Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

# 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
ul. Narutowicza 11/12, Gdańsk, Poland
Tel: +48 58 3472519, Fax: +48 58 3415821
e-mail: maciekn@eti.pg.gda.pl*

**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, 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.

Keywords: system identification, time-varying processes, noncausal estimation

## 1. INTRODUCTION

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

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

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

where $y(t)$ denotes the system output, $\boldsymbol{\varphi}(t) = [\varphi_1(t), \ldots, \varphi_n(t)]^{\mathrm{T}}$ is a known regression vector, $v(t)$ denotes white measurement noise, $\boldsymbol{\theta}(t) = [\theta_1(t), \ldots, \theta_n(t)]^{\mathrm{T}}$ 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 $\mathrm{var}[v(t)] = \sigma_v^2$.

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

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

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

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

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

where $\boldsymbol{\Sigma}(t) = \left[\sum_{i=1}^{t} \eta^{t-i}\boldsymbol{\varphi}(i)\boldsymbol{\varphi}^{\mathrm{T}}(i)\right]^{-1}$ is the inverse of the exponentially weighted regression matrix, $\varepsilon(t)$ denotes the one-step-ahead prediction error evaluated at instant $t$, 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 $\mathbf{W} = \mathrm{cov}[\mathbf{w}(t)] = \sigma_w^2\mathbf{I}_n$.

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

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

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

* This work was supported by MNiSW.

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 (Lewis, 1986), (Söderström & Stoica, 1988)

$$\widehat{\boldsymbol{\theta}}(t) = \widehat{\boldsymbol{\theta}}(t-1) + \mathbf{S}(t)\boldsymbol{\varphi}(t)\varepsilon(t)$$
$$\mathbf{S}(t) = \frac{\mathbf{P}(t-1)}{1+\boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{P}(t-1)\boldsymbol{\varphi}(t)} \tag{4}$$
$$\mathbf{P}(t) = \left[\mathbf{I}_n - \mathbf{S}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\right]\mathbf{P}(t-1) + \kappa^2\mathbf{I}_n$$

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 $\eta$ 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 become identical with the classical recursive least squares (RLS) algorithm.

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 canceling, 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.). When cancellation is performed on-line, disturbance estimates are obtained from $\widehat{d}(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}(t)$ where $\widehat{\boldsymbol{\theta}}(t) = f[\mathcal{Z}_-(t)]$ is the (causal) estimate yielded by the parameter tracking algorithm, such as EWLS or KF and $\boldsymbol{\varphi}(t)$ denotes the regression vector made up of samples drawn from the reference signal $r(t)$. 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 $\mathcal{Z} = \{y(1), \boldsymbol{\varphi}(1), \ldots, y(N), \boldsymbol{\varphi}(N)\}$. Then a more accurate estimate of $d(t)$ can be obtained from $\widetilde{d}(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\widetilde{\boldsymbol{\theta}}(t)$ where the quantity $\widetilde{\boldsymbol{\theta}}(t) = g[\mathcal{Z}]$, incorporating all past data points $\mathcal{Z}_-(t)$ and $N-t$ "future" data points $\mathcal{Z}_+(t+1) = \{y(t+1), \boldsymbol{\varphi}(t+1), \ldots, y(N), \boldsymbol{\varphi}(N)\}$, is the smoothed (noncausal) estimate of $\boldsymbol{\theta}(t)$.

Smoothing opportunities are seldom taken advantage of 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 in (Meditch, 1973), (Anderson & More, 1979), are computationally 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, the appropriately filtered KF estimates are statistically efficient, i.e. their accuracy reaches the Cramér-Rao type lower smoothing bound established recently (Niedźwiecki, 2008).

## 2. UNIFIED ANALYSIS FRAMEWORK

When the forgetting constant in the EWLS algorithm is sufficiently close to 1 and when the coefficient $\kappa$ 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 (Niedźwiecki, 2007)

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

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

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

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 $\gamma$ and for sufficiently slow changes in $\boldsymbol{\theta}(t)$ (compared to the changes in $\boldsymbol{\varphi}(t)$), the analysis of (5) can be carried using the averaging technique (Bai, Fu & Sastry, 1988), leading to the following approximation

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

which will be the basis of our further investigations.

We will analyze and optimize performance of 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 close-form expressions which explicitly relate the mean-square estimation errors to the adaptation gain $\gamma$ and second-order system statistics $\sigma_v^2$, $\sigma_w^2$ 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 $\gamma = 1-\eta$, $\mathbf{A} = \mathbf{\Phi}^{-1}$ and hence the recursive relationship (6) can be rewritten in an explicit form as

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

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

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

and $\mathbf{z}(t) = \mathbf{\Phi}^{-1}\boldsymbol{\varphi}(t)v(t)$. Note that $\{\mathbf{z}(t)\}$ is a white noise sequence with covariance matrix $\mathbf{Z} = \mathrm{cov}[\mathbf{z}(t)] = \sigma_v^2\mathbf{\Phi}^{-1}$.
According to (7), the parameter estimation error $\Delta\widehat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t)$ can be expressed in the form

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

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

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

Under (A1) - (A3) it holds that $\mathbf{S}_{\mathrm{w}}(\omega) = \mathbf{W} = \sigma_w^2\mathbf{I}_n$, $\mathbf{S}_{\mathrm{z}}(\omega) = \mathbf{Z} = \sigma_v^2\mathbf{\Phi}^{-1}$, $\forall\omega\in[-\pi,\pi]$. By means of residue calculus (Jury, 1964) one obtains

$$\begin{aligned}\mathcal{T}_{\mathrm{EWLS}} &= \frac{(1-\gamma)^2 n\sigma_w^2}{\gamma(2-\gamma)} + \frac{\gamma\sigma_v^2\mathrm{tr}\{\mathbf{\Phi}^{-1}\}}{2-\gamma}\\ &\cong \frac{n\sigma_w^2}{2\gamma} + \frac{\gamma\sigma_v^2\mathrm{tr}\{\mathbf{\Phi}^{-1}\}}{2}\end{aligned} \tag{8}$$

where the approximation is tight for sufficiently small values of $\gamma$.
The first term on the right-hand side of (8) constitutes the bias component of the mean-square error (MSE) and the second term – its variance component. Since the bias component is inversely proportional to the adaptation gain $\gamma$, whereas the variance component is proportional to $\gamma$, to obtain good tracking results one should trade off both error terms. The optimal value of $\gamma$, i.e. the one that minimizes $\mathcal{T}_{\mathrm{EWLS}}$, can be obtained by means of solving

$$\frac{\gamma_{\mathrm{opt}}^2}{1-\gamma_{\mathrm{opt}}} = \frac{c_1}{c_2} \tag{9}$$

where $c_1 = \mathrm{tr}\{\mathbf{W}\} = n\sigma_w^2$ and $c_2 = \mathrm{tr}\{\mathbf{Z}\} = \sigma_v^2\mathrm{tr}\{\mathbf{\Phi}^{-1}\}$.
Using the small gain approximations ($\gamma \ll 1$) one arrives at $\gamma_{\mathrm{opt}} \cong \sqrt{c_1/c_2}$ and

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

### 3.2 Noncausal estimation

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

$$\widetilde{\boldsymbol{\theta}}(t) = G(q^{-1})\widehat{\boldsymbol{\theta}}(t)\,. \tag{10}$$

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, 2007), a very simple form of smoothing can be obtained by setting $G(q^{-1}) = q^{\tau_o}$, where $\tau_o$ is the nominal (low frequency) delay of the filter $F(q^{-1})$. In this case $\widehat{\boldsymbol{\theta}}(t)$ is simply regarded as an estimate of $\boldsymbol{\theta}(t-\tau_o)$, instead of $\boldsymbol{\theta}(t)$. 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 $\Delta\widetilde{\boldsymbol{\theta}}(t) = \widetilde{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t)$. Note that the steady-state MSE can be in this case written down in the form

$$\mathcal{T}_{\mathrm{EWLS}}^{'} = \mathrm{E}[||\Delta\widetilde{\boldsymbol{\theta}}(t)||^2] = \frac{1}{2\pi}\int_{-\pi}^{\pi} f\left[X(e^{-j\omega})\right]d\omega \tag{11}$$

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

where $X(e^{-j\omega}) = F(e^{-j\omega})G(e^{-j\omega})$, $H(e^{-j\omega}) = 1/(1-e^{-i\omega})$ and $*$ denotes the complex conjugation.
We will look for a transfer function $X(e^{-j\omega})$ which minimizes $\mathcal{T}_{\mathrm{EWLS}}^{'}$. When no causality constraints are imposed on $X(e^{-j\omega})$, minimization of (11) is pretty straightforward – the problem can be solved by minimizing $f\left[X(e^{-j\omega})\right]$ for every value of $\omega\in[-\pi,\pi]$. Setting $\partial f/\partial X^*|_{X=X_{\mathrm{opt}}} = 0$ one obtains $X_{\mathrm{opt}} = c_1HH^*/(c_2+c_1HH^*)$, or equivalently

$$X_{\mathrm{opt}}(q^{-1}) = \frac{c_1/c_2}{c_1/c_2 + (1-q^{-1})(1-q)}\,. \tag{12}$$

Since the right-hand side of (12) can be rewritten in the form $\gamma^2/([1-(1-\gamma)q^{-1}][1-(1-\gamma)q])$, where $\gamma$ is the solution of $\gamma^2/(1-\gamma) = c_1/c_2$, one finally arrives at (cf. (9))

$$X_{\mathrm{opt}}(q^{-1}) = F_{\mathrm{opt}}(q^{-1})F_{\mathrm{opt}}(q) \tag{13}$$

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

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

$$\begin{aligned}G_{\mathrm{opt}}(q^{-1}) &= \frac{X_{\mathrm{opt}}(q^{-1})}{F_{\mathrm{opt}}(q^{-1})} = F_{\mathrm{opt}}(q)\\ &= \frac{\gamma_{\mathrm{opt}}}{1-(1-\gamma_{\mathrm{opt}})q}\,.\end{aligned} \tag{14}$$

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

$$\begin{aligned}\widetilde{\boldsymbol{\theta}}(N) &= \widehat{\boldsymbol{\theta}}(N)\\ \widetilde{\boldsymbol{\theta}}(t) &= (1-\gamma)\widetilde{\boldsymbol{\theta}}(t+1) + \gamma\widehat{\boldsymbol{\theta}}(t)\\ t &= N-1,\ldots,1\end{aligned} \tag{15}$$

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

$$\mathcal{T}_{\mathrm{EWLS}}^{'} \cong \frac{n\sigma_w^2}{4\gamma} + \frac{\gamma\sigma_v^2\mathrm{tr}\{\mathbf{\Phi}^{-1}\}}{4} \cong \frac{1}{2}\,\mathcal{T}_{\mathrm{EWLS}} \tag{16}$$

which means that, irrespective of the choice of $\gamma$, 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: $(\mathcal{I}^{'}_{\rm EWLS})_{\min} \cong (\mathcal{I}_{\rm EWLS})_{\min}/2$.

**Remark**

Note that the optimal smoothing (noncausal) gain is identical with the optimal tracking (causal) gain. This has important practical implications. When $\gamma_{\rm opt}$ 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 $\mathbf{\Phi}$, converting $\mathbf{\Phi}$ into a diagonal form: $\mathbf{Q}^{\rm T}\mathbf{Q} = \mathbf{Q}\mathbf{Q}^{\rm T} = \mathbf{I}_n$, $\mathbf{Q}^{\rm T}\mathbf{\Phi}\mathbf{Q} = \mathbf{\Lambda}$, where $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}$ is a diagonal matrix made up of the eigenvalues of $\mathbf{\Phi}$.
For the KF algorithm it holds that $\gamma = \kappa$ and $\mathbf{A} = \mathbf{\Phi}^{-1/2}$. In this case the relationship (6) can be rewritten in a closed form as (Niedźwiecki, 2007)

$$\widehat{\boldsymbol\theta}(t) \cong \mathbf{Q}\mathbf{F}(q^{-1})\mathbf{Q}^{\rm T}\boldsymbol\theta(t) + \mathbf{Q}\mathbf{\Lambda}^{-1/2}\mathbf{F}(q^{-1})\mathbf{Q}^{\rm T}\mathbf{z}(t) \quad (17)$$

where

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

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

and $\mathbf{z}(t) = \mathbf{\Phi}^{-1/2}\boldsymbol\varphi(t)v(t)$, $\mathbf{Z} = \mathrm{cov}[\mathbf{z}(t)] = \sigma_v^2\mathbf{I}_n$.
Based on (17) the following error model can be derived

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

leading to

$$\mathcal{I}_{\rm KF} = \mathrm{E}[||\Delta\widehat{\boldsymbol\theta}(t)||^2] = \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{tr}\left\{\mathbf{Q}\frac{\mathbf{F}(e^{-j\omega})-\mathbf{I}_n}{1-e^{-j\omega}} \times \mathbf{Q}^{\rm T}\mathbf{S}_{\rm w}(\omega)\mathbf{Q}\ \frac{\mathbf{F}(e^{j\omega})-\mathbf{I}_n}{1-e^{j\omega}}\ \mathbf{Q}^{\rm T}\right\}d\omega + \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{tr}\left\{\mathbf{Q}\mathbf{\Lambda}^{-1/2}\mathbf{F}(e^{-j\omega})\mathbf{Q}^{\rm T}\mathbf{S}_{\rm z}(\omega)\mathbf{Q} \times \mathbf{F}(e^{j\omega})\mathbf{\Lambda}^{-1/2}\mathbf{Q}^{\rm T}\right\}d\omega\ .$$

In the case considered $\mathbf{S}_{\rm w}(\omega) = \mathbf{W} = \sigma_w^2\mathbf{I}_n$, $\mathbf{S}_{\rm z}(\omega) = \mathbf{Z} = \sigma_v^2\mathbf{I}_n$, $\forall\omega \in [-\pi, \pi]$. For small adaptation gains one obtains

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

where the last transition stems from the fact that $\sum_{i=1}^{n} 1/\sqrt{\lambda_i} = \mathrm{tr}\{\mathbf{\Lambda}^{-1/2}\} = \mathrm{tr}\{\mathbf{\Phi}^{-1/2}\}$.
Optimization of $\mathcal{I}_{\rm KF}$ is straightforward – the minimum value of the mean-square parameter estimation error is achieved for

$$\kappa = \kappa_{\rm opt} = \frac{\sigma_w}{\sigma_v} \quad (18)$$

which is an expected result, since under (18) the algorithm (4) is a "true" Kalman filter, i.e. the optimal parameter tracking procedure. Note that

$$(\mathcal{I}_{\rm KF})_{\min} = \mathcal{I}_{\rm KF}|_{\kappa=\kappa_{\rm opt}} \cong \sigma_v\sigma_w\mathrm{tr}\{\mathbf{\Phi}^{-1/2}\}\ .$$

### *4.2 Noncausal estimation*

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

$$\widetilde{\boldsymbol\theta}(t) = \mathbf{Q}\mathbf{G}(q^{-1})\mathbf{Q}^{\rm T}\widehat{\boldsymbol\theta}(t) \quad (19)$$

where $\mathbf{G}(q^{-1}) = \mathrm{diag}\{G_1(q^{-1}), \ldots, G_n(q^{-1}\}$ and $G_i(q^{-1})$, $i = 1.\ldots, n$ denote transfer functions of the appropriately designed noncausal filters.
Combining this result with (17) and noting that the matrices $\mathbf{\Lambda}$, $\mathbf{F}(q^{-1})$ and $\mathbf{G}(q^{-1})$ are diagonal and hence they commute, one arrives at

$$\widetilde{\boldsymbol\theta}(t) \cong \mathbf{Q}\mathbf{X}(q^{-1})\mathbf{Q}^{\rm T}\boldsymbol\theta(t) + \mathbf{Q}\mathbf{\Lambda}^{-1/2}\mathbf{X}(q^{-1})\mathbf{Q}^{\rm T}\mathbf{z}(t) \quad (20)$$

where $\mathbf{X}(q^{-1}) = \mathrm{diag}\{X_1(q^{-1}), \ldots, X_n(q^{-1}\}$ and $X_i(q^{-1}) = F_i(q^{-1})G_i(q^{-1})$ , $i = 1, \ldots, n$.
Minimization of the mean-square estimation error $\mathcal{I}^{'}_{\rm KF} = \mathrm{E}[\ ||\Delta\widetilde{\boldsymbol\theta}(t)||^2]$ can be carried out in an analogous way as minimization of $\mathcal{I}^{'}_{\rm EWLS}$, performed in Section 3.2. Using the same technique one can show that $X_i^{\rm opt}(q^{-1}) = F_i^{\rm opt}(q^{-1})F_i^{\rm opt}(q)$ where

$$F_i^{\rm opt}(q^{-1}) = \frac{\kappa_{\rm opt}\sqrt{\lambda_i}}{1-(1-\kappa_{\rm opt}\sqrt{\lambda_i})q^{-1}}$$

and $\kappa_{\rm opt} = \sigma_w/\sigma_v$ denotes the optimal tracking gain. This leads to $G_i^{\rm opt}(q^{-1}) = F_i^{\rm opt}(q)$ and to the following backward-time filtering scheme, analogous to (15)

$$\widetilde{\boldsymbol\theta}(N) = \widehat{\boldsymbol\theta}(N)$$
$$\widetilde{\boldsymbol\theta}(t) = (\mathbf{I}_n - \kappa\mathbf{\Phi}^{1/2})\widetilde{\boldsymbol\theta}(t+1) + \kappa\mathbf{\Phi}^{1/2}\widehat{\boldsymbol\theta}(t) \quad (21)$$
$$t = N-1, \ldots, 1$$

obtained after adopting $G_i(q^{-1}) = F_i(q)$. Similarly as before it holds that $\mathcal{I}^{'}_{\rm KF} \cong \mathcal{I}_{\rm KF}/2$ and $(\mathcal{I}^{'}_{\rm KF})_{\min} \cong (\mathcal{I}_{\rm KF})_{\min}/2$.
Since in the steady state it holds that (Niedźwiecki, 2007) $\mathbf{S}(t) \cong \kappa\mathbf{\Phi}^{-1/2}$, where $\mathbf{S}(t)$ is the matrix recursively updated by the KF algorithm, a good estimate of $\mathbf{\Phi}^{1/2}$ can be obtained from $\widehat{\mathbf{\Phi}}^{1/2}(t) = \kappa\mathbf{S}^{-1}(t)$. 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 ($t = N/2$).
According to (Niedźwiecki, 2008), for any estimator $\widehat{\boldsymbol\theta}(t)$ of $\boldsymbol\theta(t)$, including all noncausal estimators, it holds that (under assumptions (A1) - (A4))

$$\mathrm{E}\left[\Delta\widehat{\boldsymbol\theta}(t)\Delta\widehat{\boldsymbol\theta}^{\rm T}(t)\right] \geq \frac{1}{2}\ \sigma_v\sigma_w\mathbf{\Phi}^{-1/2} = \mathbf{B}_{\rm LSB} \quad (22)$$

where $\mathbf{B}_{\rm LSB}$ denotes the Cramér-Rao type lower smoothing bound (LSB).
Note that $(\mathcal{I}^{'}_{\rm KF})_{\min} \cong \mathrm{tr}\{\mathbf{B}_{\rm LSB}\}$. Furthermore, it can be shown that $\left(\mathrm{E}[\Delta\widetilde{\boldsymbol\theta}(t)\Delta\widetilde{\boldsymbol\theta}^{\rm T}(t)]\right)_{\min} = \mathrm{E}[\Delta\widetilde{\boldsymbol\theta}(t)\Delta\widetilde{\boldsymbol\theta}^{\rm T}(t)]_{\kappa=\kappa_{\rm opt}} \cong \mathbf{B}_{\rm LSB}$, 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 more involved – Rauch-Tung-Striebel smoother (i.e. the *genuine* Kalman smoother) designed for the system (1) - (2). In Section 6 we will show that simulation experiments fully confirm this claim.

## 5. COMPARISON OF THE EWLS AND KF APPROACHES

Using the Cauchy-Shwartz inequality one obtains

$$\mathrm{tr}\{\mathbf{\Phi}^{-1/2}\} \le \sqrt{n\,\mathrm{tr}\{\mathbf{\Phi}^{-1}\}}$$

which leads to $(\mathcal{T}_{\mathrm{KF}})_{\min} \le (\mathcal{T}_{\mathrm{EWLS}})_{\min}$ and $(\mathcal{T}_{\mathrm{KF}}^{'})_{\min} \le (\mathcal{T}_{\mathrm{EWLS}}^{'})_{\min}$, where equality holds iff all eigenvalues of $\mathbf{\Phi}$ 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.

Comparison of numerical complexity of the proposed algorithms favors the EWLS approach. The efficient mechanization of the EWLS-based smoother (3)+(15) 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)+(21) are: $1.5n^2 + 5.5n$ operations and $2n^2$ operations, respectively (the cost of evaluating $\mathbf{\Phi}^{1/2}$ 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

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

### 6.1 Example 1

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

$$y(t) = \theta_1(t)u(t) + \theta_2(t)u(t-1) + v(t)$$
$$u(t) = 0.8u(t-1) + e(t)$$

where $v(t) \sim \mathcal{N}(0,1)$, $e(t) \sim \mathcal{N}(0,1)$ and $\{e(t)\}$ is an i.i.d. sequence independent of $\{v(t)\}$.

System parameters were generated using the random walk model

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

where $\mathbf{w}(t) \sim \mathcal{N}(0, 0.0001\mathbf{I}_2)$ and $\boldsymbol{\theta}(t) = [\theta_1(t),\, \theta_2(t)]^{\mathrm{T}}$.

Performance of the compared 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

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

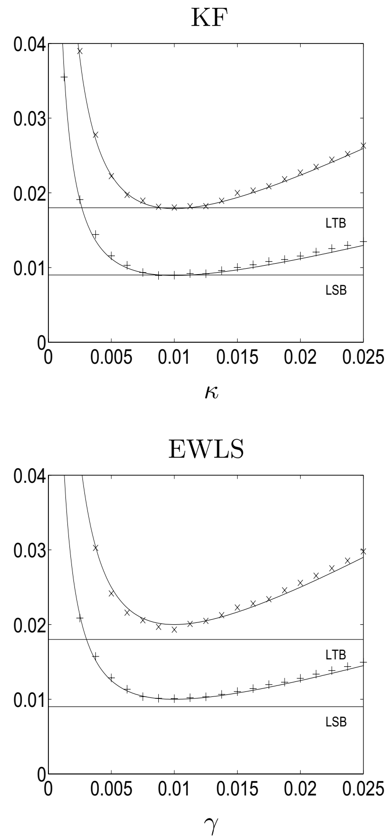


Fig. 1. Dependence of the mean-square parameter estimation errors on the adaptation gains $\kappa$ and $\gamma$ for the KF/EWLS trackers ($\times$) and the KF/EWLS-based smoothers ($+$). The lower tracking bound (LTB) and the lower smoothing bound (LSB) are indicated by horizontal lines. Solid lines show theoretical dependence of MSE on $\kappa$ and $\gamma$ for both algorithms.

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 $\gamma$ and $\kappa$.

Figure 1 shows results obtained for the KF and EWLS trackers and for the KF-based and EWLS-based smoothers 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 KF smoothing algorithm achieves the lower smoothing bound, which limits performance of any (causal or noncausal) estimation scheme. 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.

### 6.2 Example 2

In our second simulation experiment sinusoidal parameter changes were enforced (see Figure 2):

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

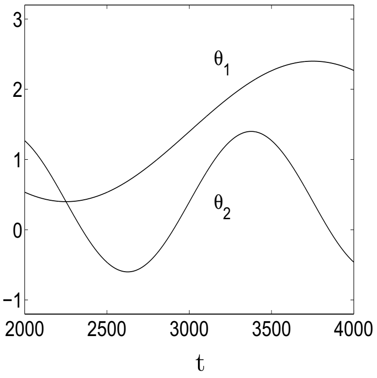


Fig. 2. Evolution of system parameters.

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

Figure 3 shows 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 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 is operated under "nonstandard" conditions.

## 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.

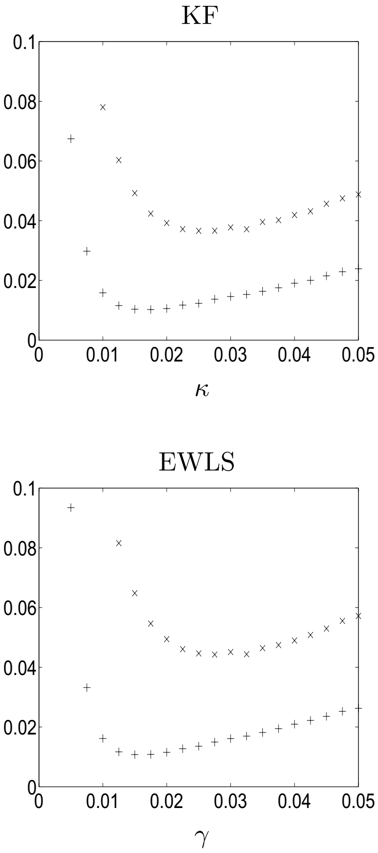


Fig. 3. Dependence of the mean-square parameter estimation errors on the adaptation gains $\kappa$ and $\gamma$ for the KF/EWLS trackers ($\times$) and the KF/EWLS-based smoothers ($+$).

## REFERENCES


Anderson, B.D.O. & J.B. Moore (1979). *Optimal Filtering.* Prentice Hall, Englewood Cliffs NJ.

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.

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

Jury, M. (1964). *Theory and Application of the $\mathcal{Z}$-transform Method.* Wiley, New York.

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

Ljung, L. & Gunnarsson S. (1985). Adaptive tracking in system identification - a survey. *Automatica*, **vol. 26**, 7–22.

Meditch, J.S. (1973). A survey of data smoothing for linear and nonlinear dynamic systems. *Automatica*, **vol. 9**, 151–162.

Niedźwiecki, M. (1990). Identification of time-varying systems using combined parameter estimation and filtering. *IEEE Trans. Acoust. Speech Signal Process.*, **vol. 38**, 679–686.

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

Niedźwiecki, M. (2007). On 'cheap smoothing' opportunities in identification of time-varying systems. Gdańsk University of Technology, Technical Report #ETI-2/2007, http://www.eti.pg.gda.pl/katedry/ksa (to appear in *Automatica*).

Niedźwiecki, M. (2008). On the lower smoothing bound in identification of time-varying systems. *Automatica*, **vol. 44**, 459–464.

Söderström, T. & P. Stoica (1988). *System Identification.* Prentice Hall. Englewood Cliffs NJ.