2019 IEEE 58th Conference on Decision and Control (CDC)
Palais des Congrès et des Expositions Nice Acropolis
Nice, France, December 11-13, 2019

Local basis function estimators for identification of nonstationary systems

Maciej Niedźwiecki, Marcin Ciołek and Artur Gańcza

***Abstract*— The problem of identification of a nonstationary stochastic system is considered and solved using local basis function approximation of system parameter trajectories. Unlike the classical basis function approach, which yields parameter estimates in the entire analysis interval, the proposed new identification procedure is operated in a sliding window mode and provides a sequence of point (rather than interval) estimates. It is shown that for the polynomial basis all computations can be carried out recursively and that two important design parameters – the number of basis functions and the size of the local analysis window – can be chosen in an adaptive way.**

I. Introduction

Identification of nonstationary stochastic systems can be carried out using different approaches, based on different methodologies. First, assuming that system parameters vary slowly with time, one can estimate them using time-localized versions of classical identification algorithms such as weighted least squares [1]–[3]. Secondly, one can approximate parameter trajectories by linear combinations of known functions of time, called basis functions (BF). In this way the problem of estimation of time-varying system parameters can be converted to a problem of estimation of time-invariant coefficients appearing in basis function based trajectory approximations [4] - [14]. Finally, one can adopt a stochastic model of parameter variation and reformulate the identification problem as a problem of estimation of a state of an appropriately defined linear dynamic system. In this case parameter estimation can be carried out using Kalman filtering/smoothing algorithms – see e.g. [15]–[17].

In this paper we combine the local estimation framework with the basis function approach. This means that, unlike [9], the BF approach is used to generate a sequence of point (rather than interval) estimates of system parameters, i.e., the estimation is carried out in the sliding window mode. The resulting local basis function (LBF) estimation algorithm, which for the polynomial basis can be put down in a recursive form, can be regarded as a generalization, to the system identification case, of the classical signal smoothing algorithm known as Savitzky-Golay filter [18], [19]. The current contribution is focused on one of the key aspects of the local BF estimation that has not been satisfactorily addressed yet: joint adaptive selection of the number of basis functions and the width of the analysis interval.

*This work was partially supported by the National Science Center under the agreement UMO-2018/29/B/ST7/00325. Computer simulations were carried out at the Academic Computer Centre in Gdańsk. All authors are with the Gdańsk University of Technology, Faculty of Electronics, Telecommunications and Informatics, Department of Automatic Control, ul. Narutowicza 11/12, Gdańsk, Poland: maciekn@eti.pg.edu.pl, marciole@pg.edu.pl, artgancz@student.pg.edu.pl

II. Problem statement

Consider a nonstationary stochastic system governed by

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

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 $n \times 1$ vector of time-varying system parameters, $\boldsymbol{\varphi}(t) = [\varphi^1(t), \ldots, \varphi^n(t)]^{\mathrm{T}}$ denotes the $n \times 1$ vector of regression variables (the contents of the regression will be specified later on), and $e(t)$ denotes white measurement noise.

In agreement with the local BF strategy, at each time instant t we will assume that in the local analysis interval $T_k(t) = [t-k, t+k]$ of width $2k+1$, centered at t, system parameters can be expressed as linear combinations of a certain number of known functions of time $f_{l|k}(i), i \in I_k = [-k, k]$ (further referred to as basis functions), namely

$$\theta^j(t+i) = \sum_{l=1}^{m} a^j_{l,m|k} f_{l|k}(i), \quad j = 1, \ldots, n, \quad i \in I_k. \tag{2}$$

The subscripts m (number of basis functions) and k (the size of the analysis interval) are explicitly shown in (2) since later both quantities will be a subject of adaptive scheduling.

We will assume that discrete-time basis functions $f_{l|k}(i)$ are obtained by "sampling" linearly independent continuous-time prototype functions $f^0_l(s), s \in [-1, 1]$, namely

$$f_{l|k}(i) = f^0_l\left(\frac{i}{k}\right), \quad l = 1, \ldots, m, \quad i \in I_k. \tag{3}$$

The subspace of the space of all square summable sequences defined on I_k spanned by the basis functions (3) will be further denoted by $\mathcal{F}_{m|k}$.

Denote by

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

the $m \times 1$ vector of basis functions. Finally, denote by

$$\boldsymbol{\psi}_{m|k}(t,i) = \boldsymbol{\varphi}(t+i) \otimes \mathbf{f}_{m|k}(i)$$

where $\otimes$ denotes Kronecker product of the corresponding vectors/matrices, the $nm \times 1$ generalized regression vector, and by

$$\boldsymbol{\alpha}_{m|k} = [a^1_{1,m|k}, \ldots, a^1_{m,m|k}, \ldots, a^n_{1,m|k}, \ldots, a^n_{m,m|k}]^{\mathrm{T}}$$

the $nm \times 1$ vector of coefficients describing evolution of system parameters in the interval $T_k(t)$. Using these shorthands, the adopted local model can be summarized as follows

$$\begin{gathered} y(t+i) = \boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,i)\boldsymbol{\alpha}_{m|k} + e(t+i) \\ \mathrm{E}[e^2(t+i)] = \rho, \quad i \in I_k. \end{gathered} \tag{4}$$

978-1-7281-1397-5/19/$31.00 ©2019 IEEE

III. Identification procedure

Identification of the model (4) can be carried out using the method of least squares, leading to the following formulae

$$\widehat{\boldsymbol{\alpha}}_{m|k}(t) = \arg\min_{\boldsymbol{\alpha}_{m|k}} \sum_{i=-k}^{k} [y(t+i) - \boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,i)\boldsymbol{\alpha}_{m|k}]^2 = \mathbf{P}^{-1}_{m|k}(t)\mathbf{p}_{m|k}(t) \quad (5)$$

where the $mn \times mn$ generalized regression matrix $\mathbf{P}_{m|k}(t)$ [assumed to be nonsingular] and the $mn \times 1$ vector $\mathbf{p}_{m|k}(t)$ are given by

$$\mathbf{P}_{m|k}(t) = \sum_{i=-k}^{k} \boldsymbol{\psi}_{m|k}(t,i)\boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,i)$$
$$\mathbf{p}_{m|k}(t) = \sum_{i=-k}^{k} y(t+i)\boldsymbol{\psi}_{m|k}(t,i). \quad (6)$$

The local estimate of ρ can be obtained from

$$\widehat{\rho}_{m|k}(t) = \frac{1}{2k+1} \sum_{i=-k}^{k} [y(t+i) - \boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,i)\widehat{\boldsymbol{\alpha}}_{m|k}(t)]^2 = \frac{1}{2k+1}[c_k(t) - \mathbf{p}^{\mathrm{T}}_{m|k}(t)\widehat{\boldsymbol{\alpha}}_{m|k}(t)] \quad (7)$$

where $c_k(t) = \sum_{i=-k}^{k} y^2(t+i)$.

Based on (5), the estimates of system parameters can be evaluated using the formula

$$\widehat{\theta}^j(t) = \sum_{l=1}^{m} \widehat{a}^j_{l,m|k} f_{l|k}(0), \quad j = 1, \ldots, n$$

which can be rewritten in a more compact form as

$$\widehat{\boldsymbol{\theta}}_{m|k}(t) = \mathbf{F}_{m|k}\widehat{\boldsymbol{\alpha}}_{m|k}(t)$$
$$\mathbf{F}_{m|k} = \mathbf{I}_n \otimes \mathbf{f}^{\mathrm{T}}_{m|k}(0). \quad (8)$$

Note that even though the estimate $\widehat{\boldsymbol{\alpha}}_{m|k}(t)$ allows one to approximate parameter trajectory in the entire analysis interval $[t-k, t+k]$, it is used to generate a point estimate (8), valid only at the instant t, i.e., the proposed estimation scheme is operated in a sliding window mode. Note also that the estimator (8) is noncausal as it relies on both "past" – with respect to t – data samples $\{y(t-i), \boldsymbol{\varphi}(t-i), i = 1, \ldots, k\}$ and "future" samples $\{y(t+i), \boldsymbol{\varphi}(t+i), i = 1, \ldots, k\}$. Even though noncausal estimators of this form cannot be used in real time applications, such as adaptive prediction or adaptive control, they are admissible in almost real time applications, such as channel equalization [20] or parametric spectrum estimation [21], where the model-based decisions can be postponed by k sampling intervals.

IV. Basic properties of LBF estimators

To obtain analytical results, we will assume that (1) is a finite impulse response system, i.e.,

$$\boldsymbol{\varphi}(t) = [u(t-1), \ldots, u(t-n)]^{\mathrm{T}}$$

where $u(t)$ denotes an observable input sequence. Furthermore, following [9], we will assume that

(A1) $\{u(t)\}$ is a zero-mean wide sense stationary Gaussian sequence, persistently exciting of order at least n, with an exponentially decaying autocorrelation function $r_u(i) = \mathrm{E}[u(t)u(t-i)]$:

$$\exists\ 0 < c_1 < \infty, 0 < \beta < 1 \ :\ \ |r_u(i)| \le c_1\beta^{|i|}, \quad \forall i$$

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

Denote by $\{\widetilde{f}_{1|k}(i), \ldots, \widetilde{f}_{m|k}(i), i \in I_k\}$ the orthonormal basis set of $\mathcal{F}_{m|k}$, obeying the condition

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

where

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

Such a set of orthonormal functions can be determined by means of applying the Gram-Schmidt procedure to $\{f_{1|k}(i), \ldots, f_{m|k}(i), i \in I_k\}$.

Let $\Omega_k(t) = \{\boldsymbol{\varphi}(t+i), e(t+i)\}$ and $\boldsymbol{\Phi} = \mathrm{E}[\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)] > 0$. Denote by

$$\bar{\boldsymbol{\theta}}_{m|k}(t) = \mathrm{E}_{\Omega_k(t)}[\widehat{\boldsymbol{\theta}}_{m|k}(t)]$$

the mean path of parameter estimates. According to [9], if true system parameters obey the model (4), i.e., if all parameter trajectories belong to the subspace $\mathcal{F}_{m|k}$, the estimator $\widehat{\boldsymbol{\theta}}_{m|k}(t)$ is – under (A1) and (A2) – unbiased

$$\bar{\boldsymbol{\theta}}_{m|k}(t) = \boldsymbol{\theta}(t) \quad (9)$$

and

$$\mathrm{cov}[\widehat{\boldsymbol{\theta}}_{m|k}(t)] \cong \frac{\rho\boldsymbol{\Phi}^{-1}}{N_{m|k}} \quad (10)$$

where

$$N_{m|k} = \left[\sum_{i=-k}^{k} h^2_{m|k}(i)\right]^{-1} \quad (11)$$

denotes the equivalent number of observations [2] and

$$h_{m|k}(i) = \widetilde{\mathbf{f}}^{\mathrm{T}}_{m|k}(0)\widetilde{\mathbf{f}}_{m|k}(i), \quad i \in I_k. \quad (12)$$

In a more realistic case, where the only assumption made about system parameter variation is that

(A3) $\{\boldsymbol{\theta}(t)\}$ is independent of $\{u(t)\}$ and $\{e(t)\}$

which does not imply that system parameters can be exactly modeled as linear combinations of basis functions, it can be shown that [9]

$$\bar{\boldsymbol{\theta}}_{m|k}(t) = \sum_{i=-k}^{k} h_{m|k}(i)\boldsymbol{\theta}(t+i) \quad (13)$$

i.e., the mean path of parameter estimates can be regarded as a result of passing the sequence of true system parameters through a linear noncausal filter with impulse response (12). For this reason the sequence $\{h_{m|k}(i), i \in I_k\}$ will be further referred to as impulse response associated with the LBF estimator – see Fig. 1.

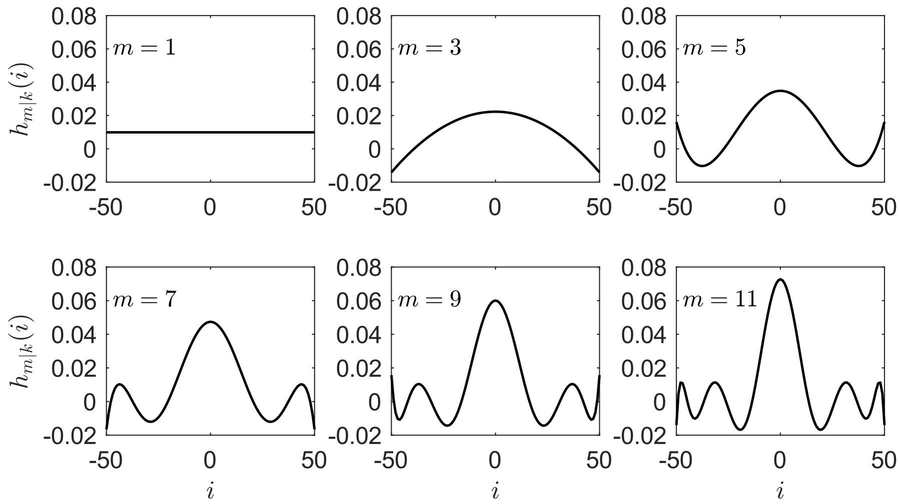

Fig. 1: Impulse responses $h_{m|k}(i)$ associated with LBF estimators of different orders ($k = 50$).

V. Recursive computability

So far our considerations were not restricted to any particular basis set. However, from the practical viewpoint it may be convenient to choose the basis that enables recursive computation of LBF estimators. This would allow one to significantly reduce computational load of the proposed approach, which for large values of m may be substantial.

It is straightforward to check that the recursive computability requirement is met if the basis set is made up of powers of time, namely if

$$\mathbf{f}_{m|k}(i) = \left[1, i/k, \ldots, (i/k)^{m-1}\right]^{\mathrm{T}}$$

which is equivalent to adopting a local Taylor series approximation of parameter trajectory. Note that in this case the basis vector is recursively (backward) computable

$$\mathbf{f}_{m|k}(i-1) = \boldsymbol{\Gamma}_{m|k}\mathbf{f}_{m|k}(i) \tag{14}$$

where

$$\boldsymbol{\Gamma}_{m|k} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ -\frac{1}{k} & 1 & & 0 \\ & & \ddots & \\ \frac{\binom{m-1}{m-1}}{(-k)^{m-1}} & \frac{\binom{m-1}{m-2}}{(-k)^{m-2}} & \ldots & 1 \end{bmatrix}$$

and $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ denotes binomial coefficient. Based on (14), one can easily design the algorithm for recursive computation of the quantities $\mathbf{P}_{m|k}(t)$ and $\mathbf{p}_{m|k}(t)$, needed to evaluate $\widehat{\boldsymbol{\theta}}_{m|k}(t)$:

$$\begin{aligned} \mathbf{P}_{m|k}(t) &= \mathbf{A}_{m|k}\big[\mathbf{P}_{m|k}(t-1) \\ &\quad - \boldsymbol{\psi}_{m|k}(t-1,-k)\boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t-1,-k)\big]\mathbf{A}^{\mathrm{T}}_{m|k} \\ &\quad + \boldsymbol{\psi}_{m|k}(t,k)\boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,k) \\ \mathbf{p}_{m|k}(t) &= \mathbf{A}_{m|k}\big[\mathbf{p}_{m|k}(t-1) \\ &\quad - y(t-k-1)\boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t-1,-k)\big] \\ &\quad + y(t+k)\boldsymbol{\psi}_{m|k}(t,k) \end{aligned} \tag{15}$$

where $\mathbf{A}_{m|k} = \mathbf{I}_n \otimes \boldsymbol{\Gamma}_{m|k}$.

Another basis set for which LBF estimators are recursively computable is made up of harmonic functions: $f_{1|k}(i) = 1$, $f_{2l|k}(i) = \sin\frac{\pi i l}{k}$, $f_{2l+1|k}(i) = \cos\frac{\pi i l}{k}$, $l = 1, \ldots, m_0$, $m = 2m_0 + 1$. It corresponds to a local Fourier series approximation of parameter trajectory.

Remark 1

In the case of the polynomial basis, one can show that $\widetilde{f}_{2l|k}(0) = 0$ for all $l \geq 1$. The straightforward consequence of this fact is that

$$h_{2l|k}(i) = h_{2l-1|k}(i), \quad \forall i \in I_k, l \geq 1. \tag{16}$$

According to (16), parameter tracking capabilities of the LBF algorithm obtained for the basis set $\{1, i/k, \ldots, (i/k)^{2l+1}\}$ are the same as those of the algorithm obtained for the basis set $\{1, i/k, \ldots, (i/k)^{2l}\}$, i.e., to observe a noticeable difference in tracking performance one should increase the order of this basis by 2 (addition of odd powers of time alone does not change tracking capabilities of the algorithm).

VI. Bias-variance tradeoff

Parameter tracking performance of the LBF estimators can be evaluated in terms of the mean square parameter estimation error (MSE) which admits the following decomposition

$$\begin{aligned} \mathrm{E}\left[\| \boldsymbol{\theta}(t) - \widehat{\boldsymbol{\theta}}_{m|k}(t) \|^2\right] &= \| \boldsymbol{\theta}(t) - \bar{\boldsymbol{\theta}}_{m|k}(t) \|^2 \\ &\quad + \mathrm{tr}\left\{\mathrm{cov}[\widehat{\boldsymbol{\theta}}_{m|k}(t)]\right\}. \end{aligned} \tag{17}$$

The first term on the right hand side of (17) can be recognized as the bias component of MSE, and the second term – as its variance component. To study dependence of bias and variance on k (determining the size of the local analysis window) and m (the number of basis functions), we will use the technique of integral approximation.

First of all, note that for the large values of k the discrete-time orthonormal basis functions can be approximately expressed in terms of their continuous-time "prototypes", namely

$$\widetilde{\mathbf{f}}_{m|k}(i) \underset{k\to\infty}{\longrightarrow} \frac{1}{\sqrt{k}}\,\widetilde{\mathbf{f}}_m^0\left(\frac{i}{k}\right) \tag{18}$$

where $\widetilde{\mathbf{f}}_m^0(s) = [\widetilde{f}_1^0(s), \ldots, \widetilde{f}_m^0(s)]^{\mathrm{T}}$ is the vector of orthonormal continuous-time functions of the subspace spanned by $\{f_1^0(s), \ldots, f_m^0(s), s \in [-1, 1]\}$, obeying the condition

$$\int_{-1}^{1} \widetilde{\mathbf{f}}_m^0(s)[\widetilde{\mathbf{f}}_m^0(s)]^{\mathrm{T}} ds = \mathbf{I}_m.$$

To justify this claim, note that after replacing $\widetilde{\mathbf{f}}_{m|k}(i)$ with $(1/\sqrt{k})\widetilde{\mathbf{f}}_m^0(i/k)$ one obtains

$$\begin{aligned}
&\frac{1}{k}\sum_{i=-k}^{k} \widetilde{\mathbf{f}}_m^0\left(\frac{i}{k}\right)\left[\widetilde{\mathbf{f}}_m^0\left(\frac{i}{k}\right)\right]^{\mathrm{T}} \\
&\qquad \cong \frac{1}{k}\int_{-k}^{k} \widetilde{\mathbf{f}}_m^0\left(\frac{\tau}{k}\right)\left[\widetilde{\mathbf{f}}_m^0\left(\frac{\tau}{k}\right)\right]^{\mathrm{T}} d\tau \\
&\qquad = \int_{-1}^{1} \widetilde{\mathbf{f}}_m^0(s)[\widetilde{\mathbf{f}}_m^0(s)]^{\mathrm{T}} ds = \mathbf{I}_m\ .
\end{aligned}$$

Combining (11), (12) and (18), one arrives at

$$N_{m|k} \cong k\left\{\int_{-1}^{1} \left\{[\mathbf{f}_m^0(0)]^{\mathrm{T}}\mathbf{f}_m^0(s)\right\}^2 ds\right\}^{-1} \tag{19}$$

Since under (A3) the covariance matrix of $\widehat{\boldsymbol{\theta}}_{m|k}(t)$ is bounded from below by (10), the variance component of the parameter MSE is inversely proportional to $N_{m|k}$ and hence, according to (19), inversely proportional to the analysis window size k. Furthermore, note that

$$\begin{aligned}
&\int_{-1}^{1} \left\{[\widetilde{\mathbf{f}}_{m+1}^0(0)]^{\mathrm{T}}\widetilde{\mathbf{f}}_{m+1}^0(s)\right\}^2 ds \\
&= \int_{-1}^{1} \left\{[\widetilde{\mathbf{f}}_m^0(0)]^{\mathrm{T}}\widetilde{\mathbf{f}}_m^0(s) + \widetilde{f}_{m+1}^0(0)\widetilde{f}_{m+1}^0(s)\right\}^2 ds \\
&= \int_{-1}^{1} \left\{[\widetilde{\mathbf{f}}_m^0(0)]^{\mathrm{T}}\widetilde{\mathbf{f}}_m^0(s)\right\}^2 ds \\
&\quad + 2\,[\widetilde{\mathbf{f}}_m^0(0)]^{\mathrm{T}}\widetilde{f}_m^0(0)\int_{-1}^{1} \widetilde{\mathbf{f}}_m^0(s)\widetilde{f}_{m+1}^0(s) ds \\
&\quad + \int_{-1}^{1} \left[\widetilde{f}_{m+1}^0(0)\widetilde{f}_{m+1}^0(s)\right]^2 ds
\end{aligned} \tag{20}$$

Since the function $\widetilde{f}_{m+1}^0(s)$ is orthogonal to $\widetilde{\mathbf{f}}_m^0(s)$, the second term on the right hand side of (20) is zero. Additionally, since the last term of (20) is nonnegative, one obtains

$$N_{m+1|k} \le N_{m|k}, \quad \forall m \ge 1$$

which means that the variance component of parameter MSE is a nondecreasing function of the number of basis functions m. Dependence of the equivalent number of observations $N_{m|k}$ on m for the polynomial basis (for $k = 50$) is shown in Fig. 2. In this case, due to (16), it holds that

$$N_{2l|k} = N_{2l-1|k}, \quad N_{2l+1|k} < N_{2l-1|k}, \quad \forall l \ge 1.$$

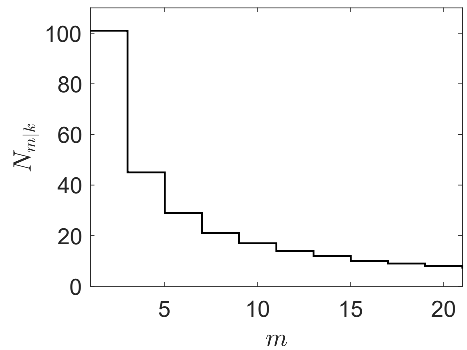

Fig. 2: Dependence of the equivalent number of observations $N_{m|k}$ on the number of basis functions ($k = 50$).

Note that $N_{m|k}$ quickly decays for growing m.

Analysis of the dependence of the bias component of parameter MSE on m and k can be performed by examining frequency response of the filter associated with the LBF estimator

$$H_{m|k}(\omega) = \sum_{i=-k}^{k} h_{m|k}(i)e^{-j\omega i} \tag{21}$$

where $\omega \in (-\pi, \pi]$ denotes normalized angular frequency.

Fig. 3 presents amplitude characteristics evaluated for the polynomial basis set for $m = 1, \ldots, 20$ [due to (16) it holds that $H_{2l|k}(\omega) = H_{2l-1|k}(\omega), \forall l \ge 1$]. Note that by increasing m, one widens the bandwidth of the associated filter, i.e., the frequency range in which parameter changes can be tracked successfully. As a result, for growing m the bias component of parameter MSE drops. To capture dependence of $H_{m|k}(\omega)$ on k note that

$$\begin{aligned}
H_{m|k}(\omega) &\cong \frac{1}{k}\sum_{i=-k}^{k} \left[\mathbf{f}_m^0(0)\right]^{\mathrm{T}} \mathbf{f}_m^0(i/k)e^{-j\omega i} \\
&\cong \frac{1}{k}\int_{-k}^{k} \left[\mathbf{f}_m^0(0)\right]^{\mathrm{T}} \mathbf{f}_m^0(\tau/k)e^{-j\omega\tau} d\tau = H_m^0(k\omega)
\end{aligned} \tag{22}$$

where

$$H_m^0(\omega) = \int_{-1}^{1} [\widetilde{\mathbf{f}}_m^0(0)]^{\mathrm{T}}\widetilde{\mathbf{f}}_m^0(s)e^{-j\omega s} ds. \tag{23}$$

According to (22), the passband of the associated filter is inversely proportional to k, i.e., for fixed m the bias component of the parameter MSE increases along with the analysis window size.

The analysis carried out above shows clearly that conditions which guarantee minimization of the estimation bias (small k and large m) and those that guarantee minimization of the estimation variance (large k and small m) are reciprocal. Since the parameter MSE (17) is the sum of bias and variance components, to achieve good tracking some sort of a compromise must be reached. Clearly, such compromise values of k and m should depend on the speed of parameter variation (degree of system nonstationarity)

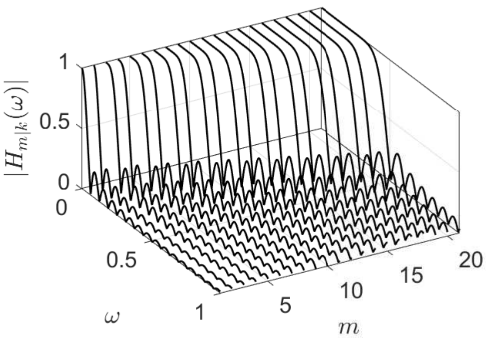

Fig. 3: Dependence of the magnitude of the frequency response $|H_{m|k}(\omega)|$ on the number of basis functions ($k = 50$).

which itself may change with time. Hence, when no prior information about the rate of parameter variation is available it is important to equip LBF algorithms with some kind of adaptive mechanism for tuning k and m.

VII. Adaptive scheduling of design parameters

The proposed adaptive tuning approach exploits the parallel estimation technique. In this framework, one considers MK LBF algorithms, equipped with different settings $m \in \mathcal{M} = \{m_1, \ldots, m_M\}$, $k \in \mathcal{K} = \{k_1, \ldots, k_K\}$, that are run simultaneously (the problem of selection of $\mathcal{K}$ is discussed in some detail in [22]). At each time instant only one of the competing algorithms is selected, i.e., the estimated parameter and variance trajectories have the form

$$\widehat{\boldsymbol{\theta}}_{\widehat{m}(t)|\widehat{k}(t)}(t), \quad \widehat{\rho}_{\widehat{m}(t)|\widehat{k}(t)}(t)$$

where

$$\{\widehat{m}(t), \widehat{k}(t)\} = \arg \min_{\substack{m \in \mathcal{M} \\ k \in \mathcal{K}}} J_{m|k}(t) \tag{24}$$

and $J_{m|k}(t)$ denotes the local decision statistic.

Our decision statistic will be based on cross-validatory analysis [22]. Denote by

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

the unbiased output interpolation error obtained when the measurement $y(t)$ collected at the instant t is eliminated from the estimation process. In this case the local estimate of $\boldsymbol{\theta}(t)$ takes the form

$$\widehat{\boldsymbol{\theta}}^{\circ}_{m|k}(t) = \mathbf{F}_{m|k}\widehat{\boldsymbol{\alpha}}^{\circ}_{m|k}(t) \tag{25}$$

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}^{\circ}_{m|k}(t) &= \arg \min_{\boldsymbol{\alpha}_{m|k}} \sum_{\substack{i=-k \\ i \neq 0}}^{k} [y(t+i) - \boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,i)\boldsymbol{\alpha}_{m|k}]^2 \\ &= [\mathbf{P}^{\circ}_{m|k}(t)]^{-1}\mathbf{p}^{\circ}_{m|k}(t)\end{aligned} \tag{26}$$

where

$$\mathbf{P}^{\circ}_{m|k}(t) = \sum_{\substack{i=-k \\ i \neq 0}}^{k} \boldsymbol{\psi}_{m|k}(t,i)\boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,i)$$

$$\mathbf{p}^{\circ}_{m|k}(t) = \sum_{\substack{i=-k \\ i \neq 0}}^{k} y(t+i)\boldsymbol{\psi}_{m|k}(t,i).$$

Selection of m and k will be based on minimization of the sum of squared interpolation errors

$$J_{m|k}(t) = \sum_{l=-L}^{L} [\varepsilon^{\circ}_{m|k}(t+l)]^2 \tag{27}$$

where L determines the size of the local decision window. Note that the statistic (27) can be computed recursively

$$J_{m|k}(t+1) = J_{m|k}(t) - [\varepsilon^{\circ}_{m|k}(t-L)]^2 + [\varepsilon^{\circ}_{m|k}(t+L+1)]^2.$$

Further reduction of computational load can be achieved by noting that the unbiased interpolation error can be evaluated in terms of the easily computable biased interpolation error

$$\varepsilon_{m|k}(t) = y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}_{m|k}(t).$$

The corresponding formula, derived in the Appendix, has the form

$$\varepsilon^{\circ}_{m|k}(t) = \frac{\varepsilon_{m|k}(t)}{1 - q_{m|k}(t)} \tag{28}$$

where

$$q_{m|k}(t) = \boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,0)\mathbf{P}^{-1}_{m|k}(t)\boldsymbol{\psi}_{m|k}(t,0). \tag{29}$$

According to (28), computation of the unbiased interpolation errors does not require evaluation of the modified parameter estimates (25).

VIII. Simulation results

In our simulation experiment parameter estimation was carried out for a nonstationary two-tap FIR system governed by

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

The input signal was autoregressive Gaussian ($r_u(i) = (0.8)^{|i|}$). System parameters were modeled as triangular chirps with two different linearly increasing beat frequencies – see 4. The shape of parameter trajectories within the simulation interval $[1, T_s]$ was fixed (discrete-time trajectories were generated by "sampling" prototype analog trajectories using different sampling periods). To test identification algorithms under different speeds of parameter variation (SoV), three values of T_s were considered: $T_s = 16000$ (slow variations), $T_s = 8000$ (medium-speed variations) and $T_s = 4000$ (fast variations). Additionally, three signal-to-noise ratios (SNR=E$\{[\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t)]^2\}/\sigma_e^2$) were considered: 5 dB ($\sigma_e^2 = 0.05$), 15 dB ($\sigma_e^2 = 0.005$) and 25 dB ($\sigma_e^2 = 0.0005$). Table I shows comparison of the mean squared parameter estimation errors obtained for 9 LBF estimators (polynomial basis) corresponding to different choices of

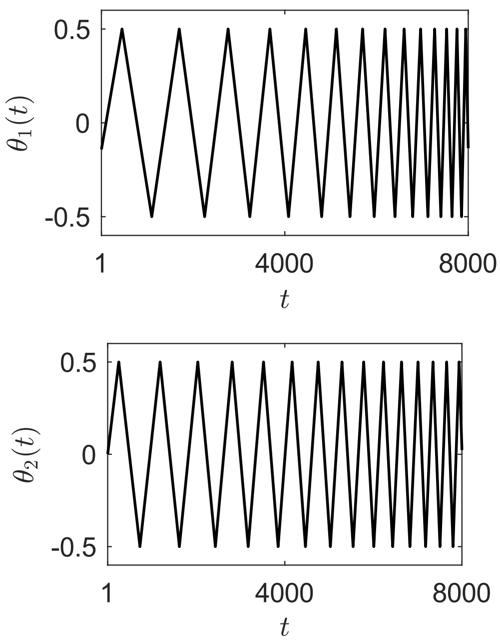

Fig. 4: Evolution of system parameters.

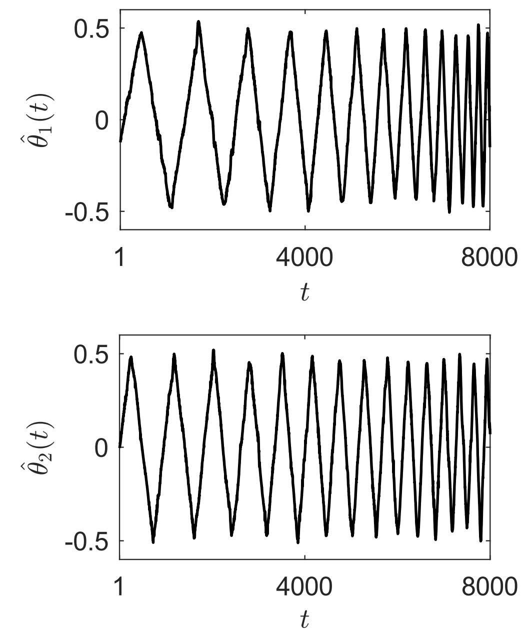

Fig. 5: Typical trajectories of parameter estimates.

design parameters: local analysis window size k (50, 100, 200) and the number of basis functions m (1, 3, 5), and for the adaptive parallel estimation scheme based on cross-validation ($L = 30$).

The results gathered in Tab. I show clearly advantages of the proposed adaptive algorithm – in all cases considered the parallel estimation scheme yielded results that were either better or only slightly inferior to those provided by the best LBF algorithm with fixed settings.

Fig. 5 shows typical realizations of parameter estimates yielded by the adaptive scheme (SNR= 15 dB). Finally, Fig. 6 shows histograms of decisions on the window size (k) and number of basis functions (m) obtained for 100 realizations of the LBF-based parallel estimation scheme (SNR=15 dB).

IX. Conclusion

The problem of identification of a nonstationary stochastic system was considered and solved using the time-localized variant of the basis function approach (LBF). LBF estimators generate a sequence of point estimates of system parameters assuming that parameter trajectories can be locally approximated by linear combinations of a certain number of known functions of time (basis functions). It was shown that for the polynomial basis LBF estimators are recursively computable. It was also shown that two important design parameters (meta-parameters) of LBF estimators – the number of basis functions and the size of the local analysis window – can be selected in an adaptive, data-dependent fashion when several competing LBF estimators, equipped with different settings, are arranged in a parallel estimation scheme and switched appropriately. The proposed selection criterion is based on the leave-one-out cross-validation approach. The resulting adaptive algorithm is computationally attractive and yields results that are either better or only slightly inferior to those provided by the best LBF algorithm with fixed settings incorporated in the parallel scheme.

References

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

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

[3] A. H. Sayed, *Fundamentals of Adaptive Filtering*, Wiley, 2003.

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

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

[6] J. M. Liporace, "Linear estimation of nonstationary signals," *J. Acoust. Soc. Amer.*, vol. 58, pp. 1288-1295, 1975.

[7] M. Hall, A. V. Oppenheim, and A. Willsky, "Time-varying parametric modeling of speech," *Signal Processing*, vol. 5, pp. 267–285, 1983.

[8] R. Charbonnier, M. Barlaud, G. Alengrin and J. Menez, "Results on AR-modelling of non-stationary signals," *Signal Processing*, vol. 12, pp. 143–151, 1987.

[9] M. Niedźwiecki, "Functional series modeling approach to identification of nonstationary stochastic systems," *IEEE Trans. Automat. Contr.*, vol. 33, pp. 955-961, 1988.

[10] M. Niedźwiecki, "Recursive functional series modeling estimators for identification of time-varying plants – more bad news than good?," *IEEE Trans. Automat. Contr.*, vol. 35, pp. 610-616, 1990.

[11] R. B. Mrad, S. D. Fassois, and J. A. Levitt, "A polynomial-algebraic method for non-stationary TARMA signal analysis - Part I: The method," *Signal Processing*, vol. 65, pp. 1–19, 1998.

[12] R. B. Zou, H. Wang H., and K. H. Chon, "A robust time-varying identification algorithm using basis functions," *Annals of Biomedical Engineering*, vol. 31, pp. 840–853, 2003.

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

TABLE I: Mean squared parameter estimation errors obtained for 9 NWBF estimators corresponding to different choices of design parameters k (50, 100, 200) and m (1, 3, 5), and for the proposed adaptive estimation scheme (A). The averages were computed for 100 process realizations, 3 speeds of parameter variation (SoV) and 3 signal-to-noise ratios (SNR).

SNR	SoV	slow			medium			fast		
	$k\backslash m$	1	3	5	1	3	5	1	3	5
5 dB	50	3.4E-03	6.7E-03	1.1E-02	6.4E-03	6.9E-03	1.1E-02	2.9E-02	9.3E-03	1.1E-02
	100	4.0E-03	3.3E-03	5.2E-03	2.1E-02	4.6E-03	5.6E-03	8.6E-02	2.5E-02	1.2E-02
	200	1.7E-02	2.6E-03	2.8E-03	7.8E-02	1.8E-02	5.7E-03	1.7E-01	8.5E-02	4.6E-02
	A		4.3E-03			5.7E-03			8.2E-03	
15 dB	50	8.4E-04	6.9E-04	1.1E-03	3.8E-03	8.4E-04	1.1E-03	2.7E-02	3.4E-03	1.6E-03
	100	2.7E-03	4.7E-04	5.6E-04	1.9E-02	1.7E-03	8.8E-04	8.5E-02	2.3E-02	7.8E-03
	200	1.6E-02	1.2E-03	6.0E-04	7.7E-02	1.6E-02	3.5E-03	1.7E-01	8.4E-02	4.4E-02
	A		5.7E-04			8.3E-04			1.6E-03	
25 dB	50	5.9E-04	8.7E-05	1.1E-04	3.6E-03	2.6E-04	1.6E-04	2.6E-02	2.8E-03	6.5E-04
	100	2.6E-03	1.8E-04	9.4E-05	1.9E-02	1.4E-03	4.3E-04	8.5E-02	2.2E-02	7.3E-03
	200	1.6E-02	1.1E-03	3.4E-04	7.7E-02	1.6E-02	3.3E-03	1.7E-01	8.3E-02	4.4E-02
	A		7.9E-05			1.5E-04			6.5E-04	

‘

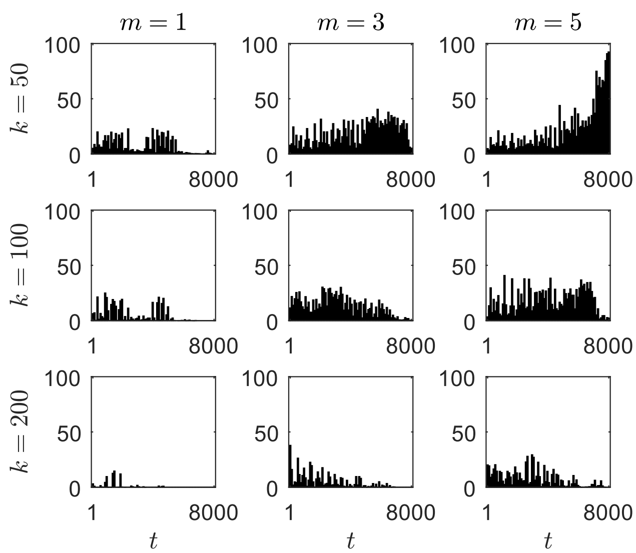

Fig. 6: Histograms of decisions on the window size (k) and the number of basis functions (m) obtained for 100 realizations of the adaptive LBF-based scheme (SNR= 15 dB, medium-speed parameter variations). Each time bin corresponds to 100 consecutive samples.

[14] M. Niedźwiecki and S. Gackowski, "New approach to noncausal identification of nonstationary FIR systems subject to both smooth and abrupt parameter changes," *IEEE Trans. Automat. Contr.*, vol. 58, pp. 1847–1853, 2013.

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

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

[17] 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.

[18] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data with simplified least squares procedures," *Annals of Chemistry*, vol. 36, pp. 1627–1639, 1964.

[19] R.W. Schafer, "What is a Savitzky-Golay filter?," *IEEE Signal Process. Mag.*, vol. 28, pp. 111-117, 2011.

[20] M. K. Tsatsanis and G. B. Giannakis, "Modeling and equalization of rapidly fading channels," *Int. J. Adaptive Contr. Signal Process.* vol. 10, pp. 159-176, 1996.

[21] R. Dahlhaus, "Locally stationary processes," *Handbook Statist.*, vol. 25, pp. 1–37, 2012.

[22] M. Niedźwiecki, M. Ciołek, and Y. Kajikawa, "On adaptive covariance and spectrum estimation of locally stationary multivariate processes," *Automatica*, vol. 82, pp. 1–12, 2017.

Appendix [derivation of (28)]

Note that

$$\mathbf{P}^{\circ}_{m|k}(t) = \mathbf{P}_{m|k}(t) - \boldsymbol{\psi}_{m|k}(t,0)\boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,0)$$
$$\mathbf{p}^{\circ}_{m|k}(t) = \mathbf{p}_{m|k}(t) - y(t)\boldsymbol{\psi}_{m|k}(t,0)$$

and

$$\varepsilon_{m|k}(t) = y(t) - \boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,0)\widehat{\boldsymbol{\alpha}}_{m|k}(t)$$
$$\varepsilon^{\circ}_{m|k}(t) = y(t) - \boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,0)\widehat{\boldsymbol{\alpha}}^{\circ}_{m|k}(t).$$

Exploiting the fact that $\mathbf{p}_{m|k}(t) = \mathbf{P}_{m|k}(t)\widehat{\boldsymbol{\alpha}}_{m|k}(t)$ and using the matrix inversion lemma [3], one arrives at

$$\begin{aligned}
&\boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,0)\widehat{\boldsymbol{\alpha}}^{\circ}_{m|k}(t) = \boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,0) \\
&\times \left[\mathbf{P}^{-1}_{m|k}(t) + \frac{\mathbf{P}^{-1}_{m|k}(t)\boldsymbol{\psi}_{m|k}(t,0)\boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,0)\mathbf{P}^{-1}_{m|k}(t)}{1-\boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,0)\mathbf{P}^{-1}_{m|k}(t)\boldsymbol{\psi}_{m|k}(t,0)}\right] \\
&\qquad \times \left[\mathbf{P}_{m|k}(t)\widehat{\boldsymbol{\alpha}}_{m|k}(t) - y(t)\boldsymbol{\psi}_{m|k}(t,0)\right] \\
&= \boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,0)\widehat{\boldsymbol{\alpha}}_{m|k}(t) + \frac{q_{m|k}(t)\boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,0)\widehat{\boldsymbol{\alpha}}_{m|k}(t)}{1-q_{m|k}(t)} \\
&- y(t)q_{m|k}(t) - y(t)\frac{q^2_{m|k}(t)}{1-q_{m|k}(t)} = \boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,0)\widehat{\boldsymbol{\alpha}}_{m|k}(t) \\
&\quad - \frac{q_{m|k}(t)}{1-q_{m|k}(t)}\left[y(t) - \boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,0)\widehat{\boldsymbol{\alpha}}_{m|k}(t)\right]
\end{aligned}$$

leading to

$$\begin{aligned}
\varepsilon^{\circ}_{m|k}(t) &= \varepsilon_{m|k}(t) + \frac{q_{m|k}(t)}{1-q_{m|k}(t)}\varepsilon_{m|k}(t) \\
&= \frac{\varepsilon_{m|k}(t)}{1-q_{m|k}(t)}\ .
\end{aligned}$$