Postprint of: Niedźwiecki M., Gackowski S., New Approach to Noncausal Identification of Nonstationary Stochastic FIR Systems Subject to Both Smooth and Abrupt Parameter Changes, IEEE Transactions on Automatic Control, Vol. 58, Iss. 7 (2013), pp. 1847-1853, DOI: 10.1109/TAC.2013.2238995

New approach to noncausal identification of nonstationary stochastic FIR systems subject to both smooth and abrupt parameter changes

Maciej Niedźwiecki and Szymon Gackowski

***Abstract*—In this paper we consider the problem of finite-interval parameter smoothing for a class of nonstationary linear stochastic systems subject to both smooth and abrupt parameter changes. The proposed parallel estimation scheme combines the estimates yielded by several exponentially weighted basis function algorithms. The resulting smoother automatically adjusts its smoothing bandwidth to the type and rate of nonstationarity of the identified system. It also allows one to account for the distribution of the measurement noise.**

***Index Terms*—Identification of nonstationary systems, parameter smoothing.**

I. Introduction

In this paper we consider the problem of noncausal identification, i.e., identification based on pre-recorded data, of a nonstationary linear stochastic system subject to a mixed-mode parameter variation – a combination of smooth persistent parameter changes and occasional parameter jumps. In the statistical literature such approach to system identification is often referred to as fixed-interval parameter smoothing.

An interesting application which admits such problem formulation is identification of rapidly fading mobile radio channels [1]. Slow time variation of the channel coefficients is due to receiver movement in space, and rapid variation is caused by abrupt changes of the spatial system configuration (e.g. due to switching between base stations). Fixed-interval parameter smoothing can be used to reconstruct trajectories of channel coefficients. The reconstructed trajectories can be later used to perform realistic simulations of a mobile communication system – they serve as 'ground truth' in demonstrations and tests.

Most of the existing work on parameter smoothing is based on the stochastic model of parameter variation known as the integrated random walk (IRW) model – the idea goes back to Shiller [2]. Adopting the IRW model of order m, one assumes that the vector of system coefficients $\boldsymbol{\theta}(t) = [\theta_1(t), \ldots, \theta_n(t)]^{\mathrm{T}}$ obeys the following equation

$$\nabla^m \boldsymbol{\theta}(t) = \left(1 - q^{-1}\right)^m \boldsymbol{\theta}(t) = \mathbf{w}(t) \tag{1}$$

where $\nabla^m(\cdot)$ denotes the m-th order difference (q^{-1} stands for the backward shift operator), and $\mathbf{w}(t)$, $\mathrm{cov}[\mathbf{w}(t)] = \sigma_w^2 \mathbf{I}_n$, is a white noise process ($\mathbf{I}_n$ denotes the $n \times n$ identity matrix).

After state space embedding the problem of estimation of parameters governed by (1) can be solved using the algorithms known as Kalman filters/smoothers [3], [4]. The estimation properties of these algorithms depend on the adopted order m of the IRW model and on the variance quotient (smoothness tradeoff parameter) $\xi = \sigma_w^2/\sigma_v^2$ – see e.g. [5]. To obtain satisfactory estimation results, both quantities should be locally 'matched' to the type and speed of parameter variation, as well as to the signal-to-noise ratio. The matching problem can be solved by means of combining results yielded by several, simultaneously operated, Kalman filters/smoothers, equipped with different smoothness constraints. For systems with abrupt parameter changes such robust parallel estimation scheme, called competitive smoother, was described in [6]. It combines, in a statistically meaningful way, the results provided by several forward-time and backward-time Kalman filters. The analogous solution for systems with smooth parameter changes, called cooperative smoother, was proposed in [7] – it combines results yielded by several Kalman smoothing algorithms. Finally, one should mention the interacting multiple model (IMM) approach, originally proposed as a solution to the problem of tracking maneuvering targets [8]. The suitably modified IMM algorithm, which can be used for the purpose of parameter tracking (based on IRW modeling), was described in [9]. In principle, the IMM smoother can be obtained by combining the results presented in [9] and [10]. However, to the best of our knowledge, the problem of IMM-based parameter smoothing has not been explored yet.

The contribution of the paper is twofold.

First, we extend the results presented in [6] and [7] to a new class of estimation algorithms - exponentially weighted basis function (EWBF) trackers/smoothers. The basis function approach to system identification is based on a deterministic model of parameter variation – it is assumed that system parameters can be modeled as linear combinations of deterministic functions of time, the so-called basis functions. The EWBF algorithms are obtained by combining the basis function based system description with the estimation technique known as exponential weighting (or exponential forgetting). Similar to the Kalman-filter-based algorithms, the EWBF algorithms belong to the class of finite memory adaptive filters. However, unlike the Kalman-filter-based algorithms, the estimation memory of EWBF trackers/smoothers can be easily quantified in terms of their design parameters, such as the number of basis functions and the forgetting constant. This makes the EWBF algorithms much easier to handle in practice.

Second, we show how one can merge results yielded by the competitive smoother with those provided by the cooperative smoother. The resulting combined smoother preserves advantages of the component algorithms – similar to the competitive smoother it sharply resolves parameter jumps, and, similar to the cooperative smoother, it accurately follows smooth parameter changes.

II. Basis function approach to parameter tracking and smoothing

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

$$y(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t) + v(t) \tag{2}$$

where $\boldsymbol{\varphi}(t) = [\varphi_1(t), \ldots, \varphi_n(t)]^{\mathrm{T}}$ denotes the vector of input (regression) variables and $v(t)$ denotes white measurement noise.

Suppose that identification of (2) is carried out based on the pre-recorded data set $\Omega(N) = \{y(i), \boldsymbol{\varphi}(i),\ i \in T\}$, where $T = [1, \ldots, N]$ denotes the observation interval of length N. Denote by

$$\{f_j(t) = t^{j-1}, j = 1, \ldots, m, t \in T\} \tag{3}$$

the set of the so-called basis functions. Adopting the basis function approach, one assumes that each time-varying system coefficient can be represented by a linear combination of basis functions (for more details on the BF approach, and its comparison with other methods of identification of nonstationary systems, see [5] and [11])

$$\theta_i(t) = \sum_{j=1}^{m} \alpha_{ij} f_j(t), \quad i = 1, \ldots, n,\ t \in T. \tag{4}$$

Manuscript received March 5, 2012; revised July 24, 2012, November 15, and December ??, 2012; accepted December 18, 2012. Date of publication January ??, 2013; date of current version December ??, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Erik Weyer.

The authors are with the Faculty of Electronics, Telecommunications and Computer Science Department of Automatic Control, Gdańsk University of Technology, ul. Narutowicza 11/12, Gdańsk, Poland (e-mail: maciekn@eti.pg.gda.pl, szymon.gackowski@gmail.com).

Digital Object Identifier ???

Powers of time are the most frequently used general-purpose basis functions (although, in principle, *any* set linearly independent square summable sequences defined on T can be used). Such a choice, which stems from the Taylor series expansion theorem and guarantees recursive computability of parameter estimates, can be traced back to Subba Rao [12]. All results presented in this paper will be derived for the basis (3).

After combining (2) with (4), system equation can be written down in the form

$$y(t) = \boldsymbol{\psi}^{\mathrm{T}}(t)\boldsymbol{\alpha} + v(t) \tag{5}$$

where $\boldsymbol{\alpha} = [\alpha_{11}, \ldots, \alpha_{1m}, \ldots, \alpha_{n1}, \ldots, \alpha_{nm}]^{\mathrm{T}}$ is the $nm \times 1$ vector of coefficients used to describe the parameter time variation, and $\boldsymbol{\psi}(t)$ denotes the $nm \times 1$ generalized regression vector: $\boldsymbol{\psi}(t) = \boldsymbol{\varphi}(t) \otimes \mathbf{f}(t)$, $\mathbf{f}(t) = [f_1(t), \ldots, f_m(t)]^{\mathrm{T}}$. The symbol $\otimes$ is used to denote a Kronecker product of two matrices/vectors.

According to (5), the time-varying process of order n can be represented by a linear time-invariant model of order nm. The values of the time-varying system parameters $\boldsymbol{\theta}(t)$ can be easily recovered from $\boldsymbol{\alpha}$ using the following compact version of (4)

$$\boldsymbol{\theta}(t) = \mathbf{Z}(t)\boldsymbol{\alpha}, \quad \mathbf{Z}(t) = \mathbf{I}_n \otimes \mathbf{f}^{\mathrm{T}}(t). \tag{6}$$

A. Parameter Tracking

It would be naive to assume that the expansion coefficients α_{ij} in (4) are constant in the entire time domain. This assumption would mean, in fact, that the analyzed system could be regarded as reducible, i.e., varying in a perfectly predictable manner. To cope with possible fluctuations in α_{ij}'s, the method of basis functions can be combined with exponential data weighting (exponential forgetting). Weighting forces the estimation to be more focused on the most recent data samples and therefore allows one to track slow variations in expansion coefficients. The forward-time ($-$) and backward-time ($+$) exponentially weighted basis function (EWBF) parameter trackers take the form

$$\widehat{\boldsymbol{\theta}}_{\pm}(t) = \mathbf{Z}(t)\widehat{\boldsymbol{\alpha}}_{\pm}(t)$$

$$\begin{aligned} \widehat{\boldsymbol{\alpha}}_{\pm}(t) &= \arg\min_{\boldsymbol{\alpha}} \sum_{i \in T_{\pm}(t)} \lambda^{|t-i|} \left[y(i) - \boldsymbol{\psi}^{\mathrm{T}}(i)\boldsymbol{\alpha} \right]^2 \\ &= \mathbf{G}_{\pm}^{-1}(t)\mathbf{g}_{\pm}(t) \end{aligned} \tag{7}$$

where λ, $0 < \lambda < 1$, denotes the forgetting constant, $T_{-}(t) = [1, \ldots, t]$, $T_{+}(t) = [t, \ldots, N]$ and the quantities $\mathbf{G}_{\pm}(t)$, $\mathbf{g}_{\pm}(t)$ are recursively computable

$$\begin{aligned} \mathbf{G}_{\pm}(t) &= \lambda \mathbf{G}_{\pm}(t \pm 1) + \boldsymbol{\psi}(t)\boldsymbol{\psi}^{\mathrm{T}}(t) \\ \mathbf{g}_{\pm}(t) &= \lambda \mathbf{g}_{\pm}(t \pm 1) + \boldsymbol{\psi}(t)y(t) \end{aligned}$$

Using the well-known matrix inversion lemma [5], one can easily set up an algorithm for recursive computation of $[\mathbf{G}_{\pm}(t)]^{-1}$.

B. Parameter Smoothing

The EWBF parameter smoother is a natural complement of the causal and anticausal estimation algorithms presented in the previous subsection. The algorithm has the following form

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_{\star}(t) &= \mathbf{Z}(t)\widehat{\boldsymbol{\alpha}}_{\star}(t) \\ \widehat{\boldsymbol{\alpha}}_{\star}(t) &= \arg\min_{\boldsymbol{\alpha}} \sum_{i \in T} \lambda^{|t-i|} \left[y(i) - \boldsymbol{\psi}^{\mathrm{T}}(i)\boldsymbol{\alpha} \right]^2 \\ &= \mathbf{G}_{\star}^{-1}(t)\mathbf{g}_{\star}(t) \end{aligned} \tag{8}$$

where

$$\mathbf{G}_{\star}(t) = \sum_{i \in T} \lambda^{|t-i|} \boldsymbol{\psi}(i)\boldsymbol{\psi}^{\mathrm{T}}(i), \;\; \mathbf{g}_{\star}(t) = \sum_{i \in T} \lambda^{|t-i|} \boldsymbol{\psi}(i)y(i).$$

It is straightforward to check that the quantities $\mathbf{G}_{\star}(t)$ and $\mathbf{g}_{\star}(t)$ can be evaluated by means of backward time filtering of the quantities $\mathbf{G}_{-}(t)$ and $\mathbf{g}_{-}(t)$, respectively

$$\begin{aligned} \mathbf{G}_{\star}(t) &= \lambda \mathbf{G}_{\star}(t+1) + (1 - \lambda^2)\mathbf{G}_{-}(t) \\ \mathbf{g}_{\star}(t) &= \lambda \mathbf{g}_{\star}(t+1) + (1 - \lambda^2)\mathbf{g}_{-}(t). \end{aligned} \tag{9}$$

For the basis (3) with only one component $f_1(t) \equiv 1$, the EWBF smoother (8) is identical with the algorithm proposed in [13].

III. Statistical properties of the exponentially weighted basis function smoother

To simplify our analysis, we will assume that an infinite observation history is available, i.e., $\mathbf{G}_{\star}(t) = \sum_{i=-\infty}^{\infty} \lambda^{|t-i|} \boldsymbol{\psi}(i)\boldsymbol{\psi}^{\mathrm{T}}(i)$, $\mathbf{g}_{\star}(t) = \sum_{i=-\infty}^{\infty} \lambda^{|t-i|} \boldsymbol{\psi}(i)y(i)$. The results of such a steady state analysis will remain valid for the algorithm (8) as long as $1 \ll t \ll N$, i.e., everywhere except the boundary regions.

Denote by $\bar{\boldsymbol{\theta}}_{\star}(t) = \mathrm{E}\left[\widehat{\boldsymbol{\theta}}_{\star}(t)\right]$ the mean path of the estimates yielded by the EWBF smoother. The mean-squared parameter estimation error can be written down in the form

$$\mathrm{E}\left[\| \, \widehat{\boldsymbol{\theta}}_{\star}(t) - \boldsymbol{\theta}(t) \, \|^2\right] = \| \, \bar{\boldsymbol{\theta}}_{\star}(t) - \boldsymbol{\theta}(t) \, \|^2 + \mathrm{tr}\left\{ \mathrm{cov}\left[\widehat{\boldsymbol{\theta}}_{\star}(t)\right] \right\}$$

i.e., it can be decomposed into the bias component and the variance component, respectively. Both components will be examined in some detail below. To arrive at analytical results, we will make the following assumptions:

(A1) The sequence of regression vectors $\{\boldsymbol{\varphi}(t)\}$ is a wide sense stationary and ergodic process with positive definite correlation matrix $\mathrm{E}[\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)] = \boldsymbol{\Phi}_0 > 0$.

(A2) $\{v(t)\}$, independent of $\{\boldsymbol{\varphi}(t)\}$, is a sequence of zero-mean, independent and identically distributed random variables with variance σ_v^2.

Both assumptions hold true for rapidly fading mobile radio channels.

A. Estimation Bias

Under the assumption (A2), one arrives at

$$\bar{\boldsymbol{\theta}}_{\star}(t) = \mathrm{E}\left[\widehat{\boldsymbol{\theta}}_{\star}(t)\right] = \mathrm{E}\left[\mathbf{Z}(t)\mathbf{G}_{\star}^{-1}(t)\mathbf{r}_{\star}(t)\right]$$

where $\mathbf{r}_{\star}(t) = \sum_{i=-\infty}^{\infty} \lambda^{|t-i|} \boldsymbol{\psi}(i)\boldsymbol{\varphi}^{\mathrm{T}}(i)\boldsymbol{\theta}(i)$.

To facilitate further analysis, we will express all quantities in terms of orthonormal basis vectors $\mathbf{f}_{\lambda}(i) = \mathbf{C}\mathbf{f}(i)$ which obey

$$\sum_{i=-\infty}^{\infty} \lambda^{|i|} \mathbf{f}_{\lambda}(i)\mathbf{f}_{\lambda}^{\mathrm{T}}(i) = \mathbf{I}_m. \tag{10}$$

The matrix $\mathbf{C}$ can be established using the classical Gram-Schmidt procedure. Additionally, we will use the following properties of the basis (3): $\mathbf{f}(i) = \mathbf{A}\mathbf{f}(i-1)$, $f(-i) = \mathbf{B}\mathbf{f}(i)$, where $\mathbf{A} = [a_{ij}]$ is a lower triangular matrix such that $a_{ij} = \binom{i-1}{i-j}$ for $i = 1, \ldots, m, j \leq i$, and $\mathbf{B} = [b_{ij}]$ is a diagonal matrix such that $b_{ii} = (-1)^{i-1}$ for $i = 1, \ldots, m$. Let $\mathbf{H}(t) = \mathbf{I}_n \otimes \mathbf{D}(t)$ where $\mathbf{D}(t) = \mathbf{C}\mathbf{A}^t\mathbf{B}$. Observe that $\mathbf{D}(t)\mathbf{f}(i) = \mathbf{f}_{\lambda}(t-i)$. Note that

$$\mathbf{Z}(t)\mathbf{G}_{\star}^{-1}(t)\mathbf{r}_{\star}(t) = \mathbf{Z}_{\lambda}\mathbf{G}_{\lambda}^{-1}(t)\mathbf{r}_{\lambda}(t)$$

where $\mathbf{Z}_\lambda = \mathbf{Z}(t)\mathbf{H}^{\mathrm{T}}(t)$, $\mathbf{G}_\lambda(t) = \mathbf{H}(t)\mathbf{G}_\star(t)\mathbf{H}^{\mathrm{T}}(t)$ and $\mathbf{r}_\lambda(t) = \mathbf{H}(t)\mathbf{r}_\star(t)$. Using the identity $(\mathbf{P}\otimes\mathbf{Q})(\mathbf{R}\otimes\mathbf{S}) = \mathbf{PR}\otimes\mathbf{QS}$, which holds for Kronecker products, one obtains

$$\begin{aligned}\mathbf{Z}_\lambda &= \mathbf{I}_n \otimes \mathbf{f}_\lambda^{\mathrm{T}}(0)\\ \mathbf{G}_\lambda(t) &= \sum_{i=-\infty}^{\infty} \lambda^{|t-i|}\left[\boldsymbol{\varphi}(i)\boldsymbol{\varphi}^{\mathrm{T}}(i)\right]\otimes\left[\mathbf{f}_\lambda(t-i)\mathbf{f}_\lambda^{\mathrm{T}}(t-i)\right]\\ \mathbf{r}_\lambda(t) &= \sum_{i=-\infty}^{\infty} \lambda^{|t-i|}\left[\boldsymbol{\varphi}(i)\otimes\mathbf{f}_\lambda(t-i)\right]\boldsymbol{\varphi}^{\mathrm{T}}(i)\boldsymbol{\theta}(i).\end{aligned}$$

Under (A1) and (A2) it holds that $\mathrm{E}\left[\mathbf{G}_\lambda(t)\right] = \mathbf{\Phi}_0 \otimes \mathbf{I}_m$. Furthermore, using the generalized law of large numbers for weighted sums of random variables given by Pruitt [14], one can show that $\lim_{\lambda\to 1}\mathbf{G}_\lambda(t) = \mathrm{E}\left[\mathbf{G}_\lambda(t)\right]$ w.p.1. Hence, when the forgetting constant λ is sufficiently close to 1, one can apply the following approximation

$$\mathbf{G}_\lambda^{-1}(t) \cong \{\mathrm{E}\left[\mathbf{G}_\lambda(t)\right]\}^{-1} = \mathbf{\Phi}_0^{-1}\otimes\mathbf{I}_m. \tag{11}$$

Using this approximation, one obtains

$$\begin{aligned}&\mathbf{Z}_\lambda\mathbf{G}_\lambda^{-1}(t)\mathbf{r}_\lambda(t)\\ &\cong \sum_{i=-\infty}^{\infty}\lambda^{|t-i|}\left[\mathbf{\Phi}_0^{-1}\boldsymbol{\varphi}(i)\boldsymbol{\varphi}^{\mathrm{T}}(i)\right]\mathbf{f}_\lambda^{\mathrm{T}}(0)\mathbf{f}_\lambda(t-i)\boldsymbol{\theta}(i)\end{aligned}$$

which leads to

$$\bar{\boldsymbol{\theta}}_\star(t) \cong \sum_{i=-\infty}^{\infty} k_\star(i)\boldsymbol{\theta}(t-i),\;\; k_\star(i) = \lambda^{|i|}\mathbf{f}_\lambda^{\mathrm{T}}(0)\mathbf{f}_\lambda(i). \tag{12}$$

According to (12), the mean path of EWBF estimates $\{\bar{\boldsymbol{\theta}}_\star(t)\}$ can be approximately viewed as a result of passing the process $\{\boldsymbol{\theta}(t)\}$ through a linear time-invariant filter. We will call the sequence $\{k_\star(t)\}$ the impulse response associated with the EWBF smoother.

The impulse response $k_\star(i)$ can be alternatively expressed in the form

$$k_\star(i) = \lambda^{|i|}\mathbf{f}^{\mathrm{T}}(0)\left[\sum_{j=-\infty}^{\infty}\lambda^{|j|}\mathbf{f}(j)\mathbf{f}^{\mathrm{T}}(j)\right]^{-1}\mathbf{f}(i). \tag{13}$$

B. *Estimation Variance*

To quantify estimation variance of the EWBF smoother, we will assume, for the time being, that the true parameter trajectory can be exactly represented by a linear combination of basis functions. It is easy to check that, in the case considered, it holds that

$$\Delta\widehat{\boldsymbol{\theta}}_\star(t) = \widehat{\boldsymbol{\theta}}_\star(t) - \boldsymbol{\theta}(t) = \mathbf{Z}(t)\mathbf{G}_\star^{-1}(t)\mathbf{h}_\star(t). \tag{14}$$

where $\mathbf{h}_\star(t) = \sum_{i=-\infty}^{\infty}\lambda^{|t-i|}\boldsymbol{\psi}(i)v(i)$.

Using (11), the error equation (14) can be rewritten in the following equivalent form

$$\Delta\widehat{\boldsymbol{\theta}}_\star(t) = \mathbf{Z}_\lambda\mathbf{G}_\lambda^{-1}(t)\mathbf{h}_\lambda(t) \cong \left[\mathbf{\Phi}_0^{-1}\otimes\mathbf{f}_\lambda^{\mathrm{T}}(0)\right]\mathbf{h}_\lambda(t) \tag{15}$$

where

$$\mathbf{h}_\lambda(t) = \mathbf{H}(t)\mathbf{h}_\star(t) = \sum_{i=-\infty}^{\infty}\lambda^{|t-i|}\left[\boldsymbol{\varphi}(i)\otimes\mathbf{f}_\lambda(t-i)\right]v(i).$$

Since $\mathrm{E}[\mathbf{h}_\lambda(t)] = 0$, it holds that $\mathrm{E}[\widehat{\boldsymbol{\theta}}_\star(t)] \cong \boldsymbol{\theta}(t)$, i.e., the smoothed estimate is approximately unbiased. Exploiting assumptions (A1) and (A2), after elementary calculations, one arrives at

$$\mathrm{E}\left[\mathbf{h}_\lambda(t)\mathbf{h}_\lambda^{\mathrm{T}}(t)\right] = \sigma_v^2\mathbf{\Phi}_0\otimes\left[\sum_{i=-\infty}^{\infty}\lambda^{2|i|}\mathbf{f}_\lambda(i)\mathbf{f}_\lambda^{\mathrm{T}}(i)\right]$$

$$\begin{aligned}\mathrm{cov}\left[\Delta\widehat{\boldsymbol{\theta}}_\star(t)\right] &\cong \left[\mathbf{\Phi}_0^{-1}\otimes\mathbf{f}_\lambda^{\mathrm{T}}(0)\right]\mathrm{E}\left[\mathbf{h}_\lambda(t)\mathbf{h}_\lambda^{\mathrm{T}}(t)\right]\left[\mathbf{\Phi}_0^{-1}\otimes\mathbf{f}_\lambda(0)\right]\\ &\cong \frac{\sigma_v^2\mathbf{\Phi}_0^{-1}}{l_\star}\end{aligned} \tag{16}$$

where the quantity $l_\star$ denotes the equivalent estimation memory length of the EWBF smoother

$$l_\star^{-1} = \mathbf{f}_\lambda^{\mathrm{T}}(0)\left[\sum_{i=-\infty}^{\infty}\lambda^{2|i|}\mathbf{f}_\lambda(i)\mathbf{f}_\lambda^{\mathrm{T}}(i)\right]\mathbf{f}_\lambda(0) = \sum_{i=-\infty}^{\infty}k_\star^2(i).$$

Equivalent estimation memory, different from the so-called effective estimation memory, characterizes the amount of information about $\boldsymbol{\theta}(t)$ which is extracted from the input/output data as a result of applying the method of weighted basis functions [5]. One can easily show that when the true parameter trajectory cannot be exactly represented by the weighted sum of basis functions, the right-hand side of (16) is the *lower bound* on $\mathrm{cov}[\Delta\widehat{\boldsymbol{\theta}}_\star(t)]$.

The results derived above parallel those obtained earlier for the EWBF tracker [15]. In the case of tracking, one obtains

$$\bar{\boldsymbol{\theta}}_\pm(t) = \mathrm{E}\left[\widehat{\boldsymbol{\theta}}_\pm(t)\right] \cong \sum_{i=0}^{\infty}k_\pm(i)\boldsymbol{\theta}(t\pm i) \tag{17}$$

where

$$k_\pm(i) = \lambda^i\mathbf{f}^{\mathrm{T}}(0)\left[\sum_{j=0}^{\infty}\lambda^j\mathbf{f}(j)\mathbf{f}^{\mathrm{T}}(j)\right]^{-1}\mathbf{f}(i),\quad i\geq 0. \tag{18}$$

The corresponding steady state equivalent memory spans can be obtained from $l_\pm = \left[\sum_{i=0}^{\infty}k_\pm^2(i)\right]^{-1}$. Note that $k_-(i)\equiv k_+(i)$ and consequently $l_- = l_+$.

Expressions for $k_\star(i)$ and $l_\star$ derived for the basis (3) for different number of basis functions ($m = 1,2,3$), and the analogous results obtained earlier for the EWBF trackers, are summarized in Table 1. Interestingly, the same expressions for $k_\star(i)$ and $l_\star$ are obtained for $m=1$ and $m=2$. This means that no bias reduction can be expected after incorporation of the linear basis function $f_2(t) = t$ (but usually there is some variance reduction compared to the first-order estimator with the same equivalent memory).

Note that the steady state equivalent memory spans take the form $l_m^\pm = c_m^\pm/(1-\lambda)$ and $l_m^\star = c_m^\star/(1-\lambda)$, where $c_m^\pm$ and $c_m^\star$ are easily computable quantities that depend on the number of basis functions and the forgetting constant, and *do not* depend on the characteristics of the input-output data.

For Kalman algorithms based on (1) the situation is different. When $m=1$, one can show that the estimation memory takes the form $l_1 = c_1(\mathbf{\Phi}_0)/\sqrt{\xi}$ where c_1 is a constant that depends on the eigenvalues of the matrix $\mathbf{\Phi}_0^{1/2}$ – see [5]. No results are currently available for $m>1$, but there are some theoretical arguments which support the claim that $l_m = c_m(\mathbf{\Phi}_0)\xi^{-1/2m}$. Since the constants c_m are not known, the practically important memory scheduling problem is difficult to solve for this class of identification algorithms.

Remark: When estimation is carried out using the exponentially weighted basis function algorithms, the forgetting constant should be chosen with caution – the rules of thumb which work for the exponentially weighted least squares scheme (i.e., the first-order EWBF tracker) cannot be mechanically extended to higher-order EWBF estimators. For example, for $\lambda = 0.99$ the equivalent memory spans of the third-order estimators are equal to $l_\pm \cong 33$ samples and $l_\star \cong 48$ samples, rather than $l_\pm \cong 200$ samples and $l_\star \cong 400$ samples, respectively, as one might expect based on the experience with exponentially weighted least squares trackers/smoothers.

Table I
Associated impulse responses and equivalent memory spans of EWBF smoothers and trackers of different orders.

EWBF smoothers

$m=1$	$k_\star(i)=\lambda^{\lvert i\rvert}\,\frac{1-\lambda}{1+\lambda}$ $l_\star=\frac{(1+\lambda)^3}{(1-\lambda)(1+\lambda^2)}\cong\frac{4}{1-\lambda}$
$m=2$	$k_\star(i)=\lambda^{\lvert i\rvert}\,\frac{1-\lambda}{1+\lambda}$ $l_\star=\frac{(1+\lambda)^3}{(1-\lambda)(1+\lambda^2)}\cong\frac{4}{1-\lambda}$
$m=3$	$k_\star(i)=\lambda^{\lvert i\rvert}\,\frac{(1-\lambda)[1+10\lambda+\lambda^2-(1-\lambda)^2 i^2]}{(1+\lambda)(1+8\lambda+\lambda^2)}$ $l_\star\cong\frac{267}{800(1-\lambda)}$

EWBF trackers

$m=1$	$k_\pm(i)=\lambda^i(1-\lambda)$ $l_\pm=\frac{1+\lambda}{1-\lambda}\cong\frac{2}{1-\lambda}$
$m=2$	$k_\pm(i)=\lambda^i(1-\lambda)[1+\lambda-(1-\lambda)i]$ $l_\pm=\frac{(1+\lambda)^3}{(1-\lambda)(1+4\lambda+5\lambda^2)}\cong\frac{4}{5(1-\lambda)}$
$m=2$	$k_\pm(i)=\lambda^i(1-\lambda)[1+\lambda+\lambda^2$ $-\frac{3}{2}(1+\lambda)(1-\lambda)i+\frac{1}{2}(1-\lambda)^2 i^2]$ $l_\pm\cong\frac{16}{33(1-\lambda)}$

IV. Cooperative EWBF smoother

Estimation properties of EWBF algorithms depend on the choice of two design parameters: the forgetting constant λ and the number of basis functions m. For 'small' values of λ and/or for 'large' values of m, the estimation memory of the EWBF algorithm decreases. The resulting short-memory trackers/smoothers are 'fast' (yield small estimation bias) but 'inacurate' (yield large estimation variance). The converse is also true – long-memory algorithms are 'accurate' but'slow'. The best results are obtained if the values of λ and m are selected so as to match the degree and type of nonstationarity of the identified system, trading off the bias and variance error components. Optimization of the design parameters is possible using parallel estimation techniques. Following this idea, consider K simultaneously running EWBF smoothers, equipped with different settings $\{\lambda_k, m_k\}$ and yielding the estimates

$$
\begin{aligned}
\widehat{\boldsymbol{\theta}}_k^\star(t) &= \mathbf{Z}_k(t)\,[\mathbf{G}_k^\star(t)]^{-1}\,\mathbf{g}_k^\star(t)\\
\mathbf{Z}_k(t) &= \mathbf{I}_n\otimes\mathbf{f}_k^{\mathrm{T}}(t)\\
\mathbf{G}_k^\star(t) &= \sum_{i=-\infty}^{\infty}\lambda_k^{|t-i|}\boldsymbol{\psi}_k(i)\boldsymbol{\psi}_k^{\mathrm{T}}(i)\\
\mathbf{g}_k^\star(t) &= \sum_{i=-\infty}^{\infty}\lambda_k^{|t-i|}\boldsymbol{\psi}_k(i)y(i).
\end{aligned}
$$

We will combine these estimates using the cooperative smoothing strategy proposed in [13].

Cooperative smoothing is a Bayesian extension of the leave-one-out cross-validation approach to model selection. In this approach, credibility of each smoother $\widehat{\boldsymbol{\theta}}_k^\star(t)$ is assessed (locally) based on observation of matching errors $e_k^\circ(t)=y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}_k^\circ(t)$ – residual errors yielded by the 'holey' smoother $\widehat{\boldsymbol{\theta}}_k^\circ(t)$ associated with $\widehat{\boldsymbol{\theta}}_k^\star(t)$. Holey smoother associated with the EWBF algorithm, i.e., the one that excludes $y(t)$ from the set of measurements used for estimation of $\boldsymbol{\theta}(t)$, takes the form:

$$
\begin{aligned}
\widehat{\boldsymbol{\theta}}_k^\circ(t) &= \mathbf{Z}_k(t)\,[\mathbf{G}_k^\circ(t)]^{-1}\,\mathbf{g}_k^\circ(t)\\
\mathbf{G}_k^\circ(t) &= \mathbf{G}_k^\star(t)-\boldsymbol{\psi}_k(t)\boldsymbol{\psi}_k^{\mathrm{T}}(t)\\
\mathbf{g}_k^\circ(t) &= \mathbf{g}_k^\star(t)-\boldsymbol{\psi}_k(t)y(t).
\end{aligned}
$$

Suppose that the measurement noise $\{v(t)\}$ is a sequence of zero-mean independent random variables obeying the generalized normal law $v\sim\mathcal{GN}(\alpha,\beta)$:

$$
p(v;\alpha,\beta)=\frac{\beta}{2\alpha\Gamma(1/\beta)}\exp\left\{-\left(\frac{|v|}{\alpha}\right)^{\beta}\right\}
$$

where $\alpha>0$ is the unknown scale parameter, $\beta\ge 1$ is the known shape parameter, and $\Gamma(\cdot)$ denotes the Euler's gamma function. Generalized normal law incorporates such practically important distributions as Gaussian ($\beta=2$), Laplace ($\beta=1$), and uniform ($\beta\to\infty$) – see [16].

According to [13], the cooperative estimate of the parameter trajectory can be obtained in the form

$$
\widehat{\boldsymbol{\theta}}_{\mathrm{A}}(t)=\sum_{k=1}^{K}\mu_k^\star(t)\widehat{\boldsymbol{\theta}}_k^\star(t),\ \ \mu_k^\star(t)\propto\left[\sum_{i\in T_{\mathrm{d}}^\star(t)}|e_k^\circ(i)|^{\beta}\right]^{-M/\beta} \tag{19}
$$

where $\mu_k^\star(t)\ge 0, k=1,\ldots,K, \sum_{k=1}^{K}\mu_k^\star(t)=1$, denote credibility coefficients (related to posterior probabilities of different parameter 'patterns'), evaluated in the local decision window $T_{\mathrm{d}}^\star(t)=[t-l,t+l]$ of (user-dependent) width $M=2l+1$.

Similarly as in [13], one can show[1] that matching errors can be expressed in terms of the residual errors $e_k(t)=y(t)-\boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}_k^\star(t)$, namely

$$
e_k^\circ(t)=\frac{e_k(t)}{1-r_k(t)} \tag{20}
$$

where

$$
r_k(t)=\boldsymbol{\psi}_k^{\mathrm{T}}(t)\,[\mathbf{G}_k^\star(t)]^{-1}\,\boldsymbol{\psi}_k(t).
$$

Hence, credibility coefficients can be evaluated without implementing the corresponding holey smoothers.

V. Competitive EWBF smoother

The smoothing formula (8) is based on an implicit assumption that parameter changes are continuous, i.e., that the true parameter trajectory obeys the polynomial model at least in some neighborhood of t. Quite obviously, the presence of jumps in the parameter trajectory violates this 'homogeneity' assumption. Note, however, that when parameter changes are infrequent, namely when the time between subsequent jumps is greater than the equivalent memory length of the forward/backward EWBF trackers, at each time instant at least one of them generates accurate estimates of the parameter trajectory. In particular, the forward time tracker yields accurate estimates immediately before the jumps, and the backward time tracker – immediately after the jumps. The degree of accurateness, i.e., 'credibility' of different estimators, can be easily judged by the magnitude of the locally observed prediction errors. Following [6], the competitive smoother that combines results yielded by K forward

[1] After replacing the regression vector $\boldsymbol{\varphi}_k(t)$ with the generalized regression vector $\boldsymbol{\psi}_k(t)$, the proof is identical with that given in [13].

time and K backward time EWBF trackers can be set up as follows

$$\widehat{\boldsymbol{\theta}}_{\mathrm{B}}(t) = \sum_{k=1}^{K} \left[\mu_k^-(t)\widehat{\boldsymbol{\theta}}_k^-(t) + \mu_k^+(t)\widehat{\boldsymbol{\theta}}_k^+(t) \right]$$

$$\mu_k^{\pm}(t) \propto \left[\sum_{i \in T_{\mathrm{d}}^{\pm}(t)} |\varepsilon_k^{\pm}(i)|^{\beta} \right]^{-M/\beta} \tag{21}$$

where $T_{\mathrm{d}}^-(t) = [t-M+1, t]$ and $T_{\mathrm{d}}^+(t) = [t, t+M-1]$ denote the local decision windows of width M, and

$$\varepsilon_k^{\pm}(i) = y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(i)\widehat{\boldsymbol{\theta}}_k^{\pm}(i \pm 1),\ k = 1, \ldots, K$$

are the corresponding forward/backward prediction errors.

The term ‘competitive smoothing’ refers to the fact that, within this approach, the forward and backward algorithms compete with each other, rather than cooperate (since, at each time instant, only one of them is assumed to be based on the correct system description).

VI. Combined EWBF smoother

While the competitive smoother yields good estimation results in the regions of parameter jumps, for smooth parameter changes it may lead to parameter ‘jitter’ - the effect caused by random switching between the forward-time and backward-time algorithms – see [6]. The converse is true for the cooperative smoother: it yields good results for smooth parameter changes, but ‘blurs’ all step-like features of the estimated parameter trajectory.

To obtain smoothing algorithm that works satisfactorily in both cases mentioned above, one can fuse the cooperative smoother (18) with the competitive smoother (21). Using the cooperative smoothing rule once again, one arrives at the following formula which will be further referred to as a combined smoother

$$\widehat{\boldsymbol{\theta}}_{\mathrm{C}}(t) = \mu_{\mathrm{A}}(t)\widehat{\boldsymbol{\theta}}_{\mathrm{A}}(t) + \mu_{\mathrm{B}}(t)\widehat{\boldsymbol{\theta}}_{\mathrm{B}}(t) \tag{22}$$

where

$$\mu_{\mathrm{A}}(t) = \frac{\eta_{\mathrm{A}}(t)}{\eta_{\mathrm{A}}(t) + \eta_{\mathrm{B}}(t)}\ , \quad \mu_{\mathrm{B}}(t) = \frac{\eta_{\mathrm{B}}(t)}{\eta_{\mathrm{A}}(t) + \eta_{\mathrm{B}}(t)}$$

$$\eta_{\mathrm{A/B}} = \left[\sum_{i \in T_{\mathrm{d}}^{\star}(t)} |e_{\mathrm{A/B}}^{\circ}(i)|^{\beta} \right]^{-M/\beta}$$

and the corresponding matching errors are given by

$$e_{\mathrm{A}}^{\circ}(t) = \sum_{k=1}^{K} \mu_k^{\star}(t) e_k^{\circ}(t)$$

$$e_{\mathrm{B}}^{\circ}(t) = \sum_{k=1}^{K} \left[\mu_k^-(t)\varepsilon_k^-(t) + \mu_k^+(t)\varepsilon_k^+(t) \right].$$

Note that the computational load of the combined smoother grows linearly with the number of component algorithms K. The analogous count for the IMM smoother proposed in [8] is proportional to K^2.

VII. Computer simulations

To check performance of smoothing algorithms, the following two-tap FIR system (inspired by channel equalization applications) was simulated

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

where $u(t) = \pm 1$, $\sigma_u^2 = 1$, denotes the pseudo-random binary signal (PRBS) – the sequence transmitted over a telecommunication channel – and $v(t)$ denotes a zero-mean white noise.

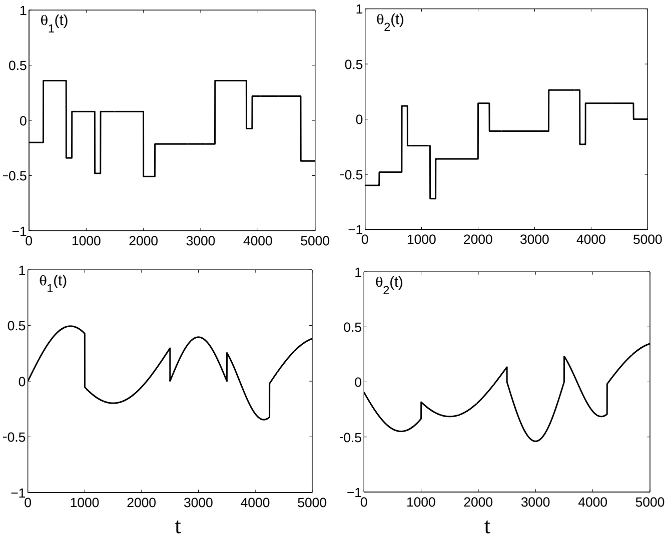

Fig. 1. Two variants of parameter changes used in computer simulations: piecewise constant (two upper plots) and piecewise sinusoidal (two lower plots).

Two types of parameter changes were considered: piecewise constant (I), and piecewise sinusoidal (II) – see Fig. 1. Due to appropriate scaling, all parameter trajectories have the same L_2 norm.

For each of the compared algorithms the steady state accumulated mean-squared parameter estimation errors $\mathrm{E}_v\{\sum_{t=101}^{4900} \parallel \widehat{\boldsymbol{\theta}}(t|5000) - \boldsymbol{\theta}(t) \parallel^2\}$ were computed (to eliminate transient effects, the evaluation was restricted to the interval $[101, 4900]$). Ensemble averaging $\mathrm{E}_v(\cdot)$ was performed over 50 realizations of the measurement noise $\{v(t)\}$. The procedure was repeated for each of 3 noise intensities ranging from $\sigma_v = 0.05$ (SNR=26 dB) to $\sigma_v = 0.30$ (SNR=10.5 dB). The distribution of noise was either Gaussian ($\beta = 2$) or Laplacian ($\beta = 1$). The width of the decision window was equal to $M = 21$.

Table 2 summarizes results obtained for 9 EWBF smoothers (S_j^k, $j, k = 1, \ldots, 3$), the cooperative smoother (S_{A}), the competitive smoother (S_{B}) and the combined smoother (S_{C}). The first 3 EWBF algorithms (S_1^1, S_1^2, S_1^3) were based on the polynomial model of order 0 ($m = 1, \lambda_1^1 = 0.818, \lambda_1^2 = 0.920, \lambda_1^3 = 0.975$). The next 3 EWBF algorithms (S_2^1, S_2^2, S_2^3) were based on the polynomial model of order 1 ($m = 2, \lambda_1^1 = 0.936, \lambda_1^2 = 0.973, \lambda_1^3 = 984$). The last 3 EWBF algorithms (S_3^1, S_3^2, S_3^3) were based on the polynomial model of order 2 ($m = 3, \lambda_1^1 = 978, \lambda_1^2 = 0.991, \lambda_1^3 = 0.995$). The forgetting constants λ_j^k of constituent smoothers were not optimized in any way – the corresponding values were chosen so that within each group of algorithms $\{S_1^1, S_2^1, S_3^1\}$, $\{S_1^2, S_2^2, S_3^2\}$ and $\{S_1^3, S_2^3, S_3^3\}$ the equivalent memory spans of the corresponding EWBF trackers were the same and equal to 10 samples, 30 samples and 90 samples, respectively. Typical estimation results (Gaussian noise, $\sigma_v = 0.2$) are shown in Fig. 2.

Note that the combined smoother is uniformly better than all other smoothers, including all component smoothers, i.e., it provides the smallest estimation errors in all cases considered. An obvious advantage of the combined smoother is its increased robustness to unknown and possibly time-varying degree of nonstationarity and the mode of variation of the identified system.

VIII. Conclusion

We have shown that, by combining the results yielded by several exponentially weighted basis function parameter trackers/smoothers (run simultaneously and equipped with different settings), one obtains a new fixed-interval smoothing algorithm with improved estimation

Table 2
Comparison of parameter estimation errors obtained for 3 EWBF smoothers based on the polynomial model of order 0 (S_1^1, S_1^2, S_1^3), 3 EWBF smoothers based on the polynomial model of order 1 (S_2^1, S_2^2, S_2^3), 3 EWBF smoothers based on the polynomial model of order 2 (S_3^1, S_3^2, S_3^3), cooperative smoother (S_{A}), competitive smoother (S_{B}) and combined smoother (S_{C}). Simulations were performed for 2 variants of parameter changes (I, II). The best results are shown in boldface.

Gaussian noise

T	σ_v	S_1^1	S_1^2	S_1^3	S_2^1	S_2^2	S_2^3	S_3^1	S_3^2	S_3^3	S_{A}	S_{B}	S_{C}
I	0.05	5.48	16.82	54.80	13.75	44.78	125.92	35.99	110.48	228.80	5.45	1.04	**1.00**
	0.15	7.02	17.31	54.97	14.39	44.98	125.99	36.24	110.56	228.88	6.19	3.81	**3.44**
	0.30	12.13	18.88	55.32	16.42	45.47	126.02	36.94	110.71	228.78	8.47	8.40	**6.91**
II	0.05	1.13	2.60	7.90	2.13	6.41	28.53	5.04	18.16	108.53	1.10	0.24	**0.14**
	0.15	2.66	3.09	8.07	2.76	6.61	28.62	5.29	18.28	108.51	1.69	1.59	**0.99**
	0.30	7.75	4.76	8.60	4.86	7.27	28.84	6.12	18.54	108.66	3.50	5.38	**3.46**

Laplacian noise

T	σ_v	S_1^1	S_1^2	S_1^3	S_2^1	S_2^2	S_2^3	S_3^1	S_3^2	S_3^3	S_{A}	S_{B}	S_{C}
I	0.05	5.47	16.81	54.80	13.74	44.77	125.93	35.98	110.49	228.82	5.38	1.38	**1.35**
	0.15	7.04	17.34	54.93	14.40	44.96	125.90	36.27	110.50	228.74	6.07	4.02	**3.69**
	0.30	12.29	19.10	55.57	16.60	45.72	126.15	37.18	110.82	228.83	8.15	8.45	**7.08**
II	0.05	1.13	2.60	7.90	2.14	6.41	28.54	5.04	18.16	108.54	1.02	0.21	**0.12**
	0.15	2.66	3.08	8.05	2.76	6.59	28.64	5.28	18.24	108.67	1.52	1.39	**0.84**
	0.30	7.83	4.84	8.66	4.95	7.35	28.81	6.23	18.51	108.68	3.03	4.79	**3.04**

capabilities. The proposed smoother can be applied to noncausal identification of nonstationary stochastic systems subject to both smooth and abrupt parameter changes. Additionally, it accounts for the distribution of measurement noise.

References

[1] G.B. Giannakis and C. Tepedelenliovglu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proc. IEEE*, vol. 86, pp. 1969–1986, 1998.

[2] R. Shiller, "A distributed lag estimator derived from smoothness priors," *Econometrica*, vol. 41, pp. 775–778, 1973.

[3] J.P. Norton, "Optimal smoothing in the identification of linear time-varying systens," *Proc. IEE*, vol. 122, 663–668, 1975.

[4] G. Kitagawa and W. Gersch, "A smoothness priors time-varying AR coefficient modeling of nonstationary covariance time series," *IEEE Trans. Automat. Contr.*, vol. 30, pp. 48–56, 1985.

[5] M. Niedźwiecki, *Identification of Time-varying Processes*. New York: Wiley, 2000.

[6] M. Niedźwiecki, "Identification of time-varying systems with abrupt parameter changes," *Automatica*, vol. 30, pp. 447–459, 1994.

[7] M. Niedźwiecki, "Locally adaptive cooperative Kalman smoothing and its application to identification of nonstationary stochastic systems," *IEEE Trans. Signal Process.*, vol. 60, pp. 48–59, 2012.

[8] H.A.P. Blom and Y. Bar-Shalom, "The interacting multiple-model algorithm for systems with Markovian switching coefficients," *IEEE Trans. Automat. Contr.*, vol. 33, pp. 780–783 , 1988.

[9] M. Niedźwiecki and P. Suchomski, "On parallel estimation approach to adaptice filtering," *Bull. Polish Acad. Sci.*, vol. 44, pp. 385–397, 1996.

[10] R.E. Helmick, W.D. Blair and S.A. Hoffman, " Fixed-interval smoothing for Markovian switching systems," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1845–1855 , 1995.

[11] A.G. Poulimenos and S.D. Fassois, "Parametric time-domain methods for non-stationary random vibration modelling and analysis – A critical survey and comparison," *Mech. Syst. Signal Process.*, vol. 20, pp. 763–816, 2006.

[12] T. Subba Rao, "The fitting of non-stationary time-series models with time-dependent parameters," *J. R. Statist. Soc. B*, vol. 32, pp. 312–322, 1970.

[13] M. Niedźwiecki and S. Gackowski, "On noncausal weighted least squares identification of nonstationary stochastic systems," *Automatica*, vol. 47, pp. 2239–2245, 2011.

[14] W. Pruitt, "Summability of independent random variables," *J. Math. Mech.*, vol. 15, pp. 769–776, 1966.

[15] M. Niedźwiecki, "Recursive functional series modelling estimators for adaptive control of time-varying plants - more bad news than good?," IEEE Trans. Automatic Control, vol 35, pp. 610–616, 1990.

[16] N. Saralees, "A generalized normal distribution," *J. Appl. Stat.*, vol. pp. 32, 685–694, 2005.

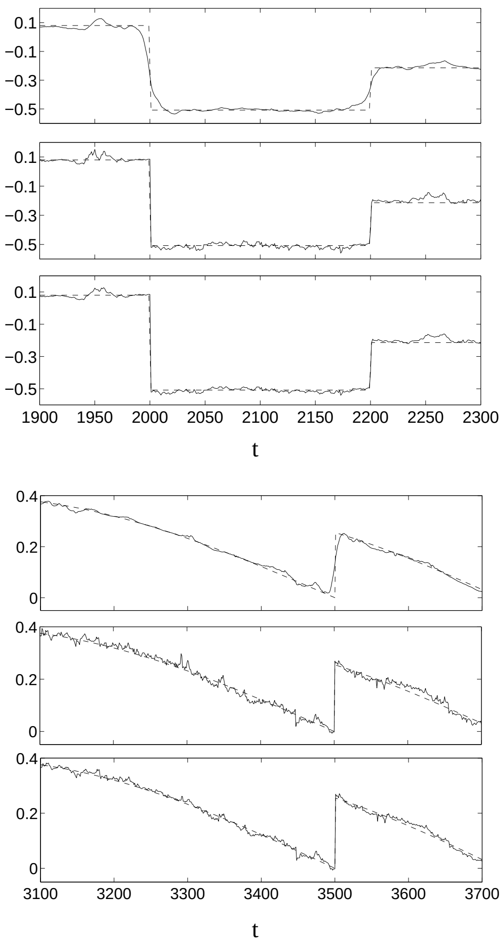

Fig. 2. A fragment of the true piecewise constant (upper figure) and piecewise sinusoidal (lower figure) parameter trajectories (broken lines), along with their estimates (solid lines) obtained using the cooperative smoother (upper plots), competitive smoother (middle plots), and combined smoother (bottom plots).