, $i = 1, \dots, n$ denote transfer functions of the appropriately designed noncausal filters.

According to (26) it holds that $\tilde{\boldsymbol{\theta}}(t) = \mathbf{Q}\mathbf{G}(q^{-1})\mathbf{Q}^T\hat{\boldsymbol{\theta}}(t)$. Combining this result with (22) and noting that the matrices Λ , $\mathbf{F}(q^{-1})$ and $\mathbf{G}(q^{-1})$ are diagonal and hence they commute, one arrives at

$$\tilde{\boldsymbol{\theta}}(t) \cong \mathbf{Q}\mathbf{X}(q^{-1})\mathbf{Q}^T\boldsymbol{\theta}(t) + \mathbf{Q}\Lambda^{-1/2}\mathbf{X}(q^{-1})\mathbf{Q}^T\mathbf{z}(t) \quad (27)$$

where $\mathbf{X}(q^{-1}) = \text{diag}\{X_1(q^{-1}), \dots, X_n(q^{-1})\}$ and $X_i(q^{-1}) = F_i(q^{-1})G_i(q^{-1})$, $i = 1, \dots, n$.

Based on (27) the mean-square estimation error can be obtained in the form

$$\begin{aligned} \mathcal{T}'_{\text{KF}} &= E[\|\Delta\tilde{\boldsymbol{\theta}}(t)\|^2] \cong c_1 \sum_{i=1}^n I_1[X_i(e^{-j\omega})] \\ &+ \sum_{i=1}^n c_{2i} I_2[X_i(e^{-j\omega})]. \end{aligned} \quad (28)$$

Minimization of \mathcal{T}'_{KF} can be carried out in an analogous way as minimization of $\mathcal{T}'_{\text{EWLS}}$, performed in Section 3.2. Using the same technique one can show that $X_i^{\text{opt}}(q^{-1}) = F_i^{\text{opt}}(q^{-1})F_i^{\text{opt}}(q)$ where

$$F_i^{\text{opt}}(q^{-1}) = \frac{\kappa_{\text{opt}}\sqrt{\lambda_i}}{1 - (1 - \kappa_{\text{opt}}\sqrt{\lambda_i})q^{-1}}, \quad i = 1, \dots, n$$

and $\kappa_{\text{opt}} = \sigma_w/\sigma_v$ denotes the optimal tracking gain. This leads to $G_i^{\text{opt}}(q^{-1}) = F_i^{\text{opt}}(q)$ and to the following backward-time filtering scheme, analogous to (19)

$$\begin{aligned} \tilde{\boldsymbol{\beta}}_i(N) &= \hat{\boldsymbol{\beta}}_i(N) \\ \tilde{\boldsymbol{\beta}}_i(t) &= (1 - \kappa\sqrt{\lambda_i})\tilde{\boldsymbol{\beta}}_i(t+1) + \kappa\sqrt{\lambda_i}\hat{\boldsymbol{\beta}}_i(t) \\ t &= N-1, \dots, 1, i = 1, \dots, n \end{aligned} \quad (29)$$

obtained after adopting $G_i(q^{-1}) = F_i(q)$. Similarly as before it holds that $\mathcal{T}'_{\text{KF}} \cong \mathcal{T}_{\text{KF}}/2$ and

$$(\mathcal{T}'_{\text{KF}})_{\min} = \mathcal{T}'_{\text{KF}}|_{\kappa=\kappa_{\text{opt}}} \cong \frac{1}{2}\sigma_v\sigma_w\text{tr}\{\boldsymbol{\Phi}^{-1/2}\}. \quad (30)$$

According to Niedźwiecki (2007a), for any estimator $\hat{\boldsymbol{\theta}}(t)$ of $\boldsymbol{\theta}(t)$, including all noncausal estimators, it holds that (under assumption (A1)–(A4))

$$E\left[\Delta\hat{\boldsymbol{\theta}}(t)\Delta\hat{\boldsymbol{\theta}}^T(t)\right] \geq \frac{1}{2}\sigma_v\sigma_w\boldsymbol{\Phi}^{-1/2} = \mathbf{B}_{\text{LSB}} \quad (31)$$

where \mathbf{B}_{LSB} denotes the lower smoothing bound (LSB). Note that $(\mathcal{T}'_{\text{KF}})_{\min} \cong \text{tr}\{\mathbf{B}_{\text{LSB}}\}$. Furthermore, it can be shown that (see Appendix)

$$\left(E[\Delta\tilde{\boldsymbol{\theta}}(t)\Delta\tilde{\boldsymbol{\theta}}^T(t)]\right)_{\min} = E[\Delta\tilde{\boldsymbol{\theta}}(t)\Delta\tilde{\boldsymbol{\theta}}^T(t)]_{\kappa=\kappa_{\text{opt}}} \cong \mathbf{B}_{\text{LSB}} \quad (32)$$

which means that the optimally tuned two-step KF algorithm is, in the steady state and under (A1)–(A4), a statistically efficient estimation procedure, achieving the same performance as the – computationally much more involved – Rauch–Tung–Striebel smoother (i.e. the *genuine* Kalman smoother) designed for the system (1) and (2). In Section 6 we will show that simulation experiments fully confirm this claim.

Remark. Observe that recursions (29) can be written down in a more compact form as

$$\tilde{\boldsymbol{\beta}}(t) = (\mathbf{I}_n - \kappa\boldsymbol{\Lambda}^{1/2})\tilde{\boldsymbol{\beta}}(t+1) + \kappa\boldsymbol{\Lambda}^{1/2}\hat{\boldsymbol{\beta}}(t). \quad (33)$$

Premultiplying both sides of (33) with \mathbf{Q} , and noting that $\hat{\boldsymbol{\beta}}(t) = \mathbf{Q}\boldsymbol{\theta}(t)$, $\tilde{\boldsymbol{\theta}}(t) = \mathbf{Q}\tilde{\boldsymbol{\beta}}(t)$, one obtains

$$\tilde{\boldsymbol{\theta}}(t) = \mathbf{Q}(\mathbf{I}_n - \kappa\boldsymbol{\Lambda}^{1/2})\mathbf{Q}^T\tilde{\boldsymbol{\theta}}(t+1) + \kappa\mathbf{Q}\boldsymbol{\Lambda}^{1/2}\mathbf{Q}^T\hat{\boldsymbol{\theta}}(t).$$

Finally, since $\mathbf{Q}\boldsymbol{\Lambda}^{1/2}\mathbf{Q}^T = \boldsymbol{\Phi}^{1/2}$, one arrives at the following equivalent form of the filtering procedure

$$\begin{aligned} \tilde{\boldsymbol{\theta}}(N) &= \hat{\boldsymbol{\theta}}(N) \\ \tilde{\boldsymbol{\theta}}(t) &= (\mathbf{I}_n - \kappa\boldsymbol{\Phi}^{1/2})\tilde{\boldsymbol{\theta}}(t+1) + \kappa\boldsymbol{\Phi}^{1/2}\hat{\boldsymbol{\theta}}(t) \\ t &= N-1, \dots, 1. \end{aligned} \quad (34)$$

The modified filtering scheme (34) allows one to avoid factorization of the covariance matrix $\boldsymbol{\Phi}$ — the step that must be performed in order to use (26) and (29). Actually, since in the steady state it holds that (Niedźwiecki, 2007b) $\mathbf{S}(t) \cong \kappa\boldsymbol{\Phi}^{-1/2}$, where $\mathbf{S}(t)$ is the matrix recursively updated by the KF algorithm, a good estimate of $\boldsymbol{\Phi}^{1/2}$ can be obtained from

$$\hat{\boldsymbol{\Phi}}^{1/2}(t) = \kappa\mathbf{S}^{-1}(t). \quad (35)$$

When the regression sequence is wide-sense stationary and ergodic (as assumed here), inversion of $\mathbf{S}(t)$ has to be performed only once, e.g. in the middle of the analysis interval ($t = N/2$).

5. Comparison of the EWLS and KF approaches

Using the Cauchy–Schwartz inequality one obtains

$$\text{tr}\{\boldsymbol{\Phi}^{-1/2}\} = \sum_{i=1}^n \frac{1}{\sqrt{\lambda_i}} \leq \left(n \sum_{i=1}^n \frac{1}{\lambda_i}\right)^{1/2} = \sqrt{n\text{tr}\{\boldsymbol{\Phi}^{-1}\}}$$

which leads to (cf. (10), (21), (25) and (30))

$$(\mathcal{T}_{\text{KF}})_{\min} \leq (\mathcal{T}_{\text{EWLS}})_{\min}, \quad (\mathcal{T}'_{\text{KF}})_{\min} \leq (\mathcal{T}'_{\text{EWLS}})_{\min}$$

where equality holds iff all eigenvalues are identical. Of course, both relationships are a straightforward consequence of the fact that under (A1)–(A4) the optimally tuned KF algorithms are the best, from the statistical viewpoint, tracking/smoothing procedures. This optimality feature of the KF algorithms should not be overemphasized. One should remember that optimality holds for a specific class of systems with randomly drifting coefficients. When system parameters do not change according to the random walk model, the KF tracking/smoothing algorithms are neither more nor less appropriate than the analogous EWLS algorithms — see Section 5.5 and 7.7 in Niedźwiecki (2000) for a more detailed discussion of this problem.

Comparison of numerical complexity of the proposed algorithms favors the EWLS approach. The efficient mechanization of the EWLS-based smoother (3) + (19) requires $2n^2 + 5n$ multiply-add operations per time update in the tracking (forward-time) loop, and only $2n$ operations per time update in the smoothing (backward-time) loop. The analogous counts for the KF-based smoother (4) + (34) are: $1.5n^2 + 5.5n$ operations and $2n^2$ operations, respectively (the cost of evaluating $\boldsymbol{\Phi}^{1/2}$ was not included, since such operation is performed only once). In contrast with this, the Rauch–Tung–Striebel smoother requires $O(n^3)$ operations per time update.

6. Computer simulations

Three simulation experiments, adopted from Niedźwiecki (2007b), were performed to check properties of the analyzed algorithms.

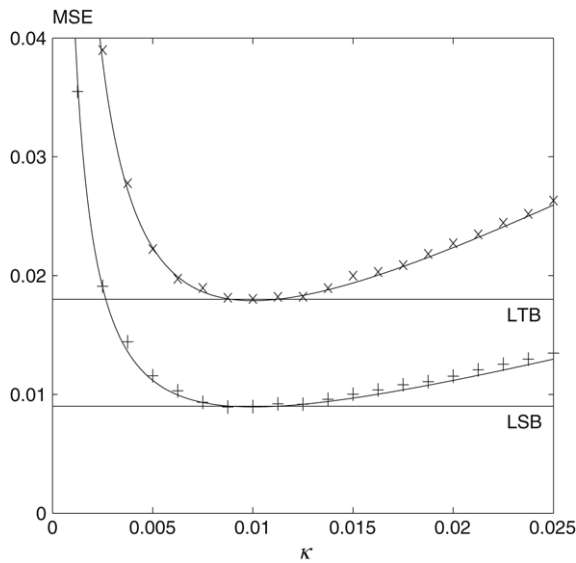


Fig. 1. Dependence of the mean-square parameter estimation error, observed for a nonstationary FIR system with randomly drifting coefficients (slow variations), on the adaptation gain κ for the KF tracker (\times) and the KF-based smoother ($+$). The lower tracking bound (LTB) and the lower smoothing bound (LSB) are indicated by horizontal lines. Solid lines show theoretical dependence of MSE on κ for both algorithms.

6.1. Example 1

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

$$y(t) = \theta_1(t)u(t) + \theta_2(t)u(t - 1) + v(t), \quad v(t) \sim \mathcal{N}(0, 1)$$

$$u(t) = 0.8u(t - 1) + e(t), \quad e(t) \sim \mathcal{N}(0, 1)$$

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

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

where $\theta(t) = [\theta_1(t), \theta_2(t)]^T$ and $\sigma_w = 0.01$.

Performance of the compared estimators was quantified in terms of the associated mean-square errors. The MSE of an estimator $\hat{\theta}(t)$ was evaluated by means of combined time and ensemble averaging. First, for each realization of $\{\theta(t)\}$, $\{u(t)\}$ and $\{v(t)\}$, the following steady-state performance index was computed

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

The obtained results were next averaged over 200 realizations of $\{\theta(t)\}$ 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 γ and κ .

Fig. 1 shows results obtained for the KF tracker and for the KF-based smoother derived in the paper. Note very good fit between the theoretical MSE curves and the results of computer simulations. In agreement with theory, the optimally tuned smoothing algorithm achieves the lower smoothing

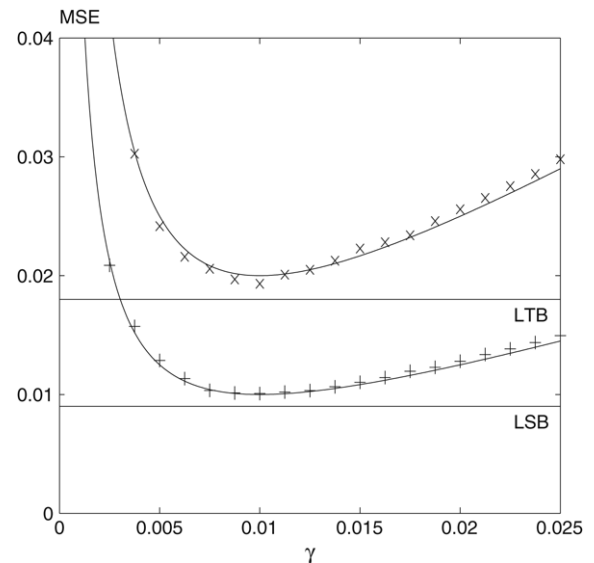


Fig. 2. Dependence of the mean-square parameter estimation error, observed for a nonstationary FIR system with randomly drifting coefficients (slow variations), on the adaptation gain γ for the EWLS tracker (\times) and the EWLS smoother ($+$). The lower tracking bound (LTB) and the lower smoothing bound (LSB) are indicated by horizontal lines. Solid lines show theoretical dependence of MSE on γ for both algorithms.

bound, which limits performance of any (causal or noncausal) estimation scheme.

The analogous results for the EWLS filter and smoother are shown in Fig. 2. Note that the optimally tuned EWLS smoother yields mean-square errors that are well below the lower tracking bound and pretty close to the lower smoothing bound. Hence, in the case considered, it may be deservedly called suboptimal.

Figs. 3 and 4 show what happens when, in the same system, parameters change at a *very fast* rate, namely when the variance σ_w^2 is set to 0.01, i.e. $\sigma_w = 0.1$. This forces one to work with pretty large adaptation gains (note that $\gamma > 0.1$ entails $\eta < 0.9$ — such small values of the forgetting factor are seldom used in practice). The quantitative effect that can be seen under such harsh conditions is discrepancy between the theoretical curves and the measured errors for large values of adaptation gains, something that is expected, as all analytical expressions were derived under the small gain hypothesis. On the qualitative level, however, nothing has changed: smoothing still allows one to reduce the mean-square errors by (approximately) the factor of 2.

6.2. Example 2

In our second simulation experiment sinusoidal parameter changes were enforced (see Fig. 5):

$$\theta_1(t) = 1.5 + \sin(2\pi t/3000)$$

$$\theta_2(t) = 0.5 + \sin(2\pi t/1500).$$

The remaining simulation details (input, noise) were kept unchanged. 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). This experiment was intended to check how the proposed algorithms cope with

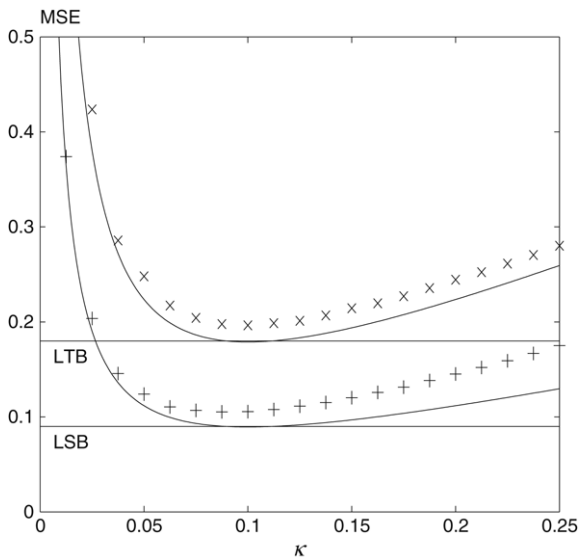


Fig. 3. Dependence of the mean-square parameter estimation error, observed for a nonstationary FIR system with randomly drifting coefficients (fast variations), on the adaptation gain κ for the KF tracker (\times) and the KF-based smoother ($+$). The lower tracking bound (LTB) and the lower smoothing bound (LSB) are indicated by horizontal lines. Solid lines show theoretical dependence of MSE on κ for both algorithms.

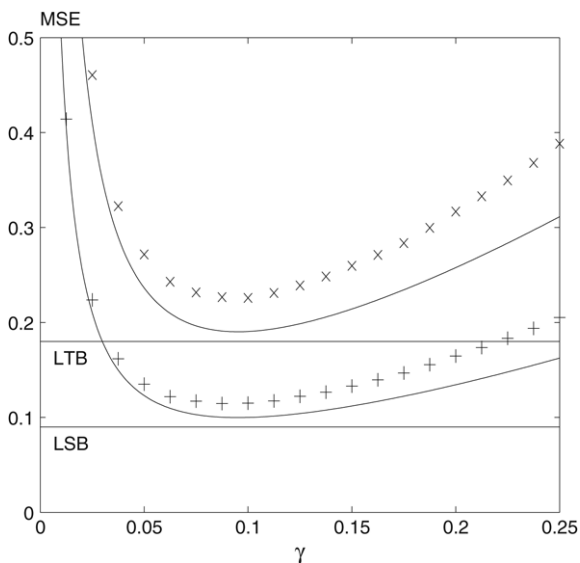


Fig. 4. Dependence of the mean-square parameter estimation error, observed for a nonstationary FIR system with randomly drifting coefficients (fast variations), on the adaptation gain γ for the EWLS tracker (\times) and the EWLS smoother ($+$). The lower tracking bound (LTB) and the lower smoothing bound (LSB) are indicated by horizontal lines. Solid lines show theoretical dependence of MSE on γ for both algorithms.

deterministically time-varying systems, i.e. systems that clearly violate assumption (A3).

Figs. 6 and 7 show the plots of the mean-square estimation errors obtained for different KF and EWLS estimation algorithms (no theoretical curves are shown as in this case they are not available). 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

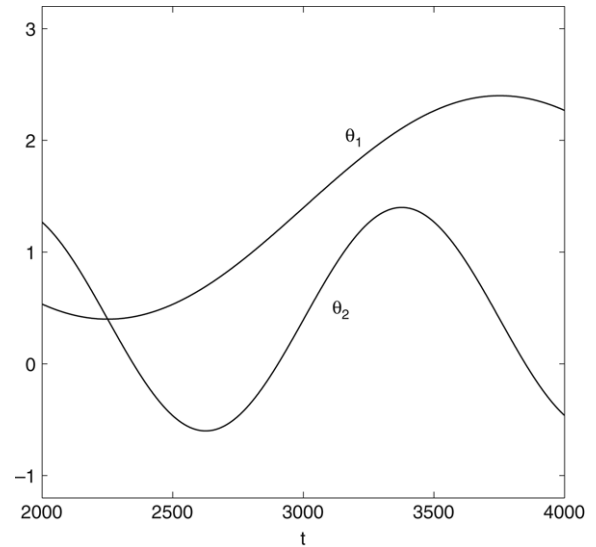


Fig. 5. Evolution of system parameters.

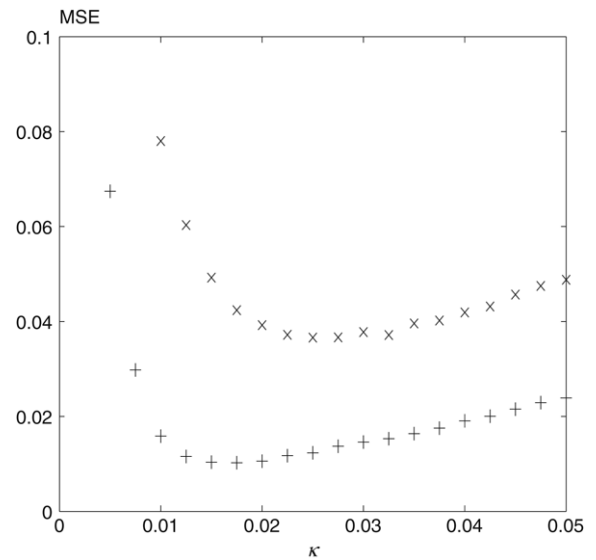


Fig. 6. Dependence of the mean-square parameter estimation error, observed for a nonstationary FIR system with sinusoidally time-varying coefficients, on the adaptation gain κ for the KF tracker (\times) and the KF-based smoother ($+$).

smoothing, are higher for the deterministically (slowly) time-varying system than for the system with randomly drifting coefficients. Note also that in this case the KF-based smoother shows no advantage over the EWLS smoother, which is pretty understandable as it operates under “nonstandard” conditions.

6.3. Example 3

The smoothing formulas (19) and (34) were derived under the assumption that the covariance matrix of the regression vector is time-invariant. While this assumption is fulfilled for nonstationary FIR (finite-impulse response) systems subject to a wide-sense stationary excitation, it is clearly *not* satisfied for nonstationary ARX (autoregressive with exogenous inputs) systems. We will show that despite this formal limitation, the EWLS-based smoother – observe that the smoothing formula

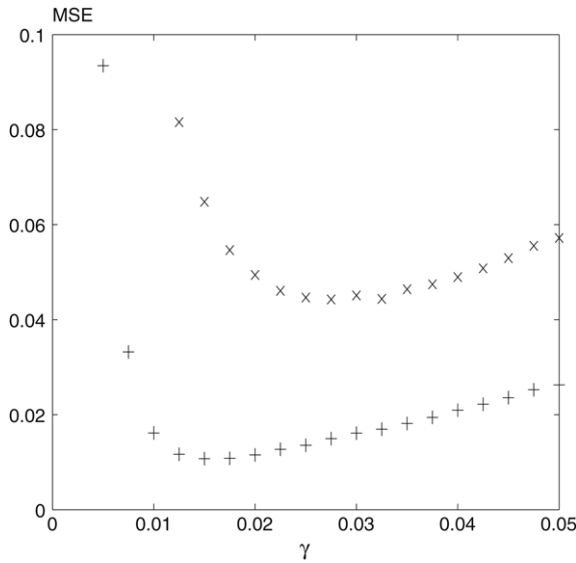


Fig. 7. Dependence of the mean-square parameter estimation error, observed for a nonstationary FIR system with sinusoidally time-varying coefficients, on the adaptation gain γ for the EWLS tracker (\times) and the EWLS smoother ($+$).

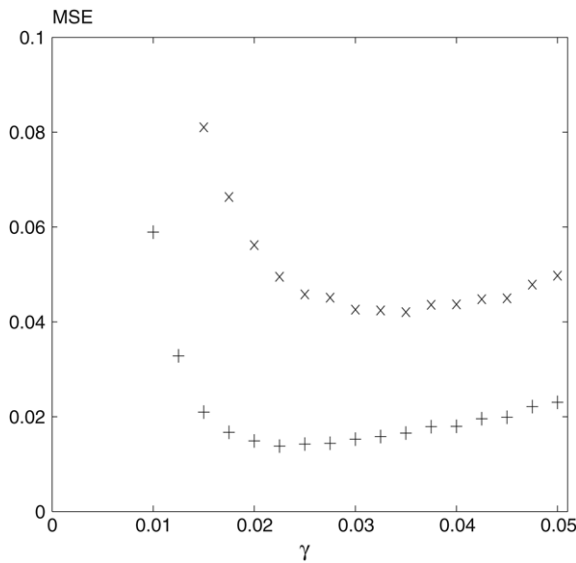


Fig. 8. Dependence of the mean-square parameter estimation error, observed for a nonstationary AR signal, on the adaptation gain γ for the EWLS tracker (\times) and the EWLS smoother ($+$).

(19) does not depend on the covariance structure of $\boldsymbol{\varphi}(t)$ – can be successfully used to identify ARX processes. Fig. 8 shows the plots of the mean-square parameter estimation errors (for 200 realizations of the driving noise sequence) obtained in the course of identification of a second-order autoregressive signal governed by

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

$$v(t) \sim \mathcal{N}(0, 1) \tag{36}$$

where

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

Note that in this case the covariance matrix of the regression vector $\boldsymbol{\varphi}(t) = [y(t-1), y(t-2)]^T$, which depends on the autoregressive coefficients $\theta_1(t)$ and $\theta_2(t)$, is time-varying. Even though for the process (36) the assumption (A2) is clearly violated, the EWLS-based smoother yields very good results, allowing one to achieve at least two-fold reduction of estimation errors for the considered range of adaptation gains γ .

7. Conclusion

We have considered the problem of identification of a linear dynamic system with randomly varying coefficients. When identification can be performed off-line, which is allowed in certain applications, estimation of time-varying system parameters can be based on both past and future data samples. Such noncausal (smoothing) estimation schemes offer considerable performance improvements compared with their causal (tracking) variants. We have shown theoretically, and confirmed by means of computer simulations, that statistically efficient or near-efficient smoothed estimates can be obtained by backward-time filtering of the estimates yielded by the well-known and widely-used exponentially weighted least squares and Kalman filter-based parameter trackers. The proposed algorithms have low computational requirements and are easy to implement.

Appendix. Derivation of (32)

Using (27) one obtains

$$\Delta \tilde{\boldsymbol{\theta}}(t) \cong \mathbf{Q} \frac{\mathbf{X}(q^{-1}) - \mathbf{I}_n}{1 - q^{-1}} \mathbf{Q}^T \mathbf{w}(t) + \mathbf{Q} \boldsymbol{\Lambda}^{-1/2} \mathbf{X}(q^{-1}) \mathbf{Q}^T \mathbf{z}(t).$$

Since the processes $\{\mathbf{w}(t)\}$ and $\{\mathbf{z}(t)\}$ are orthogonal and $\mathbf{W} = \sigma_w^2 \mathbf{I}_n$, $\mathbf{Z} = \sigma_v^2 \mathbf{I}_n$, it holds that

$$\begin{aligned} \mathbb{E} [\Delta \tilde{\boldsymbol{\theta}}(t) \Delta \tilde{\boldsymbol{\theta}}^T(t)] &= \sigma_w^2 \mathbf{Q} \mathbf{I}_1 \left[X(e^{-j\omega}) \right] \mathbf{Q}^T \\ &\quad + \sigma_v^2 \mathbf{Q} \boldsymbol{\Lambda}^{-1} \mathbf{I}_2 \left[X(e^{-j\omega}) \right] \mathbf{Q}^T \end{aligned}$$

where: $\mathbf{I}_i [X(e^{-j\omega})] = \text{diag}\{I_i[X_1(e^{-j\omega})], \dots, I_i[X_n(e^{-j\omega})]\}$, $i = 1, 2$. It is straightforward to check that

$$I_1 [X_i(e^{-j\omega})] \cong \frac{1}{4\kappa \sqrt{\lambda_i}}, \quad I_2 [X_i(e^{-j\omega})] \cong \frac{\kappa \sqrt{\lambda_i}}{4}.$$

Hence

$$\mathbf{I}_1 [X(e^{-j\omega})] \cong \frac{1}{4\kappa} \boldsymbol{\Lambda}^{-1/2}, \quad \mathbf{I}_2 [X(e^{-j\omega})] \cong \frac{\kappa}{4} \boldsymbol{\Lambda}^{1/2}$$

leading to

$$\mathbb{E} [\Delta \tilde{\boldsymbol{\theta}}(t) \Delta \tilde{\boldsymbol{\theta}}^T(t)] \cong \mathbf{Q} \boldsymbol{\Lambda}^{-1/2} \mathbf{Q}^T \left(\frac{\sigma_w^2}{4\kappa} + \frac{\sigma_v^2 \kappa}{4} \right). \tag{37}$$

Observing that $\mathbf{Q} \boldsymbol{\Lambda}^{-1/2} \mathbf{Q}^T = \boldsymbol{\Phi}^{-1/2}$ and evaluating (37) at $\kappa = \kappa_{\text{opt}} = \sigma_w / \sigma_v$ one arrives at (32).

References

- Bai, E.-W., Fu, L.-C., & Sastry, S. (1988). Averaging analysis for discrete time and sampled data adaptive systems. *IEEE Transactions on Circuits and Systems*, 35, 137–148.
- Guo, L., & Ljung, L. (1995). Performance analysis of general tracking algorithms. *IEEE Transactions on Automatic Control*, 40, 1388–1402.
- Haykin, S. (1996). *Adaptive filter theory*. Englewood Cliffs, NJ: Prentice Hall.
- Jury, M. (1964). *Theory and application of the Z-transform method*. New York: Wiley.
- Lewis, F. L. (1986). *Optimal estimation*. New York: Wiley.
- Ljung, L., & Gunnarsson, S. (1990). Adaptive tracking in system identification—A survey. *Automatica*, 26, 7–22.
- Meditch, J. S. (1973). A survey of data smoothing for linear and nonlinear dynamic systems. *Automatica*, 9, 151–162.
- Niedźwiecki, M. (1990). Identification of time-varying systems using combined parameter estimation and filtering. *IEEE Transactions on Acoustics Speech and Signal Processing*, 38, 679–686.
- Niedźwiecki, M. (2000). *Identification of time-varying processes*. New York: Wiley.
- Niedźwiecki, M. (2007a). On the lower smoothing bound in identification of time-varying systems. *Automatica*, 44, 459–464.
- <http://www.eti.pg.gda.pl/katedry/ksa/pracownicy/Maciej.Niedzwiecki>.
- Niedźwiecki, M. (2007b). On ‘cheap smoothing’ opportunities in identification of time-varying systems. *Technical report # ETI-2/2007*, Gdańsk University of Technology, *Automatica*, in press (doi:10.1016/j.automatica.2007.09.007). <http://www.eti.pg.gda.pl/katedry/ksa/pracownicy/Maciej.Niedzwiecki>.



Maciej Niedźwiecki was born in Poznań, Poland in 1953. He received the M.Sc. and Ph.D. degrees from the Gdańsk University of Technology, Gdańsk, Poland, and the Dr. Hab. (D.Sc.) degree from the Technical University of Warsaw, Warsaw, Poland, in 1977, 1981 and 1991, respectively.

He spent three years as a Research Fellow with the Department of Systems Engineering, Australian National University, 1986–1989. In 1990–1993 he served as a Vice Chairman of Technical Committee on Theory of the International Federation of Automatic Control (IFAC). He is the author of the book *Identification of Time-varying Processes* (Wiley, 2000).

He works as a Professor and Head of the Department of Automatic Control, Faculty of Electronics, Telecommunications and Computer Science, Gdańsk University of Technology. His main areas of research interests include system identification, signal processing and adaptive systems.