A new look at the statistical identification of nonstationary systems

Maciej Niedźwiecki [a], Marcin Ciołek [a], Artur Gańcza [a]

[a] *Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control, Gdańsk University of Technology, Narutowicza 11/12, 80-233 Gdańsk , Poland, Tel: + 48 58 3472519; fax: +48 58 3415821*

Abstract

The paper presents a new, two-stage approach to identification of linear time-varying stochastic systems, based on the concepts of preestimation and postfiltering. The proposed preestimated parameter trajectories are unbiased but have large variability. Hence, to obtain reliable estimates of system parameters, the preestimated trajectories must be further filtered (postfiltered). It is shown how one can design and optimize such postfilters using the basis function framework. The proposed solution to adaptive tuning of postfilter settings is based on parallel estimation and cross-validatory analysis. When compared with the classical solutions to the problem of parameter tracking, the new approach offers, without compromising good tracking performance, significant computational savings, higher numerical robustness and greater flexibility.

Key words: identification of nonstationary systems, basis functions, parallel estimation, cross-validatory analysis

1 Introduction

Finite memory estimators, used to identify linear time-varying (LTV) stochastic systems, can be designed in several ways. First, if system parameters vary sufficiently slowly, their estimation can be carried out using localized versions of classical identification methods, such as weighted least squares (WLS). Secondly, one can adopt an explicit model ("hypermodel") of parameter variation, either deterministic or stochastic. In the first case, system parameters can be modeled as linear combinations of known functions of time, called basis functions (BF) (Rao, 1970), (Mendel, 1973), (Liporace, 1975), (Grenier, 1981), (Hall & Oppenheim, 1983), (Niedźwiecki, 1988b), (Mrad et al., 1998), (Zou et al., 2003), (Poulimenos & Fassois, 2006). In the second case the problem of parameter tracking is formulated as a problem of statistical filtering in the state space (Norton, 1975), (Young, 1984), (Kitagawa & Gersch, 1985), (Niedźwiecki, 2012).

The approach described in this paper differs from the ones mentioned above. It is based on two concepts: preestimation and postfiltering. Preestimates are raw, very noisy but unbiased (approximately) estimates of parameter trajectories. Due to the second property they can be also called, using the term coined by (Bellegarda & Farden, 1988), the maximum bandwidth estimators of system parameters. Preestimators are "unprejudiced", in the sense that their unbiasedness property holds no matter how system parameters change over time. Therefore, using preestimators, one can "X-ray" the structure of system parameter variation without making any assumptions about its functional form, degree of smoothness etc.

The second step of the proposed identification procedure, called postfiltering, consists in denoising the preestimated parameter trajectories. We will show how one can design postfilters which yield approximately the same parameter tracking results as the state-of-the-art (computationally much more demanding) local basis function (LBF) algorithms proposed recently by (Niedźwiecki & Ciołek, 2019). We will also show how design parameters of such postfilters can be adaptively adjusted to the rate and form of system nonstationarity. The proposed adjustment mechanism is based on parallel estimation and cross-validation.

The two-stage, preestimate/postfilter approach has several advantages over the classical identification procedures. It offers, without compromising good tracking performance, significant computational savings (both preestimators and postfilters are computationally cheap and can be made recursively computable), higher numerical robustness and greater flexibility. The proposed approach generalizes, to the noncausal estimation case, some ear-

lier results and concepts developed for causal estimators (Niedźwiecki, 1990), (Niedźwiecki & Kłaput, 2002). It should be also stressed that it bears no resemblance to the multi-step least squares methods developed recently for the purpose of identification of time-invariant systems (Galrinho et al., 2014).

2   Two-stage identification procedure

Consider a nonstationary linear system governed by

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

where $t = \ldots, -1, 0, 1, \ldots$ denotes discrete (normalized) time, $\boldsymbol{\theta}(t) = [\theta_1(t), \ldots, \theta_n(t)]^{\mathrm{T}}$ denotes the vector of time-varying system coefficients, $\boldsymbol{\varphi}(t) = [u(t-1), \ldots, u(t-n)]^{\mathrm{T}}$ denotes regression vector made up of past values of the input signal $u(t)$, and $\{e(t)\}$ denotes measurement noise. Furthermore, assume that:

(A1) $\{u(t)\}$ is a zero-mean wide sense stationary Gaussian sequence with an exponentially decaying autocorrelation function $r_u(i) = \mathrm{E}[u(t)u(t-i)]: \exists\, 0 < \alpha < \infty,\ 0 < \gamma < 1\ :\ |r_u(i)| \leq \alpha\gamma^{|i|},\ \forall i$.

(A2) $\{e(t)\}$, independent of $\{u(t)\}$, is a sequence of zero-mean independent and identically distributed random variables with variance σ_e^2.

(A3) $\{\boldsymbol{\theta}(t)\}$ is a uniformly bounded sequence, independent of $\{u(t)\}$ and $\{e(t)\}$.

2.1   Preestimation

To arrive at the concept of preestimation we will refer to some known properties of exponentially weighted least squares (EWLS) estimators defined as

$$\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t) = \arg\min_{\boldsymbol{\theta}} \sum_{i=0}^{t-1} \lambda_0^i [y(t-i) - \boldsymbol{\varphi}^{\mathrm{T}}(t-i)\boldsymbol{\theta}]^2$$
$$= \mathbf{R}^{-1}(t)\mathbf{r}(t) \qquad (2)$$

where $\mathbf{R}(t) = \sum_{i=0}^{t-1} \lambda_0^i \boldsymbol{\varphi}(t-i)\boldsymbol{\varphi}^{\mathrm{T}}(t-i)$, $\mathbf{r}(t) = \sum_{i=0}^{t-1} \lambda_0^i \boldsymbol{\varphi}(t-i)y(t-i)$, and λ_0, $0 < \lambda_0 < 1$, denotes the so-called forgetting constant. The EWLS estimates can be computed recursively using the well-known recursive least squares algorithm (Söderström & Stoica, 1988).

According to (Niedźwiecki, 1988a), under assumptions (A1)–(A3) it holds that

$$\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t) \cong \frac{1}{L_t}\boldsymbol{\Phi}_0^{-1}\sum_{i=0}^{t-1}\lambda_0^i\boldsymbol{\varphi}(t-i)y(t-i) \qquad (3)$$

where $L_t = \sum_{i=0}^{t-1}\lambda_0^i = \lambda_0 L_{t-1} + 1$ denotes the effective width of the exponential window and $\boldsymbol{\Phi}_0 = \mathrm{E}[\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)] > 0$. Using (3), one arrives at

$$L_t\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t) \cong \lambda_0 L_{t-1}\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1) + \boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t)y(t) \qquad (4)$$

which allows one to define the preestimate $\boldsymbol{\theta}^*(t)$ of $\boldsymbol{\theta}(t)$ in the form

$$\boldsymbol{\theta}^*(t) = L_t\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t) - \lambda_0 L_{t-1}\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1)$$
$$\cong \boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t)y(t) = \boldsymbol{\theta}^\dagger(t). \qquad (5)$$

For large values of t the effective window width reaches its steady state value equal to $L_\infty = 1/(1 - \lambda_0)$. In this case the preestimate (5) can be evaluated using the following simplified (steady state) formula

$$\boldsymbol{\theta}^*(t) = \frac{1}{1 - \lambda_0}[\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t) - \lambda_0\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1)]. \qquad (6)$$

Note that

$$\mathrm{E}[\boldsymbol{\theta}^*(t)] \cong \mathrm{E}[\boldsymbol{\theta}^\dagger(t)] = \mathrm{E}[\boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)]\boldsymbol{\theta}(t)$$
$$+ \mathrm{E}[\boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t)e(t)].$$

where, here and later, the expectation is over $\Omega_k(t) = \{\boldsymbol{\varphi}(t+i), e(t+i), i \in I_k\}$. Hence, under assumptions (A1)–(A3), it holds that $\mathrm{E}[\boldsymbol{\theta}^*(t)] \cong \boldsymbol{\theta}(t)$, meaning that $\boldsymbol{\theta}^*(t)$ is an (approximately) unbiased, i.e., maximum bandwidth, estimator of $\boldsymbol{\theta}(t)$, which can be written down as

$$\boldsymbol{\theta}^*(t) = \boldsymbol{\theta}(t) + \mathbf{z}^*(t) \qquad (7)$$

where $\mathbf{z}^*(t)$ is a zero-mean noise. The unbiasedness property comes at the cost of a large variability. In the case of the oracle preestimate $\boldsymbol{\theta}^\dagger(t)$, which requires prior knowledge of the covariance matrix $\boldsymbol{\Phi}_0$, it holds that

$$\mathbf{z}^\dagger(t) = \boldsymbol{\theta}^\dagger(t) - \boldsymbol{\theta}(t) = \boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t)e(t)$$
$$+ [\boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) - \mathbf{I}_n]\boldsymbol{\theta}(t). \qquad (8)$$

Under Gaussian assumptions [cf. (A1)], one obtains

$$\mathrm{E}\{\mathbf{z}^\dagger(t)[\mathbf{z}^\dagger(t)]^{\mathrm{T}}\} = \mathrm{cov}[\boldsymbol{\theta}^\dagger(t)] = \sigma_e^2\boldsymbol{\Phi}_0^{-1}$$
$$+ \boldsymbol{\theta}^{\mathrm{T}}(t)\boldsymbol{\Phi}_0\boldsymbol{\theta}(t)\boldsymbol{\Phi}_0^{-1} + \boldsymbol{\theta}(t)\boldsymbol{\theta}^{\mathrm{T}}(t). \qquad (9)$$

Note that the covariance matrix of $\mathbf{z}^\dagger(t)$ depends on time-varying parameters of the identified system. The situation does not change if the covariance matrix $\boldsymbol{\Phi}_0$ in (5) is replaced with its exponentially weighted estimate $\widehat{\boldsymbol{\Phi}}_0(t) = \mathbf{R}(t)/L_t$, i.e., if the preestimate is defined in the following "direct" form

$$\widetilde{\boldsymbol{\theta}}(t) = \widehat{\boldsymbol{\Phi}}_0^{-1}(t)\boldsymbol{\varphi}(t)e(t) \qquad (10)$$

leading to

$$\widetilde{\mathbf{z}}(t) = \widetilde{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t) = \widehat{\boldsymbol{\Phi}}_0^{-1}(t)\boldsymbol{\varphi}(t)e(t)$$
$$+ [\widehat{\boldsymbol{\Phi}}_0^{-1}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) - \mathbf{I}_n]\boldsymbol{\theta}(t). \qquad (11)$$

However, in the case of the proposed "indirect" preestimate $\boldsymbol{\theta}^*(t)$, obtained by inverse filtering of the EWLS estimate of $\theta(t)$, the situation is different. It can be easily checked that $\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t) = \widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1) + \mathbf{R}^{-1}(t)\boldsymbol{\varphi}(t)\varepsilon(t)$, where $\varepsilon(t) = y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1)$ denotes the one-step-ahead prediction error. Consequently

$$\boldsymbol{\theta}^*(t) = \widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1) + \widehat{\boldsymbol{\Phi}}_0^{-1}(t)\boldsymbol{\varphi}(t)\varepsilon(t). \qquad (12)$$

Combining (12) with (1), one arrives at

$$\begin{aligned}
\mathbf{z}^*(t) = \boldsymbol{\theta}^*(t) - \boldsymbol{\theta}(t) &= \widehat{\boldsymbol{\Phi}}_0^{-1}(t)\boldsymbol{\varphi}(t)e(t) \\
&+ [\widehat{\boldsymbol{\Phi}}_0^{-1}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) - \mathbf{I}_n][\boldsymbol{\theta}(t) - \widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1)].
\end{aligned} \qquad (13)$$

It is instructive to compare the estimation error expressions derived for the direct and indirect preestimates (note that both expressions are exact, not approximate). While the second term on the right hand side of (11) depends on $\boldsymbol{\theta}(t)$, the analogous term in (13) depends on $\boldsymbol{\theta}(t) - \widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1)$. Since typically $||\boldsymbol{\theta}(t) - \widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1)|| \ll ||\boldsymbol{\theta}(t)||$, one can expect variability of indirect preestimates to be much smaller and less dependent on $\boldsymbol{\theta}(t)$ than variability of direct preestimates. Simulation results fully support this claim.

It was observed that the best preestimation results can be obtained for small values of L_∞, such as $L_\infty = 10$ ($\lambda_0 = 0.9$) – even if the number of estimated coefficients is large. It is not difficult to explain this. Since $\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t)$ is a causal estimator, incorporating only past data samples $\{y(i), u(i), i \leq t\}$, when L_∞ is increased, the mean square deviation of $\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t)$ from $\boldsymbol{\theta}(t)$ becomes quickly dominated by the bias error caused mainly by the fact that the estimated parameter trajectory lags behind the true trajectory. Hence, to achieve a good bias-variance trade-off, the value of L_∞ should be relatively small. For small values of L_∞ the preestimation noise is dominated by the first term on the right hand side of (13), equal to $\mathbf{z}_0(t) = \widehat{\boldsymbol{\Phi}}_0^{-1}(t)\boldsymbol{\varphi}(t)e(t) \cong \boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t)e(t)$, which can be recognized as white noise with covariance matrix $\boldsymbol{\Phi}_0^{-1}\sigma_e^2$.

The preestimation technique will be illustrated with the results obtained for a nonstationary two-tap FIR system governed by

$$y(t) = \theta_1(t)u(t-1) + \theta_2(t)u(t-2) + e(t). \qquad (14)$$

The applied stationary input signal was autoregressive Gaussian: $u(t) = 0.8u(t-1) + v(t)$, $\mathrm{var}[v(t)] = 1$, where $\{v(t)\}$ denotes white noise independent of $\{e(t)\}$. System parameter $\theta_1(t)$ was modeled as a sinusoidal linear chirp, and parameter $\theta_2(t)$ – as an inverted triangular linear chirp (see Fig. 1). The variance of the measurement noise was set to $\sigma_e^2 = 0.01$, which corresponds to the average signal-to-noise ratio SNR=25 dB

$$\mathrm{SNR} = \frac{1}{T_s} \sum_{i=1}^{T_s} \frac{\mathrm{E}\{[\boldsymbol{\varphi}^{\mathrm{T}}(i)\boldsymbol{\theta}(i)]^2\}}{\sigma_e^2} = \frac{1}{T_s} \sum_{i=1}^{T_s} \frac{\boldsymbol{\theta}^T(i)\boldsymbol{\Phi}_0\boldsymbol{\theta}(i)}{\sigma_e^2}$$
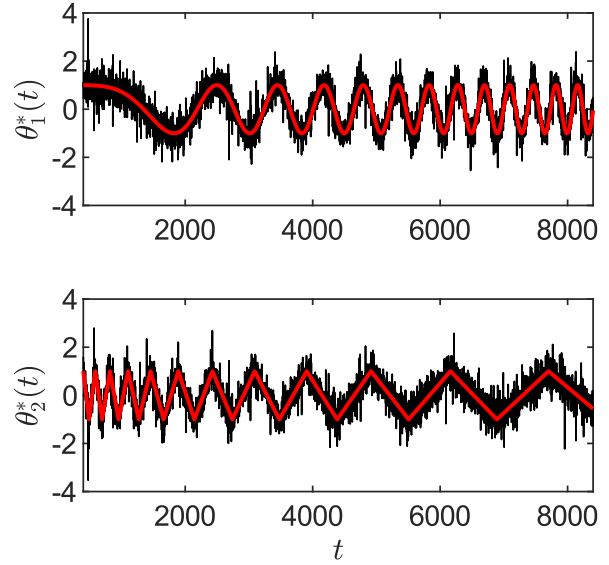

Fig. 1. Preestimated parameter trajectories of a nonstationary two-tap finite impulse response system (SNR=25 dB). Preestimates (black lines) are superimposed on true parameter trajectories (red lines).

where T_s denotes the length of the simulation interval.

The preestimates obtained via inverse filtering of EWLS estimates ($\lambda_0 = 0.9$) are shown in Fig. 1. Note that preestimates provide interesting insights into the structure of parameter variation, and they do so without making any assumptions about the speed and mode of parameter variation. Additionally, such a "prescreening" is provided separately for each system coefficient, which allows one to individually adjust the post-processing scheme (as different components of the parameter vector may require using different algorithm settings).

2.2 Postfiltering

Consider the j-th component of $\boldsymbol{\theta}(t)$. According to (7), the preestimated trajectory $\{\theta_j^*(t)\}$ can be regarded as a noisy version of the true trajectory $\{\theta_j(t)\}$

$$\theta_j^*(t) = \theta_j(t) + z_j^*(t)$$

where $z_j^*(t)$ is a zero-mean noise – the j-th component of $\mathbf{z}^*(t)$. Denoising of $\{\theta_j^*(t)\}$ will be carried out using the local basis function (LBF) approach (Niedźwiecki & Ciołek, 2019). In this approach each parameter trajectory $\{\theta_j(t)\}$ is modeled, in the local analysis interval $T_k(t) = [t-k, t+k]$ centered at t, as a linear combination of known, linearly independent functions of time $F_{m|k} = \{f_{1|k}(i), \ldots, f_{m|k}(i), i \in I_k = [-k, k]\}$, further referred to as basis functions, i.e., it is assumed that

$$\begin{aligned}
\theta_j(t+i) &= \mathbf{f}_{m|k}^{\mathrm{T}}(i)\boldsymbol{\beta}_{j;m|k}, \quad i \in I_k \\
j &= 1, \ldots, n.
\end{aligned} \qquad (15)$$

where $\mathbf{f}_{m|k}(i) = [f_{1|k}(i), \ldots, f_{m|k}(i)]^{\mathrm{T}}$.

The most frequently used basis functions are powers of time (polynomial basis), sine and cosine functions (Fourier basis), Slepian (prolate spheroidal) functions and Walsh functions (Grenier, 1981), (Zou et al., 2003). The subspace spanned by basis functions will be denoted by $\mathcal{F}_{m|k}$.

Denote by $\{w_k(i), i \in I_k\}$, $w_k(0) = 1$, a nonnegative, symmetric bell-shaped window of width $2k+1$ which will be used to put more emphasis on data gathered at instants close to t. Without any loss of generality we will assume that basis functions form a w-orthonormal basis of $\mathcal{F}_{m|k}$ obeying the condition

$$\sum_{i=-k}^{k} w_k(i)\mathbf{f}_{m|k}(i)\mathbf{f}_{m|k}^{\mathrm{T}}(i) = \mathbf{I}_m. \qquad (16)$$

Furthermore, to allow for asymptotic reasoning, we will assume that the basis set $F_{m|k}$ is a result of orthonormalization (carried out using, for example, the Gram-Schmidt procedure) of the set $G_{m|k} = \{g_{1|k}(i), \ldots, g_{m|k}(i), i \in I_k\}$ obtained by sampling the continuous-time basis generating functions $g_l^0(s), s \in [-1, 1]$: $g_{l|k}(i) = g_l^0(i/k), l = 1, \ldots, m, i \in I_k$. When $g_l^0(s) = s^{l-1}$, one arrives at the frequently used polynomial basis

$$G_{m|k} = \left\{ 1, i/k, \ldots, (i/k)^{m-1}, i \in I_k \right\}. \qquad (17)$$

Note that in this case the model (15) can be regarded as a local Taylor series approximation of the parameter trajectory $\{\theta_j(t)\}$. Finally, we will assume that $w_k(i) = w^0(i/k)$, where $w^0(s), s \in [-1, 1]$ denotes the continuous-time window generating function.

The local estimate of $\boldsymbol{\beta}_{j;m|k}$ can be obtained using the method of weighted least squares, leading to

$$\widehat{\theta}_{j;m|k}(t) = \mathbf{f}_{m|k}^{\mathrm{T}}(0)\widehat{\boldsymbol{\beta}}_{j;m|k}(t) \qquad (18)$$

$$\begin{aligned} &\widehat{\boldsymbol{\beta}}_{j;m|k}(t) \\ &= \arg\min_{\boldsymbol{\beta}_{j;m|k}} \sum_{i=-k}^{k} w_k(i)[\theta_j^*(t+i) - \mathbf{f}_{m|k}^{\mathrm{T}}(i)\boldsymbol{\beta}_{j;m|k}]^2 \\ &= \mathbf{Q}_{m|k}^{-1}\mathbf{q}_{j;m|k}(t) \end{aligned} \qquad (19)$$

where $\mathbf{Q}_{m|k} = \sum_{i=-k}^{k} w_k(i)\mathbf{f}_{m|k}(i)\mathbf{f}_{m|k}^{\mathrm{T}}(i) = \mathbf{I}_m$ and

$$\mathbf{q}_{j;m|k}(t) = \sum_{i=-k}^{k} w_k(i)\mathbf{f}_{m|k}(i)\theta_j^*(t+i). \qquad (20)$$

Combining (18), (19) and (20), one arrives at the solution which – for the reasons explained in the next section – will

be further referred to as fast local basis function (fLBF) estimator

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_{m|k}^{\mathrm{fLBF}}(t) &= [\widehat{\theta}_{1;m|k}(t), \ldots, \widehat{\theta}_{n;m|k}(t)]^{\mathrm{T}} \\ &= \sum_{i=-k}^{k} h_{m|k}(i)\boldsymbol{\theta}^*(t+i). \end{aligned} \qquad (21)$$

where

$$h_{m|k}(i) = w_k(i)\mathbf{f}_{m|k}^{\mathrm{T}}(0)\mathbf{f}_{m|k}(i), \quad i \in I_k. \qquad (22)$$

It can be easily shown that if a constant function $f(i) = c, \forall i \in I_k$, belongs to the subspace $\mathcal{F}_{m|k}$, as in the case of the basis (17), it holds that

$$\sum_{i=-k}^{k} h_{m|k}(i) = 1 \qquad (23)$$

which means that $h_{m|k}(i)$ is an impulse response of a low-pass FIR filter. For the polynomial basis (17) and cosinusoidal window $w_k(i) = \cos\frac{\pi i}{2k}$ the corresponding impulse responses are shown in Fig. 2.

2.3 Bias and variance

Combining (21) with (1) and (5), one arrives under (A1)–(A3) at

$$\begin{aligned} \bar{\boldsymbol{\theta}}_{m|k}^{\mathrm{fLBF}}(t) &= \mathrm{E}[\widehat{\boldsymbol{\theta}}_{m|k}^{\mathrm{fLBF}}(t)] \\ &\cong \sum_{i=-k}^{k} h_{m|k}(i)\boldsymbol{\Phi}_0^{-1}\mathrm{E}[\boldsymbol{\varphi}(t+i)\boldsymbol{\varphi}^{\mathrm{T}}(t+i)]\boldsymbol{\theta}(t+i) \\ &\quad + \sum_{i=-k}^{k} h_{m|k}(i)\boldsymbol{\Phi}_0^{-1}\mathrm{E}[\boldsymbol{\varphi}(t+i)e(t+i)] \\ &= \sum_{i=-k}^{k} h_{m|k}(i)\boldsymbol{\theta}(t+i) \end{aligned} \qquad (24)$$

which means that the mean path of fLBF estimates can be approximately viewed as an output of a linear noncausal filter with impulse response $\{h_{m|k}(i)\}$, excited by the process $\{\boldsymbol{\theta}(t)\}$. This relationship, which holds for *any* parameter trajectory, has a straightforward geometric interpretation: the right hand side of (24) is a result of orthogonal projection, evaluated at the instant t, of $\{\boldsymbol{\theta}(t+i), i \in I_k\}$ on the subspace $\mathcal{F}_{m|k}$. It can be used to quantify bias errors.

When system parameters can be exactly modeled as linear combinations of basis functions, i.e., when system coefficients obey (15), one obtains [cf. (16)]

$$\begin{aligned} \bar{\theta}_{j;m|k}(t) &= \mathbf{f}_{m|k}^{\mathrm{T}}(0)\left[\sum_{i=-k}^{k} w_k(i)\mathbf{f}_{m|k}(i)\mathbf{f}_{m|k}^{\mathrm{T}}(i)\right]\boldsymbol{\beta}_{j;m|k} \\ &= \mathbf{f}_{m|k}^{\mathrm{T}}(0)\boldsymbol{\beta}_{j;m|k} = \theta_j(t) \end{aligned}$$
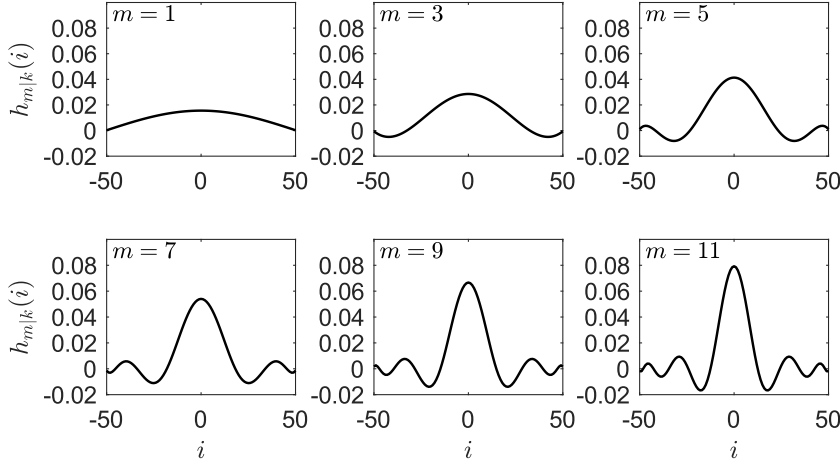
Fig. 2. Impulse responses $h_{m|k}(i)$ associated with fLBF estimators of different orders (polynomial basis, cosinusoidal window, $k = 50$).

and consequently

$$\bar{\boldsymbol{\theta}}_{m|k}^{\text{fLBF}}(t) = \boldsymbol{\theta}(t) \qquad (25)$$

which means that the fLBF estimator is unbiased.

The variance errors can be quantified as follows (see Appendix)

$$\text{cov}[\widehat{\boldsymbol{\theta}}_{m|k}^{\text{fLBF}}(t)] \cong \frac{\sigma_e^2 \boldsymbol{\Phi}_0^{-1}}{N_{m|k}} + \text{O}\left(\frac{1}{k}\right) \qquad (26)$$

where

$$N_{m|k} = \left[\sum_{i=-k}^{k} h_{m|k}^2(i) \right]^{-1} \qquad (27)$$

is the quantity which will be further referred to as equivalent estimation memory of the fLBF estimator.

3 Relation to the original local basis function approach

In the original (direct) LBF approach, proposed recently in (Niedźwiecki & Ciołek, 2019), the estimate of $\boldsymbol{\theta}(t)$ is obtained by minimizing the local sum of squared output modeling errors

$$\widehat{\boldsymbol{\theta}}_{m|k}(\cdot) = \arg\min_{\boldsymbol{\theta}(\cdot)} \sum_{i=-k}^{k} [y(t+i) - \boldsymbol{\varphi}^{\text{T}}(t+i)\boldsymbol{\theta}(t+i)]^2 \qquad (28)$$

under the constraints (15). The resulting point estimate has the form

$$\widehat{\boldsymbol{\theta}}_{m|k}^{\text{LBF}}(t) = [\mathbf{I}_n \otimes \mathbf{f}_{m|k}^{\text{T}}(0)]\mathbf{P}_{m|k}^{-1}(t)\mathbf{p}_{m|k}(t) \qquad (29)$$

where $\mathbf{P}_{m|k}(t) = \sum_{i=-k}^{k} w_k(i)\boldsymbol{\psi}_{m|k}(t,i)\boldsymbol{\psi}_{m|k}^{\text{T}}(t,i)$ and $\mathbf{p}_{m|k}(t) = \sum_{i=-k}^{k} w_k(i)y(t+i)\boldsymbol{\psi}_{m|k}(t,i)$ (provided that the matrix $\mathbf{P}_{m|k}(t)$ is nonsingular). The symbol $\otimes$ denotes Kronecker product of the respective matrices/vectors and $\boldsymbol{\psi}_{m|k}(t,i) = \boldsymbol{\varphi}(t+i) \otimes \mathbf{f}_{m|k}(i)$ denotes the generalized regression vector. The estimation formula (29) can be regarded as a generalization, to the system identification case, of the classical signal smoothing formula known as Savitzky-Golay filter (Schafer, 2011).

One can show – under assumptions (A1)–(A2) – that for growing k the regression matrix $\mathbf{P}_{m|k}(t)$ converges in the mean squared sense to a constant matrix $\bar{\mathbf{P}}_m = \text{E}[\mathbf{P}_{m|k}(t)] = \bar{\boldsymbol{\Phi}}_0 \otimes \mathbf{I}_m$ (Niedźwiecki, 1988b). This justifies the following approximation valid for sufficiently large values of k:

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}_{m|k}^{\text{LBF}}(t) &\cong [\mathbf{I}_n \otimes \mathbf{f}_{m|k}^{\text{T}}(0)]\bar{\mathbf{P}}_m^{-1}\mathbf{p}_{m|k}(t) \\
&= [\mathbf{I}_n \otimes \mathbf{f}_{m|k}^{\text{T}}(0)][\boldsymbol{\Phi}_0^{-1} \otimes \mathbf{I}_m] \times \\
&\quad \times \sum_{i=-k}^{k} w_k(i)y(t+i)[\boldsymbol{\varphi}(t+i) \otimes \mathbf{f}_{m|k}(i)] \\
&= \boldsymbol{\Phi}_0^{-1} \sum_{i=-k}^{k} h_{m|k}(i)y(t+i)\boldsymbol{\varphi}(t+i) \cong \widehat{\boldsymbol{\theta}}_{m|k}^{\text{fLBF}}(t) \quad (30)
\end{aligned}$$

The last transition in (30) stems from the identity $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{A}\mathbf{C} \otimes \mathbf{B}\mathbf{D}$ which holds for Kronecker products. In the light of this approximate equivalence relationship it is not surprising that the bias and variance formulas derived in Section 2 for fLBF estimators are approximately the same as those established in (Niedźwiecki & Ciołek, 2019) for LBF estimators.

Fig. 3 shows the results of postfiltering of the preestimated trajectories depicted in Fig. 1 ($m = 5$, $k = 100$, cosinusoidal window). Note that the fLBF estimates obtained in this way are hardly distinguishable from the LBF es-
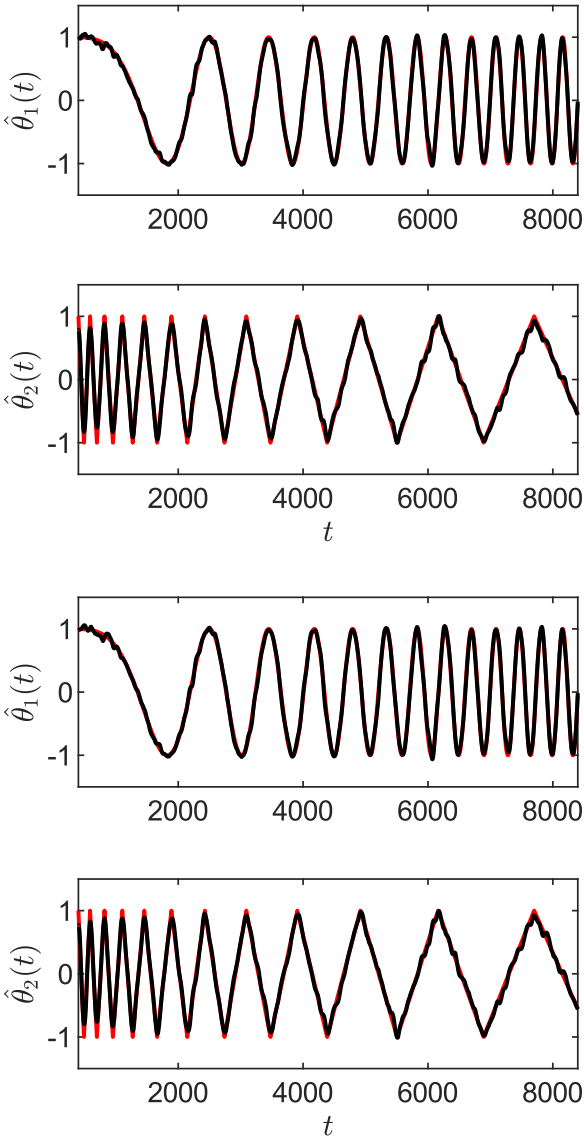
Fig. 3. Comparison of LBF estimates (two upper plots) and fLBF estimates (two lower plots) for $k = 100$ and $m = 5$. Estimated trajectories (black lines) are superimposed on true trajectories (red lines).

timates (more simulation results confirming this observation can be found in (Supplement, 2019)). This is quite remarkable considering the fact that LBF estimators are computationally much more demanding. Actually, computation of the LBF estimate (29) requires inversion, at each time instant t, of the $mn \times mn$-dimensional generalized regression matrix $\mathbf{P}_{m|k}(t)$. In contrast with this, the computational load of the fLBF estimator (21) is very low. First, the matrix $\mathbf{R}(t)$ in (2) has much smaller dimensions ($n \times n$) and can be inverted in a recursive way. Secondly, for the polynomial basis (17) and cosinusoidal window $w_k(i) = \cos \frac{\pi i}{2k}$, Hann window $w_k(i) = 0.5 \left[1 + \cos \frac{\pi i}{k} \right]$, or rectangular window, the fLBF estimates can be computed in a recursive way, cf. (Niedźwiecki & Ciołek, 2019).

Finally, we note that when the number of estimated coefficients n is large and the width of the local analysis interval $2k+1$ is small, the matrix $\mathbf{P}_{m|k}(t)$ is often poorly conditioned, which may result in some sort of bursting phenomena. Unlike the LBF algorithm, the fLBF algorithm is free of this negative effect, i.e., it is numerically robust – see (Supplement, 2019).

4 Adaptive selection of the number of basis functions and the window size

As shown in (Niedźwiecki & Ciołek, 2019), when the number of basis functions m is fixed and the window size k is increased, the variance component of the mean squared parameter estimation error (MSE) decreases and the bias component increases. Similarly, for a fixed window size, increasing the number of basis functions results in reduction of the bias component but, at the same time, in increase of the variance component. Since MSE is the sum of its bias and variance contributions, it is clear that to guarantee good tracking performance, the values of m and k should be chosen so as to trade-off both error components. Moreover, depending on the way system parameters change with time, the bias/variance compromise may require choosing different values of m and k in different time intervals. In this section we will propose the method that allows for adaptive scheduling of m and k.

4.1 Parallel estimation scheme

The proposed solution is based on parallel estimation. We will consider MK fLBF algorithms, equipped with different settings $m \in \mathcal{M} = \{m_1, \ldots, m_M\}$, $k \in \mathcal{K} = \{k_1, \ldots, k_K\}$. These algorithms are run simultaneously and at each time instant only one of the competing estimates is selected, i.e., the estimated parameter trajectory has the form

$$\widehat{\boldsymbol{\theta}}(t) = \widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}_{\widehat{m}(t)|\widehat{k}(t)}(t)$$

where

$$\{\widehat{m}(t), \widehat{k}(t)\} = \arg \min_{\substack{m \in \mathcal{M} \\ k \in \mathcal{K}}} J_{m|k}(t) \qquad (31)$$

and $J_{m|k}(t)$ denotes the local decision statistic. We will consider decision strategy which is based on the leave-one-out cross-validation technique (Allen, 1974), namely we will set

$$J_{m|k}(t) = \sum_{i=-L}^{L} [\varepsilon^{\circ}_{m|k}(t+i)]^2 \qquad (32)$$

where L decides upon the size of the local decision window of width $2L + 1$, and $\varepsilon^{\circ}_{m|k}(t)$ denotes the leave-one-out interpolation error of the system output (deleted residual)

$$\varepsilon^{\circ}_{m|k}(t) = y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)[\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}_{m|k}(t)]^{\circ} \qquad (33)$$

evaluated in terms of the holey fLBF estimator of $\boldsymbol{\theta}(t)$, i.e., the one that eliminates from the estimation process the measurement $y(t)$ collected at the instant t

$$[\widehat{\boldsymbol{\theta}}_{m|k}^{\mathrm{fLBF}}(t)]^{\circ} = [\widehat{\theta}_{1;m|k}^{\circ}(t), \ldots, \widehat{\theta}_{n;m|k}^{\circ}(t)]^{\mathrm{T}}$$
$$\widehat{\theta}_{j;m|k}^{\circ}(t) = \mathbf{f}_{m|k}^{\mathrm{T}}(0)\widehat{\boldsymbol{\beta}}_{j;m|k}^{\circ}(t), \ j = 1, \ldots, n$$

where

$$\widehat{\boldsymbol{\beta}}_{j;m|k}^{\circ}(t) = \arg\min_{\boldsymbol{\beta}_{j;m|k}} \sum_{\substack{i=-k \\ i\neq 0}}^{k} [\theta_j^*(t+i) - \mathbf{f}_{m|k}^{\mathrm{T}}(i)\boldsymbol{\beta}_{j;m|k}]^2$$

$$= [\mathbf{Q}_{m|k}^{\circ}]^{-1}\mathbf{q}_{j;m|k}^{\circ}(t) \qquad (34)$$

and (note that $w_k(0) = 1$)

$$\mathbf{Q}_{m|k}^{\circ} = \sum_{\substack{i=-k \\ i\neq 0}}^{k} w_k(i)\mathbf{f}_{m|k}(i)\mathbf{f}_{m|k}^{\mathrm{T}}(i) = \mathbf{I}_m - \mathbf{f}_{m|k}(0)\mathbf{f}_{m|k}^{\mathrm{T}}(0)$$

$$\mathbf{q}_{j;m|k}^{\circ}(t) = \sum_{\substack{i=-k \\ i\neq 0}}^{k} w_k(i)\mathbf{f}_{m|k}(i)\theta_j^*(t+i)$$

$$= \widehat{\boldsymbol{\beta}}_{j;m|k}(t) - \mathbf{f}_{m|k}(0)\theta_j^*(t).$$

Using the matrix inversion lemma (Söderström & Stoica, 1988), one obtains

$$[\widehat{\boldsymbol{\theta}}_{m|k}^{\mathrm{fLBF}}(t)]^{\circ} = \mathbf{f}_{m|k}^{\mathrm{T}}(0) \left[\mathbf{I}_m + \frac{\mathbf{f}_{m|k}(0)\mathbf{f}_{m|k}^{\mathrm{T}}(0)}{1 - \mathbf{f}_{m|k}^{\mathrm{T}}(0)\mathbf{f}_{m|k}(0)} \right] \times$$
$$\times [\widehat{\boldsymbol{\beta}}_{j;m|k}(t) - \mathbf{f}_{m|k}(0)\theta_j^*(t)]$$

which, after straightforward calculations, leads to

$$[\widehat{\boldsymbol{\theta}}_{m|k}^{\mathrm{fLBF}}(t)]^{\circ} = \frac{1}{1 - c_{m|k}} \left[\widehat{\boldsymbol{\theta}}_{m|k}(t) - c_{m|k}\boldsymbol{\theta}^*(t) \right]$$
$$\varepsilon_{m|k}^{\circ}(t) = \frac{1}{1 - c_{m|k}} [\varepsilon_{m|k}(t) - c_{m|k}\varepsilon^*(t)] \qquad (35)$$

where $\varepsilon_{m|k}(t) = y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}_{m|k}(t)$, $\varepsilon^*(t) = y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}^*(t)$ and $c_{m|k} = \mathbf{f}_{m|k}^{\mathrm{T}}(0)\mathbf{f}_{m|k}(0)$. According to (35), the cross-validation statistic can be evaluated without actually implementing the holey estimation scheme. It should be stressed that the cross-validation based decision rule described above differs from the one derived in (Niedźwiecki & Ciołek, 2019) for LBF estimators.

Note that while denoising of preestimates is carried out using classical *signal processing* tools, selection of the best smoothing variant relies on comparison of local system modeling errors (33), i.e., it is based on *system identification* inference.

4.2 Selection of design parameters

The problem of selection of the analysis window sizes $k_1, \ldots, k_K$ for a bank of competing algorithms was stud-

Table 1
Parameter settings corresponding to three speeds of parameter variation (SoV)

SoV	slow	medium	fast
T_s	40000	20000	10000
$\omega_{1|T_s}$	0.015	0.03	0.06
$\omega_{2|T_s}$	0.02	0.04	0.08

ied analytically in (Niedźwiecki et al., 2017b). As shown there, to maximize robustness of the parallel estimation scheme, the adopted values of k should form a geometric progression: $k_i = \alpha k_{i-1}, \ i = 2, \ldots, \alpha > 1$. The recommended values of α range between 1.57 (for smooth parameter trajectories) and 2.43 (for random walk type trajectories). If nothing is known *a priori* about the way signal parameters change, one can adopt a compromise value $\alpha = 2$.

If needed, the order n of the FIR model can be selected adaptively at the preestimation stage using the modified (localized) version of the Akaike's final prediction error (FPE) criterion (Niedźwiecki & Ciołek, 2017a).

5 Simulation results

In our simulation experiment the two-tap FIR system (14) was analyzed, with parameters modeled as sinusoidal chirps with linearly increasing instantaneous frequencies (see Fig. 4). The shape of parameter trajectories within the simulation interval $[1, T_s]$ was fixed, i.e., different variants of discrete-time trajectories were generated by "sampling" the prototype analog trajectories at different sampling rates:

$$\theta_i(t) = 0.5 \cos[\phi_i(t)]$$
$$\phi_i(t) = \sum_{s=1}^{t} \omega_i(s), \quad \omega_i(s) = \frac{s}{T_s}\omega_{i|T_s} \qquad (36)$$
$$i = 1, 2, \quad t = 1, \ldots, T_s$$

– all further details are summarized in Table 1. To test identification algorithms under different speeds of parameter variation (SoV), three values of T_s were considered: $T_s = 40000$ (slow variations), $T_s = 20000$ (medium-speed variations) and $T_s = 10000$ (fast variations). Additionally, two average signal-to-noise ratios were applied: 15 dB ($\sigma_e^2 = 1.25 \cdot 10^{-2}$) and 25 dB ($\sigma_e^2 = 1.25 \cdot 10^{-3}$).

Table 2 shows the mean squared parameter estimation errors obtained for 9 LBF/fLBF estimators corresponding to different choices of design parameters k (50, 100, 200, $\alpha = 2$) and m (1, 3, 5), and for the adaptive parallel estimation schemes with model selection based on cross-validation. The averages were computed for 100 process realizations, 3 speeds of parameter variation (SoV) and 2 average signal-to-noise ratios. The value of L was set to 30 and the preestimation parametr λ_0 was set to 0.9. Since $k_{\max} + L + n = 232$, evaluation of the compared
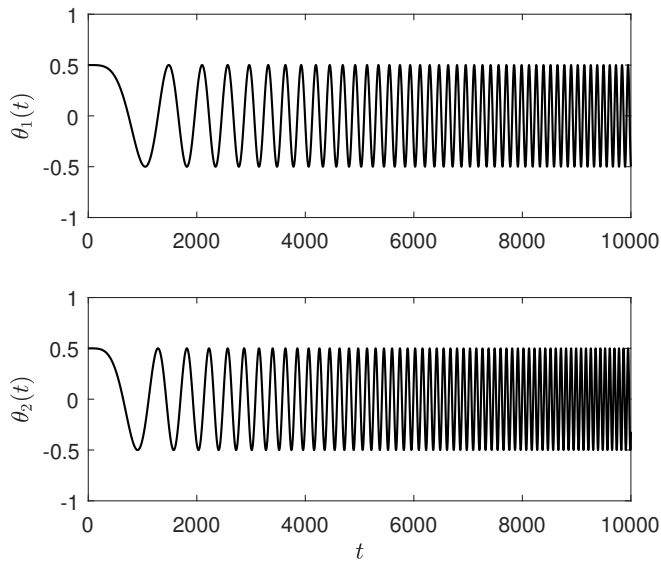
Fig. 4. True parameter trajectories.

algorithms was started at the instant $t = 233$ and ended at the instant $t = T_s - 233$ so that, irrespective of the value of k, all competitors could be checked on the same data set.

A careful examination of the results shown in Table 2 leads to the following conclusions

(1) For all choices of m, k and SoV, the fLBF algorithms yield results that only marginally differ (either up or down) from those provided by their, computationally much more demanding, LBF "prototypes".
(2) In all cases considered, the parallel estimation schemes yield results that are better or only slightly worse than those provided by the best LBF/fLBF algorithms with fixed settings.

Simulation experiments were next repeated for a different pattern of window sizes (60, 90, 135, 200, $\alpha = 1.5$), different sets of basis functions (harmonic, Slepian), and different shapes of the weighting sequence (rectangular, Hann). The obtained results, not shown here because of the lack of space, were very similar to those summarized in Table 2, which confirms that all design choices mentioned above are by no means critical, i.e., the proposed parallel estimation scheme is robust.

6   The bigger picture - implicit filtering versus explicit filtering

On the qualitative level, the current paper extends and further explores intriguing dualities between identification of nonstationary stochastic systems and signal processing, originally pointed out in (Niedźwiecki, 1990) and (Niedźwiecki & Kłaput, 2002). We have shown that implicit filtering, imposed by the LBF scheme, is to some extent interchangeable with explicit filtering (postfiltering)

of parameter preestimates. This observation has several interesting implications. First of all, it allows one to design fast versions of LBF algorithms, with the same parameter tracking capabilities but significantly lower computational complexity and higher numerical robustness. Second, it can be easily extended to multivariate systems. Third, it offers increased estimation flexibility, since optimization of the fLBF algorithm (the choice of design parameters m and k) can be carried out independently for each system coefficient. Last but not least, the proposed approach opens up several new opportunities for parameter tracking, such as wavelet-based denoising of preestimated trajectories, postfiltering using a bank of arbitrary lowpass filters, nonlinear (e.g. median) postfiltering, or postfiltering using various data-adaptive algorithms developed for the purpose of signal processing (Orfanidis, 2010), (Fitzgerald et al., 2000).

7   Conclusion

A new estimation paradigm for identification of linear time-varying FIR systems, based on the concepts of preestimation and postfiltering, was proposed. The resulting two-stage identification algorithm compares favorably with the existing solutions to the problem of parameter tracking as it offers – without compromising good tracking performance – greater flexibility, higher numerical robustness and significantly lower computational complexity than the classical estimation schemes.

References

Allen, D.M. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometrics*, (16), 125–127.

Bellegarda, J.R. & Farden, D.C. (1988). Constrained time-varying system modelling. In *Proc. 1988 American Contol Conference*, Atlanta, USA, 1295–1300.

Fitzgerald, W.J., Smith, R.L., Walden, A.T. & Young, P.C. Eds. (2000). *Nonlinear and Nonstationary Signal Processing*. Cambridge University Press.

Galrinho, M., Rojas, C. R. & Hjalmarsson, H. (2014). A weighted least-squares method for parameter estimation in structured models. In Proc. *53rd IEEE Conference on Decision and Control*, Los Angeles, USA, 3322-3327.

Grenier, Y. (1981). Time-dependent ARMA modeling of nonstationary signals. *IEEE Trans. Acoust. Speech Signal Process.*, (31), 899–911.

Hall, M., Oppenheim A. V. & Willsky A. (1983). Time-varying parametric modeling of speech. *Signal Processing*, (5), 267–285.

Isserilis L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12, 134-139.

Kitagawa, G. & Gersch, W. (1985). A smoothness priors time-varying AR coefficient modeling of nonstationary covariance time series. *IEEE Trans. Autom. Control*, (30), 48–56.

Liporace, J. M. (1975). Linear estimation of nonstationary signals. *J. Acoust. Soc. Amer.*, (58), 1288–1295.

Mendel, J. M. (1973). *Discrete Techniques of Parameter Estimation: The Equation Error Formulation*. New York: Marcel Dekker, 1973.

Table 2

Mean squared parameter estimation errors (FIR system) obtained for 9 LBF/fLBF estimators corresponding to different choices of design parameters k (50, 100, 200) and m (1, 3, 5), and for the adaptive parallel estimation schemes (A) based on cross-validation. The averages were computed for 100 process realizations, 3 speeds of parameter variation (SoV) and 2 average signal-to-noise ratios (SNR). The best results in each group are shown in boldface. MATLAB codes and input data allowing one to reproduce simulation results are available from (Supplement, 2019).

LBF

SNR	SoV	slow			medium			fast		
	$k\backslash m$	1	3	5	1	3	5	1	3	5
15 dB	50	1.47E-03	1.37E-03	2.15E-03	6.34E-03	1.39E-03	2.14E-03	4.00E-02	3.13E-03	**2.18E-03**
	100	5.25E-03	6.78E-04	1.01E-03	4.00E-02	2.42E-03	**1.03E-03**	1.41E-01	5.09E-02	8.99E-03
	200	3.96E-02	2.06E-03	**5.11E-04**	1.41E-01	5.44E-02	9.87E-03	2.00E-01	1.62E-01	1.19E-01
	A	9.06E-04			1.15E-03			1.67E-03		
25 dB	50	9.04E-04	1.37E-04	2.14E-04	5.81E-03	1.66E-04	2.15E-04	3.92E-02	1.86E-03	**2.40E-04**
	100	4.98E-03	8.80E-05	1.02E-04	3.96E-02	1.84E-03	**1.25E-04**	1.40E-01	5.03E-02	8.01E-03
	200	3.94E-02	1.77E-03	**7.21E-05**	1.41E-01	5.40E-02	9.48E-03	1.99E-01	1.61E-01	1.19E-01
	A	9.96E-05			1.36E-04			2.00E-04		

fLBF

SNR	SoV	slow			medium			fast		
	$k\backslash m$	1	3	5	1	3	5	1	3	5
15 dB	50	9.91E-04	1.47E-03	2.23E-03	4.54E-03	1.77E-03	2.72E-03	3.46E-02	**4.23E-03**	4.49E-03
	100	4.15E-03	7.42E-04	1.10E-03	3.63E-02	2.31E-03	**1.33E-03**	1.34E-01	4.73E-02	9.25E-03
	200	3.75E-02	1.90E-03	**5.72E-04**	1.37E-01	5.17E-02	9.35E-03	1.96E-01	1.55E-01	1.13E-01
	A	9.37E-04			1.35E-03			3.46E-03		
25 dB	50	3.92E-04	2.34E-04	3.80E-04	3.95E-03	5.57E-04	9.02E-04	3.39E-02	2.94E-03	**2.59E-03**
	100	3.85E-03	1.28E-04	1.70E-04	3.59E-02	1.71E-03	**4.12E-04**	1.34E-01	4.67E-02	8.26E-03
	200	3.73E-02	1.59E-03	**1.10E-04**	1.37E-01	5.13E-02	8.90E-03	1.96E-01	1.55E-01	1.13E-01
	A	1.28E-04			4.00E-04			2.14E-03		

'

Mrad, R. B., Fassois S. D. & Levitt J. A. (1998). A polynomial-algebraic method for non-stationary TARMA signal analysis Part I: The method. *Signal Processing*, (65), 1–19.

Niedźwiecki, M. (1988a). First order tracking properties of weighted least squares estimators. *IEEE Trans. Automat. Contr.*, (33), 94–96.

Niedźwiecki, M. (1988b). Functional series modeling approach to identification of nonstationary stochastic systems. *IEEE Trans. Automat. Contr.*, (33), 955–961.

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

Niedźwiecki, M. & Kłaput T. (2002). Fast recursive basis function estimators for identification of nonstationary systems. *IEEE Transactions on Signal Processing*, (50), 1925–1934.

Niedźwiecki M. (2012). Locally adaptive cooperative Kalman smoothing and its application to identification of nonstationary stochastic systems. *IEEE Trans. Signal Process.*, (60), 48–59.

Niedźwiecki, M. & Ciołek, M. (2017a) Akaike's final prediction error criterion revisited. In Proc. *40th International Conference on Telecommunications and Signal Processing*, Barcelona, Spain, 237-242.

Niedźwiecki M., Ciołek M. & Kajikawa Y. (2017b). On adaptive covariance and spectrum estimation of locally stationary multivariate processes. *Automatica*, (82), 1–12.

Niedźwiecki, M. & Ciołek, M. (2019). Generalized Savitzky-Golay filters for identification of nonstationary systems. *Automatica*, (108), 108477, 1–8.

Norton, J. P. (1975). Optimal smoothing in the identification of linear time-varying systems. *Proc. IEE*, (122), 663–668.

Orfanidis, S. J. (2010). *Introduction to Signal Processing.* Rutgers ECE.

Poulimenos, A. G. & Fassois S. D. (2006). Parametric time-domain methods for non-stationary random vibration modelling and analysis - a critical survey and comparison. *Mechanical Systems and Signal Processing*, (20), 763–816.

Rao, T. Subba (1970). The fitting of nonstationary time-series models with time-dependent parameters. *J. R. Statist. Soc. B*, (32), 312–322.

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

Schafer R. W. (2011). What is a Savitzky-Golay filter?. *IEEE Signal Process. Mag.*, (28), 111–117.

Supplement: supplementary material (including MATLAB codes) [Online]. Available at: https://eti.pg.edu.pl/katedra-systemow-automatyki/Automatica

Young, P.C. (1984). *Recursive Estimation and Time Series Analysis.* New York: Springer Verlag.

Zou, R. B., Wang H. & Chon K. H. (2003). A robust time-varying identification algorithm using basis functions. *Annals of Biomedical Engineering*, (31), 840–853.

Appendix [derivation of (26)]

Combining (1), (5), (21) and (24), one arrives at

$$\Delta\widehat{\boldsymbol{\theta}}_{m|k}^{\text{fLBF}}(t) = \widehat{\boldsymbol{\theta}}_{m|k}^{\text{fLBF}}(t) - \bar{\boldsymbol{\theta}}_{m|k}^{\text{fLBF}}(t) = \boldsymbol{\Delta}_1(t) + \boldsymbol{\Delta}_2(t)$$

where

$$\boldsymbol{\Delta}_1(t) = \sum_{i=-k}^{k} h_{m|k}(i)\boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t+i)e(t+i)$$

$$\boldsymbol{\Delta}_2(t) =$$
$$= \sum_{i=-k}^{k} h_{m|k}(i)[\boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t+i)\boldsymbol{\varphi}^{\text{T}}(t+i) - \mathbf{I}_n]\boldsymbol{\theta}(t+i).$$

Since $\{e(t)\}$ is zero-mean and independent of $\{\boldsymbol{\varphi}(t)\}$, it holds that $\mathrm{E}[\boldsymbol{\Delta}_1(t)\boldsymbol{\Delta}_2^{\text{T}}(t)] = \mathrm{E}[\boldsymbol{\Delta}_2(t)\boldsymbol{\Delta}_1^{\text{T}}(t)] = 0$ and

$$\mathrm{cov}[\widehat{\boldsymbol{\theta}}_{m|k}^{\text{fLBF}}(t)] = \mathrm{E}\{\Delta\widehat{\boldsymbol{\theta}}_{m|k}^{\text{fLBF}}(t)[\Delta\widehat{\boldsymbol{\theta}}_{m|k}^{\text{fLBF}}(t)]^{\text{T}}\}$$
$$= \mathrm{E}[\boldsymbol{\Delta}_1(t)\boldsymbol{\Delta}_1^{\text{T}}(t)] + \mathrm{E}[\boldsymbol{\Delta}_2(t)\boldsymbol{\Delta}_2^{\text{T}}(t)]$$

Note that

$$\mathrm{E}[\boldsymbol{\Delta}_1(t)\boldsymbol{\Delta}_1^{\text{T}}(t)] = \mathrm{E}\Big[\sum_{i=-k}^{k}\sum_{j=-k}^{k} h_{m|k}(i)h_{m|k}(j)\times$$
$$\times\, e(t+i)e(t+j)\boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t+i)\boldsymbol{\varphi}^{\text{T}}(t+j)\boldsymbol{\Phi}_0^{-1}\Big]$$
$$= \sum_{i=-k}^{k} h_{m|k}^2(i)\mathrm{E}[e^2(t+i)]\boldsymbol{\Phi}_0^{-1}\mathrm{E}[\boldsymbol{\varphi}(t+i)\boldsymbol{\varphi}^{\text{T}}(t+i)]\boldsymbol{\Phi}_0^{-1}$$
$$= \frac{\sigma_e^2\boldsymbol{\Phi}_0^{-1}}{N_{m|k}}\,.$$

The second component of the covariance matrix has the form

$$\mathrm{E}[\boldsymbol{\Delta}_2(t)\boldsymbol{\Delta}_2^{\text{T}}(t)] = \sum_{i=-k}^{k}\sum_{j=-k}^{k} h_{m|k}(i)h_{m|k}(j)\mathrm{E}[\boldsymbol{\Sigma}(i,j)] \geq 0$$

where

$$\boldsymbol{\Sigma}(i,j) = \boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t+i)\boldsymbol{\varphi}^{\text{T}}(t+i)\boldsymbol{\theta}(t+i)\boldsymbol{\theta}^{\text{T}}(t+j)\times$$
$$\times\, \boldsymbol{\varphi}(t+j)\boldsymbol{\varphi}^{\text{T}}(t+j)\boldsymbol{\Phi}_0^{-1}$$
$$- \boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t+i)\boldsymbol{\varphi}^{\text{T}}(t+i)\boldsymbol{\theta}(t+i)\boldsymbol{\theta}^{\text{T}}(t+j)$$
$$- \boldsymbol{\theta}(t+i)\boldsymbol{\theta}^{\text{T}}(t+j)\boldsymbol{\varphi}(t+j)\boldsymbol{\varphi}^{\text{T}}(t+j)\boldsymbol{\Phi}_0^{-1}$$
$$+ \boldsymbol{\theta}(t+i)\boldsymbol{\theta}^{\text{T}}(t+j).$$

Using the well-known properties of higher order moments of Gaussian variables (Isserilis, 1918), one can easily derive the following formula which holds for zero-mean jointly normally distributed $n \times 1$ random vectors $\mathbf{x}$, $\mathbf{y}$, $\mathbf{w}$, $\mathbf{v}$ and a $n \times n$ matrix $\mathbf{A}$

$$\mathrm{E}[\mathbf{x}\mathbf{y}^{\text{T}}\mathbf{A}\mathbf{w}\mathbf{v}^{\text{T}}] = \mathrm{E}[\mathbf{x}\mathbf{y}^{\text{T}}]\mathbf{A}\mathrm{E}[\mathbf{w}\mathbf{v}^{\text{T}}] + \mathrm{E}[\mathbf{x}\mathbf{w}^{\text{T}}]\mathbf{A}^{\text{T}}\mathrm{E}[\mathbf{y}\mathbf{v}^{\text{T}}]$$
$$+ \mathrm{E}[\mathbf{x}\mathbf{v}^{\text{T}}]\mathrm{E}[\mathbf{y}^{\text{T}}\mathbf{A}\mathbf{w}].$$

Using this formula, after elementary calculations, one obtains

$$\mathrm{E}[\boldsymbol{\Sigma}(i,j)] = \boldsymbol{\Phi}_0^{-1}\boldsymbol{\Phi}_{i-j}\boldsymbol{\theta}(t+j)\boldsymbol{\theta}^{\text{T}}(t+i)\boldsymbol{\Phi}_{i-j}\boldsymbol{\Phi}_0^{-1}$$
$$+ \boldsymbol{\Phi}_0^{-1}\boldsymbol{\Phi}_{i-j}\boldsymbol{\theta}^{\text{T}}(t+i)\boldsymbol{\Phi}_{i-j}\boldsymbol{\theta}(t+j)\boldsymbol{\Phi}_0^{-1}$$

where $\boldsymbol{\Phi}_{i-j} = \mathrm{E}[\boldsymbol{\varphi}(t+i)\boldsymbol{\varphi}^{\text{T}}(t+j)] = \boldsymbol{\Phi}_{j-i}^{\text{T}}$.

One can show that

$$\mathbf{f}_{m|k}(i) \;\xrightarrow[k\to\infty]{}\; \frac{1}{\sqrt{k}}\,\mathbf{f}_m^0\left(\frac{i}{k}\right)$$

where $\mathbf{f}_m^0(s) = [f_1^0(s),\ldots,f_m^0(s)]^{\text{T}}$ is the vector of w-orthonormal continuous-time basis generating functions obeying $\int_{-1}^{1} w^0(s)\mathbf{f}_m^0(s)[\mathbf{f}_m^0(s)]^{\text{T}}ds = \mathbf{I}_m$. Actually, note that after replacing $\mathbf{f}_{m|k}(i)$ and $w_k(i)$ with $(1/\sqrt{k})\mathbf{f}_m^0(i/k)$ and $w^0(i/k)$, respectively, one obtains

$$\frac{1}{k}\sum_{i=-k}^{k} w^0\left(\frac{i}{k}\right)\mathbf{f}_m^0\left(\frac{i}{k}\right)\left[\mathbf{f}_m^0\left(\frac{i}{k}\right)\right]^{\text{T}}$$
$$\cong \frac{1}{k}\int_{-k}^{k} w^0\left(\frac{\tau}{k}\right)\mathbf{f}_m^0\left(\frac{\tau}{k}\right)\left[\mathbf{f}_m^0\left(\frac{\tau}{k}\right)\right]^{\text{T}} d\tau$$
$$= \int_{-1}^{1} w^0(s)\mathbf{f}_m^0(s)[\mathbf{f}_m^0(s)]^{\text{T}}ds = \mathbf{I}_m\,.$$

This means that $|f_{j|k}(t)| = \mathrm{O}(1/\sqrt{k})$, $\forall t, 1 \leq j \leq m$, and hence $h_{m|k}(i) = \mathrm{O}(1/k), \forall i$. Denote by δ any element of the matrix $\mathrm{E}[\boldsymbol{\Delta}_2(t)\boldsymbol{\Delta}_2^{\text{T}}(t)]$. When parameter trajectory is uniformly bounded, it holds that

$$|\delta| \leq \frac{c}{k^2}\sum_{i=-k}^{k}\sum_{j=-k}^{k}\gamma^{2|i-j|} = \mathrm{O}(1/k)$$

where c is a constant that does not depend on k and t, and γ is the rate of decay of the autocorrelation function of $u(t)$ postulated in (A1). This completes derivation of (26).

Maciej Niedźwiecki received the M.Sc. and Ph.D. degrees from the Technical University of Gdańsk, 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). His main areas of research interests include system identification, statistical signal processing and adaptive systems.

Dr. Niedźwiecki is currently a member of the IFAC committees on Modeling, Identification and Signal Processing and on Large Scale Complex Systems, and a member of the Automatic Control and Robotics Committee of the Polish Academy of Sciences (PAN). He works as a Professor and Head of the Department of Automatic Control, Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology.

Marcin Ciołek received the M.Sc. and Ph.D. degrees from the Gdańsk University of Technology (GUT), Gdańsk, Poland, in 2010 and 2017, respectively. Since 2017, he has been working as an Adjunct Professor in the Department of Automatic Control, Faculty of Electronics, Telecommunications and Informatics, GUT. His professional interests include speech, music and biomedical signal processing.

Artur Gańcza received the M.Sc. degree from the Gdańsk University of Technology (GUT), Gdańsk, Poland, in 2019. He is currently a Ph.D. student at GUT, with the Department of Automatic Control, Faculty of Electronics, Telecommunications and Informatics. His professional interests include speech recognition, system identification and adaptive signal processing.