Postprint of: Niedźwiecki M., Ciołek M., Generalized Savitzky–Golay filters for identification of nonstationary systems, Automatica, Vol. 108 (2019), 108477, DOI: 10.1016/j.automatica.2019.06.029

Generalized Savitzky-Golay filters for identification of nonstationary systems

Maciej Niedźwiecki [a], Marcin Ciołek [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 problem of identification of nonstationary systems using noncausal estimation schemes is considered and a new class of identification algorithms, combining the basis functions approach with local estimation technique, is described. Unlike the classical basis function estimation schemes, the proposed local basis function estimators are not used to obtain interval approximations of the parameter trajectory, but provide a sequence of point estimates corresponding to consecutive instants of time. Based on the results of theoretical analysis conducted for nonstationary finite impulse response systems the paper proposes two mechanisms for adaptive selection of the number of basis functions and the size of the local analysis window.

Key words: basis functions, identification and smoothing of nonstationary systems, Savitzky-Golay filters

1 Introduction

When system model is used for prediction purposes, the identification task can be formulated as the problem of finding the predictive (usually nonlinear) mapping from the space of regression (input) variables to the space of system outputs. In this case identification can be efficiently carried out using explicit (Billings, 2013) or implicit (machine-learning-based) (Liu, Principe & Haykin, 2011) nonlinear modeling techniques. In the current paper, which *is not* prediction-oriented as it involves noncausal estimation techniques, we consider the case where the linear form of the process description is enforced by the underlying application, such as channel equalization (Tsatsanis & Giannakis, 1996) or parametric spectrum estimation (Dahlhaus, 2012), and the identification objective is to track parameters of the corresponding ground truth model with the greatest possible accuracy. Note that in both applications mentioned above noncausal estimation is feasible.

The existing classical approaches to identification of nonstationary stochastic systems/signals (along the lines indicated above) can be coarsely divided into model-free and model-based solutions. In the model-free case no explicit description of process parameter variation is adopted. Instead, assuming that unknown process parameters vary "sufficiently slowly", so that the process can be regarded as locally stationary, identification can be carried out using the localized (weighted or windowed) versions of the least squares or maximum likelihood estimators, or using the stochastic gradient algorithms (Sayed, 2003). When the model-based approach is taken, an explicit model of parameter changes is incorporated, either deterministic or stochastic (Norton, 1975). In the deterministic setup, parameter trajectories are approximated by linear combinations of a certain number of known functions of time, called basis functions (BF). Parameter estimates can be obtained by means of local estimation of the best-fitting approximation coefficients (Rao, 1970), (Mendel, 1973), (Liporace, 1975), (Grenier, 1981), (Hall, 1983), (Niedźwiecki, 1988), (Tsatsanis & Giannakis, 1996), (Mrad, Fassois & Lewitt, 1998), (Zou et al., 2003), (Poulimenos & Fassois, 2006).

In this paper we will describe and analyze a new class of identification algorithms, combining the BF approach with local estimation technique. Unlike (Niedźwiecki, 1988), the proposed local basis functions (LBF) estimators are not used to obtain interval approximations of the parameter trajectory, but are regarded as a source of point estimates corresponding to a particular time instant t. This means that the estimation procedure is carried out independently for each value of t, based on the input/output data gathered in the local analysis interval centered at t. Since the estimates obtained in this way are noncausal, i.e., at each time instant t they rely on both "past" and "future" (relative to t) process observations, they cannot be used in real-time applications

⋆ This work was 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.

Email addresses: maciekn@eti.pg.edu.pl (Maciej Niedźwiecki), marcin.ciolek@pg.edu.pl (Marcin Ciołek).

such as adaptive prediction or adaptive control. However, many other applications exist, such as the ones indicated above, that are not time-critical in the sense that the model-based decisions can be delayed by a certain number of sampling intervals. Such a processing mode is often referred to as almost real time. Noncausal estimation schemes can significantly reduce the bias component of the mean square parameter estimation error. Owing to this, their estimation performance is usually considerably better than that of the comparable causal schemes.

On the qualitative level, the LBF approach can be considered an extension, to the process identification case, of the signal smoothing technique known as Savitzky-Golay (SG) filtering (Schafer, 2011) (filtering successive subsets of adjacent data points with a low-degree polynomial by the method of least squares). The paper extends, in several directions, the earlier work on doubly exponentially weighted basis function estimators (Niedźwiecki & Gackowski, 2013).

2 Estimation scheme

2.1 Notation

Consider a nonstationary stochastic process governed by the equation

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

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 unknown n-dimensional vector of time-varying process parameters, $\boldsymbol{\varphi}(t) = [\varphi^1(t), \ldots, \varphi^n(t)]^{\mathrm{T}}$ denotes regression vector and $e(t)$ denotes white noise. The more explicit structure of the regression vector, as well as detailed assumptions about the sequences $\{\boldsymbol{\theta}(t)\}$, $\{\boldsymbol{\varphi}(t)\}$ and $e(t)\}$ will be discussed later.

The identification approach pursued in this paper will be based on the method of basis functions, namely, we will assume that in the local time interval $T_k(t) = [t-k, t+k]$ of length $l_k = 2k+1$, centered at t, each process parameter can be expressed as a linear combination of m linearly independent basis functions, namely

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

where

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

and $f^0_l(s), s \in [-1, 1], l = 1, \ldots, m$ denote continuous-time square integrable basis generating functions defined on the interval $[-1, 1]$. 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 $\mathcal{F}_{m|k}$.

If no prior knowledge about the nature of parameter variation is available, selection of the basis has to rely on some general approximation guidelines. The most frequently used general purpose bases are comprised of powers of time (Taylor series approximation)

$$f^0_l(s) = s^{l-1}, \quad l = 1, \ldots, m \tag{4}$$

or harmonic functions (Fourier series approximation)

$$\begin{gathered} f^0_1(s) = 1, \; f^0_{2l}(s) = \sin[\pi s l], \quad f^0_{2l+1}(s) = \cos[\pi s l] \\ l = 1, \ldots, m_0, \; m = 2m_0 + 1. \end{gathered} \tag{5}$$

Denote by $\boldsymbol{\alpha}^j_{m|k} = [a^j_{1,m|k}, \ldots, a^j_{m,m|k}]^{\mathrm{T}}$ the vector of coefficients describing evolution of the j-th system parameter $\theta^j(t)$ in the interval $T_k(t)$ and let $\boldsymbol{\alpha}_{m|k} = \left[(\boldsymbol{\alpha}^1_{m|k})^{\mathrm{T}}, \ldots, (\boldsymbol{\alpha}^n_{m|k})^{\mathrm{T}}\right]^{\mathrm{T}}$. Finally, denote by $\boldsymbol{\psi}_{m|k}(t, i) = \boldsymbol{\varphi}(t+i) \otimes \mathbf{f}_{m|k}(i)$ the generalized regression vector where $\otimes$ denotes Kronecker product of the respective vectors/matrices and $\mathbf{f}_{m|k}(i) = [f_{1|k}(i), \ldots, f_{m|k}(i)]^{\mathrm{T}}$ is the vector of basis functions.

Combining (1) with (2) and using the shorthand notation introduced above, the local process model which will be the subject of identification, can be written down in the following compact form

$$\begin{gathered} y(t+i) = \boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t, i)\boldsymbol{\alpha}_{m|k} + e(t+i) \\ \mathrm{E}[e^2(t+i)] = \rho, \quad i \in I_k. \end{gathered} \tag{6}$$

It should be stressed that the local process description obtained after replacing the quantities $\boldsymbol{\alpha}_{m|k}$ and ρ in (6) with the corresponding estimates $\widehat{\boldsymbol{\alpha}}_{m|k}(t)$ and $\widehat{\rho}_{m|k}(t)$ will be further regarded as a pointwise, rather than interval, process model, valid *only* at the instant t ($i = 0$). This means that unlike (Niedźwiecki, 1988), the parameter fitting procedure, described below, will be carried out independently for each value of t. Such a technique is usually called sliding window approach.

2.2 Identification procedure

Our identification procedure will be based on the method of weighted least squares (WLS)

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

where $\{w_k(i), i = -k, \ldots, k\}$, $w_k(0) = 1$, denotes a nonnegative, symmetric bell-shaped window of width $2k+1$ used for localization purposes. We will further assume that $w_k(i) = w^0(i/k)$, where $w^0(s), s \in [-1, 1]$ denotes the continuous-time window generating function.

Straightforward calculations lead to

$$\widehat{\boldsymbol{\alpha}}_{m|k}(t) = \mathbf{P}^{-1}_{m|k}(t)\mathbf{p}_{m|k}(t) \tag{7}$$

where

$$\begin{aligned}\mathbf{P}_{m|k}(t) &= \sum_{i=-k}^{k} w_k(i)\boldsymbol{\psi}_{m|k}(t,i)\boldsymbol{\psi}_{m|k}^{\mathrm{T}}(t,i)\\ \mathbf{p}_{m|k}(t) &= \sum_{i=-k}^{k} w_k(i)y(t+i)\boldsymbol{\psi}_{m|k}(t,i)\end{aligned} \tag{8}$$

– provided, of course, that the matrix $\mathbf{P}_{m|k}(t)$ is nonsingular. The estimate of the noise variance can be obtained in the form

$$\begin{aligned}\widehat{\rho}_{m|k}(t) &= \frac{1}{L_k}\sum_{i=-k}^{k} w_k(i)[y(t+i) - \boldsymbol{\psi}_{m|k}^{\mathrm{T}}(t,i)\widehat{\boldsymbol{\alpha}}_{m|k}(t)]^2\\ &= \frac{1}{L_k}[c_k(t) - \mathbf{p}_{m|k}^{\mathrm{T}}(t)\widehat{\boldsymbol{\alpha}}_{m|k}(t)]\end{aligned} \tag{9}$$

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

$$L_k = \sum_{i=-k}^{k} w_k(i) \cong k\int_{-1}^{1} w^0(s)ds \tag{10}$$

denotes the effective window width.

Finally, based on (2), the instantaneous estimate of $\boldsymbol{\theta}(\cdot)$ can be evaluated

$$\widehat{\boldsymbol{\theta}}_{m|k}(t) = \mathbf{F}_{m|k}\widehat{\boldsymbol{\alpha}}_{m|k}(t), \quad \mathbf{F}_{m|k} = \mathbf{I}_n \otimes \mathbf{f}_{m|k}^{\mathrm{T}}(0). \tag{11}$$

The estimators $\widehat{\boldsymbol{\theta}}_{m|k}(t)$ and $\widehat{\rho}_{m|k}(t)$ will be further referred to as local basis function (LBF) estimators.

3 Basic properties of LBF estimators

3.1 Basis orthonormalization

To simplify our further developments, we will rewrite the LBF algorithm in terms of the orthonormal basis set of $\mathcal{F}_{m|k}$. Define the inner (dot) product of the sequences $f(i)$ and $g(i)$, $i \in I_k$ in the form $\langle f, g\rangle = \sum_{i=-k}^{k} w(i)f(i)g(i)$, $\parallel f \parallel^2 = \langle f, f\rangle$. The basis set of $\mathcal{F}_{m|k}$ orthonormal with respect to the adopted inner product, further referred to as w-orthonormal basis set of $\mathcal{F}_{m|k}$, will be denoted by $\widetilde{f}_{1|k}(i), \ldots, \widetilde{f}_{m|k}(i), i \in I_k$. Note that such a basis set obeys the condition $\sum_{i=-k}^{k} w_k(i)\widetilde{\mathbf{f}}_{m|k}(i)\widetilde{\mathbf{f}}_{m|k}^{\mathrm{T}}(i) = \mathbf{I}_m$ where $\widetilde{\mathbf{f}}_{m|k}(i) = [\widetilde{f}_{1|k}(i), \ldots, \widetilde{f}_{m|k}(i)]^{\mathrm{T}}$. The w-orthonormal basis can be obtained from the original basis by means of applying the Gram-Schmidt procedure, or using the following linear transformation $\widetilde{\mathbf{f}}_{m|k}(i) = \mathbf{Q}_{m|k}^{-1/2}\,\mathbf{f}_{m|k}(i)$ where $\mathbf{Q}_{m|k}^{-1/2} = [\mathbf{Q}_{m|k}^{1/2}]^{-1}$, $\mathbf{Q}_{m|k} = \sum_{i=-k}^{k} w_k(i)\mathbf{f}_{m|k}(i)\mathbf{f}_{m|k}^{\mathrm{T}}(i)$ and $\mathbf{Q}_{m|k}^{1/2}$ denotes any square root of the matrix $\mathbf{Q}_{m|k}$.

After this modification the LBF estimator of $\boldsymbol{\theta}(t)$ takes the form

$$\widetilde{\boldsymbol{\theta}}_{m|k}(t) = \widetilde{\mathbf{F}}_{m|k}\widetilde{\boldsymbol{\alpha}}_{m|k}(t), \quad \widetilde{\mathbf{F}}_{m|k} = \mathbf{I}_n \otimes \widetilde{\mathbf{f}}_{m|k}^{\mathrm{T}}(0) \tag{12}$$

where $\widetilde{\boldsymbol{\alpha}}_{m|k}(t) = \widetilde{\mathbf{P}}_{m|k}^{-1}(t)\widetilde{\mathbf{p}}_{m|k}(t)$ and the quantities $\widetilde{\mathbf{P}}_{m|k}(t)$ and $\widetilde{\mathbf{p}}_{m|k}(t)$ are defined in an analogous way as $\mathbf{P}_{m|k}(t)$ and $\mathbf{p}_{m|k}(t)$, respectively.

One can easily show that $\widetilde{\boldsymbol{\theta}}_{m|k}(t) = \widehat{\boldsymbol{\theta}}_{m|k}(t)$ which means that, as expected, the change of the basis of the subspace $\mathcal{F}_{m|k}$ has no influence on the obtained estimation results.

The noise variance estimate takes the form analogous to (9)

$$\widetilde{\rho}_{m|k}(t) = \frac{1}{L_k}[c_k(t) - \widetilde{\mathbf{p}}_{m|k}^{\mathrm{T}}(t)\widetilde{\boldsymbol{\alpha}}_{m|k}(t)] = \widehat{\rho}_{m|k}(t). \tag{13}$$

3.2 Static characteristics

To obtain analytical results, we will assume that (1) is a finite impulse response (FIR) system, i.e., $\boldsymbol{\varphi}(t) = [u(t-1), \ldots, u(t-n)]^{\mathrm{T}}$ where $u(t)$ denotes an observable input sequence. Furthermore, 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.

(A3) In the time interval $T_k(t)$ the true process parameters obey (2), i.e., all parameter trajectories belong to the subspace $\mathcal{F}_{m|k}$ spanned by the basis functions.

Note that under (A3) it holds that

$$y(t+i) = \widetilde{\boldsymbol{\psi}}_{m|k}^{\mathrm{T}}(t,i)\boldsymbol{\alpha}_{m|k}^0 + e(t+i), \quad i \in I_k \tag{14}$$

where $\boldsymbol{\alpha}_{m|k}^0 = \left[\mathbf{I}_n \otimes \mathbf{Q}_{m|k}^{\mathrm{T}/2}\right]\boldsymbol{\alpha}_{m|k}$ denotes the vector of true basis function coefficients and $\widetilde{\boldsymbol{\psi}}_{m|k}(t,i) = \left[\mathbf{I}_n \otimes \mathbf{Q}_{m|k}^{-1/2}\right]\boldsymbol{\psi}_{m|k}(t,i) = \boldsymbol{\varphi}(t+i) \otimes \widetilde{\mathbf{f}}_{m|k}(i)$.

Denote by $\Delta\widetilde{\boldsymbol{\alpha}}_{m|k}(t) = \widetilde{\boldsymbol{\alpha}}_{m|k}(t) - \boldsymbol{\alpha}_{m|k}^0$ the estimation error. It is straightforward to check that

$$\begin{aligned}\Delta\widetilde{\boldsymbol{\alpha}}_{m|k}(t) &= \widetilde{\mathbf{P}}_{m|k}^{-1}(t)\boldsymbol{\xi}_{m|k}(t)\\ \boldsymbol{\xi}_{m|k}(t) &= \sum_{i=-k}^{k} w_k(i)\widetilde{\boldsymbol{\psi}}_{m|k}(t,i)e(t+i).\end{aligned} \tag{15}$$

Using a similar technique as that adopted in (Niedźwiecki, 1988), one can show that under the assumption (A1) the matrix $\widetilde{\mathbf{P}}_{m|k}(t)$ converges in the mean squared sense to a constant matrix $\lim_{k\to\infty} \widetilde{\mathbf{P}}_{m|k}(t) = \boldsymbol{\Phi} \otimes \mathbf{I}_m = \bar{\mathbf{P}}_m$ m.s., where $\boldsymbol{\Phi} = \mathrm{E}[\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)] > 0$. This limiting result justifies the following approximation valid for sufficiently long analysis intervals

$$\Delta\widetilde{\boldsymbol{\alpha}}_{m|k}(t) \cong \bar{\mathbf{P}}_m^{-1}\boldsymbol{\xi}_{m|k}(t). \tag{16}$$

All further analytical results will be based on (16). The errors that arise in consequence of adopting this approximation can be analyzed in an analogous way as in (Niedźwiecki, 1988).

Note that according to assumption (A2) it holds that $\mathrm{E}[\boldsymbol{\xi}_{m|k}(t)] = 0$, which entails $\mathrm{E}[\Delta\widetilde{\boldsymbol{\alpha}}_{m|k}(t)] \cong 0$ and consequently $\mathrm{E}[\widetilde{\boldsymbol{\theta}}_{m|k}(t)] = \mathrm{E}[\widehat{\boldsymbol{\theta}}_{m|k}(t)] \cong \boldsymbol{\theta}(t)$. Hence, in the case considered, the LBF estimator $\widetilde{\boldsymbol{\theta}}_{m|k}(t)$ is (approximately) unbiased. To obtain expression for its covariance matrix, note that

$$\begin{aligned}\mathrm{cov}[\widetilde{\boldsymbol{\alpha}}_{m|k}(t)] &= \mathrm{E}[\Delta\widetilde{\boldsymbol{\alpha}}_{m|k}(t)\Delta\widetilde{\boldsymbol{\alpha}}_{m|k}^{\mathrm{T}}(t)] \\ &\cong \bar{\mathbf{P}}_m^{-1}\ \mathrm{E}[\boldsymbol{\xi}_{m|k}(t)\boldsymbol{\xi}_{m|k}^{\mathrm{T}}(t)]\ \bar{\mathbf{P}}_m^{-1}\end{aligned} \tag{17}$$

Since $\bar{\mathbf{P}}_m^{-1} = \boldsymbol{\Phi}^{-1} \otimes \mathbf{I}_m$ and $\mathrm{E}[\boldsymbol{\xi}_{m|k}(t)\boldsymbol{\xi}_{m|k}^{\mathrm{T}}(t)] = \rho\ [\boldsymbol{\Phi} \otimes \mathbf{W}_{m|k}]$ where $\mathbf{W}_{m|k} = \sum_{i=-k}^{k} w_k^2(i)\widetilde{\mathbf{f}}_{m|k}(i)\widetilde{\mathbf{f}}_{m|k}^{\mathrm{T}}(i)$, one arrives at

$$\begin{aligned}\mathrm{cov}[\widetilde{\boldsymbol{\alpha}}_{m|k}(t)] &\cong \rho\ [\boldsymbol{\Phi}^{-1} \otimes \mathbf{I}_m]\ [\boldsymbol{\Phi} \otimes \mathbf{W}_{m|k}][\boldsymbol{\Phi}^{-1} \otimes \mathbf{I}_m] \\ &= \rho[\boldsymbol{\Phi}^{-1} \otimes \mathbf{W}_{m|k}]\end{aligned} \tag{18}$$

which stems from the identity $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$. Finally,

$$\begin{aligned}&\mathrm{cov}[\widetilde{\boldsymbol{\theta}}_{m|k}(t)] = \mathrm{cov}[\widehat{\boldsymbol{\theta}}_{m|k}(t)] = \widetilde{\mathbf{F}}_{m|k}\mathrm{cov}[\widetilde{\boldsymbol{\alpha}}_{m|k}(t)]\widetilde{\mathbf{F}}_{m|k}^{\mathrm{T}} \\ &\cong \rho[\mathbf{I}_n \otimes \widetilde{\mathbf{f}}_{m|k}^{\mathrm{T}}(0)][\boldsymbol{\Phi}^{-1} \otimes \mathbf{W}_{m|k}][\mathbf{I}_n \otimes \widetilde{\mathbf{f}}_{m|k}(0)] = \frac{\rho\boldsymbol{\Phi}^{-1}}{N_{m|k}}\end{aligned} \tag{19}$$

where

$$\begin{aligned}N_{m|k} &= [\widetilde{\mathbf{f}}_{m|k}^{\mathrm{T}}(0)\mathbf{W}_{m|k}\widetilde{\mathbf{f}}_{m|k}(0)]^{-1} \\ &= \left\{\sum_{i=-k}^{k}[w_k(i)\widetilde{\mathbf{f}}_{m|k}^{\mathrm{T}}(0)\widetilde{\mathbf{f}}_{m|k}(i)]^2\right\}^{-1}\end{aligned} \tag{20}$$

is the quantity that will be further referred to as equivalent number of observations. The name stems from the fact that in the time invariant case the covariance matrix of the LBF estimator is the same as the covariance matrix of the LS estimator incorporating $N_{m|k}$ data points. Note that $N_{m|k}$ depends on m and differs from the effective number of observations, given by (10). It is is proportional to the window size k [which can be shown using the integral approximation technique – cf. (10)] and inversely proportional to the number of basis functions m – the latter effect is illustrated in Fig. 1.

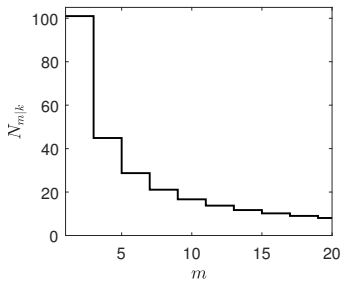

Fig. 1. Dependence of the equivalent number of observations $N_{m|k}$ on the number of basis functions [polynomial basis, rectangular window of width 101].

3.3 Dynamic characteristics

To study parameter matching characteristics of LBF estimators we will again use the approximation $\widetilde{\mathbf{P}}_{m|k}^{-1}(t) \cong \bar{\mathbf{P}}_m^{-1}$. However, unlike previously, the only assumption we will make about the sequence $\{\boldsymbol{\theta}(t)\}$ is that

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

Note that (A3*) *does not* imply that parameter trajectories belong to the subspace $\mathcal{F}_{m|k}$, i.e., that system parameters can be exactly modeled as linear combinations of basis functions.

Denote by $\bar{\boldsymbol{\theta}}_{m|k}(t) = \mathrm{E}_{\Omega_k(t)}[\widetilde{\boldsymbol{\theta}}_{m|k}(t)]$ the average path of parameter estimates, where the expectation is carried over $\Omega_k(t) = \{\boldsymbol{\varphi}(t+i), e(t+i), i \in I_k\}$. It is easy to derive the following relationship that holds under (A1), (A2) and (A3*)

$$\begin{aligned}\bar{\boldsymbol{\theta}}_{m|k}(t) &\cong [\boldsymbol{\Phi}^{-1} \otimes \widetilde{\mathbf{f}}_{m|k}^{\mathrm{T}}(0)] \sum_{i=-k}^{k} w_k(i)[\boldsymbol{\Phi}\boldsymbol{\theta}(t+i)] \otimes \widetilde{\mathbf{f}}_{m|k}(i) \\ &= \sum_{i=-k}^{k} h_{m|k}(i)\boldsymbol{\theta}(t+i)\end{aligned} \tag{21}$$

where

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

is the sequence that will be further referred to as the impulse response associated with the LBF estimator. According to (21), the mean path of parameter estimates $\{\bar{\boldsymbol{\theta}}_{m|k}(t)\}$ can be approximately viewed as an output of a linear noncausal filter with impulse response $\{h_{m|k}(i)\}$ excited by the process $\{\boldsymbol{\theta}(t)\}$.

Note that [cf. (20)] $\sum_{i=-k}^{k} h_{m|k}^2(i) = N_{m|k}$. Additionally, it can be shown that if a constant function $f(i) = 1, \forall i \in I_k$, belongs to the subspace $\mathcal{F}_{m|k}$, it holds that $\sum_{i=-k}^{k} h_{m|k}(i) = 1$ which means that (22) is an impulse response of a lowpass filter.

Based on (21), one can easily quantify matching capabilities of $\widetilde{\boldsymbol{\theta}}_{m|k}(t)$ in the frequency domain. Suppose that

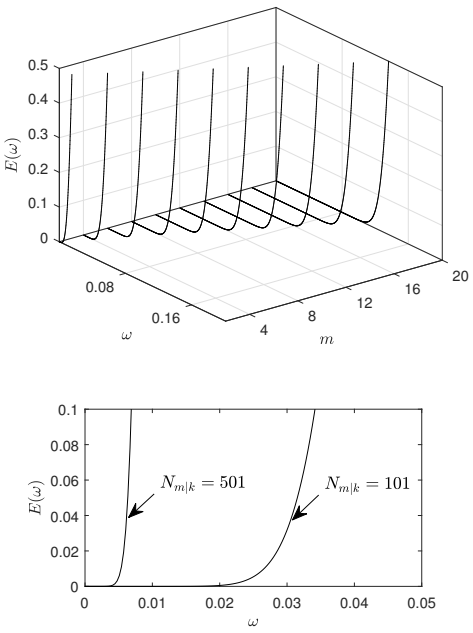

Fig. 2. Dependence of parameter matching characteristics (polynomial basis, rectangular window) on the number of basis functions m (for $N_{m|k} = 101$, upper figure) and on equivalent number of observations $N_{m|k}$ (for $m = 5$, lower figure).

$\theta^l(t)$ – the l-th component of $\boldsymbol{\theta}(t)$ – is a wide-sense stationary random process with a spectral density function $S^l_\theta(\omega), \omega \in (-\pi, \pi]$. Then the mean squared value of the bias error can be expressed in the form

$$\mathrm{E}_\Theta\{[\bar{\theta}^l_{m|k}(t) - \theta^l(t)]^2\} \cong \frac{1}{\pi}\int_0^\pi E_{m|k}(\omega) S^l_\theta(\omega) d\omega \quad (23)$$

where averaging is carried out over the set Θ of different realizations of the parameter trajectory,

$$E_{m|k}(\omega) = |1 - H_{m|k}(\omega)|^2 \,, \quad (24)$$

and

$$H_{m|k}(\omega) = \sum_{i=-k}^{k} h_{m|k}(i) e^{-j\omega i} \quad (25)$$

denotes frequency response associated with the LBF estimator. The function $E_{m|k}(\omega)$ will be further referred to as parameter matching characteristic of $\widetilde{\boldsymbol{\theta}}_{m|k}(\cdot)$. It is clear from (23) that good estimation performance can be achieved only in the case where the spectral density function $S^l_\theta(\omega)$ matches the passband region of $E_{m|k}(\omega)$. The passband widens with growing number of basis functions. Using the integral approximation technique it is also possible to show that its width is inversely proportional to the window size k (and hence to $N_{m|k}$). Both effects can be clearly seen in Fig. 2.

3.4 Recursive computability

LBF estimators are computationally expensive. The computational load can be substantially reduced if the basis set and window shape are chosen so as to guarantee recursive computability of $\widehat{\boldsymbol{\theta}}_{m|k}(t) = \widetilde{\boldsymbol{\theta}}_{m|k}(t)$. The simplest choice which guarantees this property is the polynomial (4) or harmonic (5) basis and cosinusoidal window $w_k(i) = \cos\frac{\pi i}{2k}$. First, note that the vector of powers of time is recursively forward/backward computable. In particular, one can easily check that

$$\mathbf{f}_{m|k}(i-1) = \boldsymbol{\Gamma}_{m|k}\mathbf{f}_{m|k}(i) \quad (26)$$

where

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

and $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ denotes binomial coefficient. The same relationship holds for the harmonic basis $\mathbf{f}_{m|k}(i) = \left[1, \sin\frac{\pi i}{k}, \cos\frac{\pi i}{k}, \dots, \sin\frac{\pi i m_0}{k}, \cos\frac{\pi i m_0}{k}\right]^{\mathrm{T}}$, $m = 2m_0 + 1$. In this case $\boldsymbol{\Gamma}_{m|k}$ is a block diagonal matrix of the form

$$\boldsymbol{\Gamma}_{m|k} = \mathrm{bl\ diag}\{1, \mathbf{G}_{1|k}, \dots, \mathbf{G}_{m_0|k}\}$$

where

$$\mathbf{G}_{l|k} = \begin{bmatrix} \cos\frac{\pi l}{k} & -\sin\frac{\pi l}{k} \\ \sin\frac{\pi l}{k} & \cos\frac{\pi l}{k} \end{bmatrix}.$$

Second, observe that $w_k(i) = \mathrm{Re}\{v_k(i)\}$ where $v_k(i) = e^{j\frac{\pi i}{2k}}$ denotes the recursively computable complex-valued window

$$v_k(i-1) = \gamma_k v_k(i), \quad \gamma_k = e^{-j\frac{\pi}{2k}}. \quad (27)$$

Exploiting (26) and (27), one can compute the quantities $\mathbf{P}_{m|k}(t)$ and $\mathbf{p}_{m|k}(t)$ recursively using the following algorithms

$$\mathbf{P}_{m|k}(t) = \mathrm{Re}\{\mathbf{R}_{m|k}(t)\}, \quad \mathbf{p}_{m|k}(t) = \mathrm{Re}\{\mathbf{r}_{m|k}(t)\}$$

$$\begin{aligned} \mathbf{R}_{m|k}(t) &= \sum_{i=-k}^{k} v_k(i)\boldsymbol{\psi}_{m|k}(t,i)\boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,i) \\ &= \gamma_k[\mathbf{I}_n \otimes \boldsymbol{\Gamma}_{m|k}]\Big\{\mathbf{R}_{m|k}(t-1) \\ &- v_k(-k)\mathbf{A}(t-k-1) \otimes \mathbf{B}_{m|k}(-k)\Big\}[\mathbf{I}_n \otimes \boldsymbol{\Gamma}^{\mathrm{T}}_{m|k}] \\ &+ v_k(k)\mathbf{A}(t+k) \otimes \mathbf{B}_{m|k}(k) \end{aligned} \quad (28)$$

$$\begin{aligned}\mathbf{r}_{m|k}(t) &= \sum_{i=-k}^{k} v_k(i)y(t+i)\boldsymbol{\psi}_{m|k}(t,i) \\ &= \gamma_k[\mathbf{I}_n \otimes \boldsymbol{\Gamma}_{m|k}]\Big\{\mathbf{r}_{m|k}(t-1) \\ &- v_k(-k)\mathbf{c}(t-k-1)\otimes \mathbf{f}_{m|k}(-k)\Big\} \\ &+ v_k(k)\mathbf{c}(t+k)\otimes \mathbf{f}_{m|k}(k)\end{aligned} \tag{29}$$

where $\mathbf{A}(t) = \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)$, $\mathbf{c}(t) = y(t)\boldsymbol{\varphi}(t)$ and $\mathbf{B}_{m|k}(i) = \mathbf{f}_{m|k}(i)\mathbf{f}_{m|k}^{\mathrm{T}}(i)$, $i \in I_k$.

Another bell-shaped window which in combination with the polynomial or harmonic basis allows for recursive computation of $\mathbf{P}_{m|k}(t)$ and $\mathbf{p}_{m|k}(t)$ (albeit at a higher computational cost) is the classical Hann, i.e., raised cosine, window $w_k(i) = 0.5\left[1+\cos\frac{\pi i}{k}\right] = 0.5\left\{1+\mathrm{Re}[e^{j\pi i/k}]\right\}$. Finally, recursive computability is guaranteed if the adopted window is rectangular $w_k(i) = 1, i \in I_k$, i.e., if no weighting is applied.

Remark

The sliding window substract-add recursive algorithms, such as the ones presented above, are not exponentially stable, but only marginally stable (since γ_k and all eigenvalues of $\boldsymbol{\Gamma}_k$ are located on the unit circle in the complex plane), and hence they diverge at a slow (linear) rate when the number of time steps becomes very large. This effect is due to unbounded accumulation of roundoff errors. For this reason the derived recursive algorithms should be periodically reset by direct (nonrecursive) computation of the quantities $\mathbf{P}_{m|k}(t)$ and $\mathbf{p}_{m|k}(t)$.

3.5 Comparison with the interval basis function approach

In this section we will show why the proposed point estimators $\widehat{\boldsymbol{\theta}}_{m|k}(t)$ yield better results than the classical interval ones $\boldsymbol{\theta}^*_{m|k}(t+l|t), l \in I_k$, which approximate parameter trajectory in the entire analysis interval $T_k(t) = [t-k, t+k]$.

First of all, note that the interval estimates can be obtained using the formulas

$$\begin{aligned}\boldsymbol{\theta}^*_{m|k}(t+l|t) &= \mathbf{F}^*_{m|k}(l)\widehat{\boldsymbol{\alpha}}_{m|k}(t) \\ \mathbf{F}^*_{m|k}(l) &= \mathbf{I}_n \otimes \mathbf{f}^{\mathrm{T}}_{m|k}(l) \\ l &\in I_k.\end{aligned} \tag{30}$$

In the case of BF estimation with non-preferential, i.e., rectangular windowing (which is a natural choice in such a case), one can derive the following relationships

$$\mathrm{cov}[\boldsymbol{\theta}^*_{m|k}(t+l|t)] \cong \frac{\rho\boldsymbol{\Phi}^{-1}}{N^*_{m|k}(l)}\ ,\ N^*_{m|k}(l) = [\widetilde{\mathbf{f}}^{\mathrm{T}}_{m|k}(l)\widetilde{\mathbf{f}}_{m|k}(l)]^{-1}$$

$$\mathrm{E}_{\Omega_k(t)}[\boldsymbol{\theta}^*_{m|k}(t+l|t)] \cong \sum_{i=-k}^{k} h^*_{m|k}(i,l)\boldsymbol{\theta}(t+i) \tag{31}$$

$$h^*_{m|k}(i,l) = \widetilde{\mathbf{f}}^{\mathrm{T}}_{m|k}(l)\widetilde{\mathbf{f}}_{m|k}(i)$$

which, under the same assumptions as those used before, parallel the relationships (19) and (21), respectively. Note that $\boldsymbol{\theta}^*_{m|k}(t|t) = \widehat{\boldsymbol{\theta}}_{m|k}(t)$, $\mathbf{F}^*_{m|k}(0) = \mathbf{F}_{m|k}$, $N^*_{m|k}(0) = N_{m|k}$ and $h^*_{m|k}(i,0) = h_{m|k}(i)$.

Similarly as in the case of point estimators, one can define parameter matching characteristic at the location l in the form

$$E^*_{m|k}(\omega, l) = |1 - H^*_{m|k}(\omega, l)|^2$$

where

$$H^*_{m|k}(\omega, l) = \sum_{i=-k}^{k} h^*_{m|k}(i,l)e^{-j\omega i}$$

denotes the "frequency response" of $\boldsymbol{\theta}^*_{m|k}(t+l|t)$.

Fig. 3 shows dependence of the equivalent number of observations $N^*_{m|k}(l)$ and parameter matching characteristic $E^*_{m|k}(\omega, l)$ of the estimator $\boldsymbol{\theta}^*_{m|k}(t+l|t)$ on the location parameter l. The plots were obtained for polynomial bases of order 3 and 5. Note that the estimation bandwidth monotonically decreases for growing $|l|$, and that the same happens, although not in a monotonic way, in the case of equivalent number of observations. The straightforward consequence of these facts is significant increase of both bias and variance components of the mean squared parameter matching errors at both ends of the analysis interval – the hardly surprising effect considering the fact that only one-sided information (only the "past" samples or only the "future" ones) is available at both ends of $T_k(t)$. In contrast with this, the point estimators $\widehat{\boldsymbol{\theta}}_{m|k}(t) = \boldsymbol{\theta}^*_{m|k}(t+l|t)|_{l=0}$, which benefit from the access to both "past" and "future" data samples, provide the best estimation accuracy.

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

The mean squared parameter matching error can be decomposed into the bias component $\sigma^b_{m|k}(t)$ and variance component $\sigma^v_{m|k}(t)$

$$\mathrm{E}[\|\ \widehat{\boldsymbol{\theta}}_{m|k}(t) - \boldsymbol{\theta}(t)\ \|^2] = \sigma^b_{m|k}(t) + \sigma^v_{m|k}(t) \tag{32}$$

where $\sigma^b_{m|k}(t) = \|\ \bar{\boldsymbol{\theta}}_{m|k}(t) - \boldsymbol{\theta}(t)\ \|^2$ and $\sigma^v_{m|k}(t) = \mathrm{tr}\{\mathrm{cov}[\widehat{\boldsymbol{\theta}}_{m|k}(t)]\}$. Under assumptions (A1), (A2) and (A3*) the lower bound on $\mathrm{cov}[\widehat{\boldsymbol{\theta}}_{m|k}(t)]$ is set by (19).

As shown in Sections 3.2 and 3.3, when the number of basis functions is fixed and the window size is increased, the variance component [see (19)] decreases and the bias

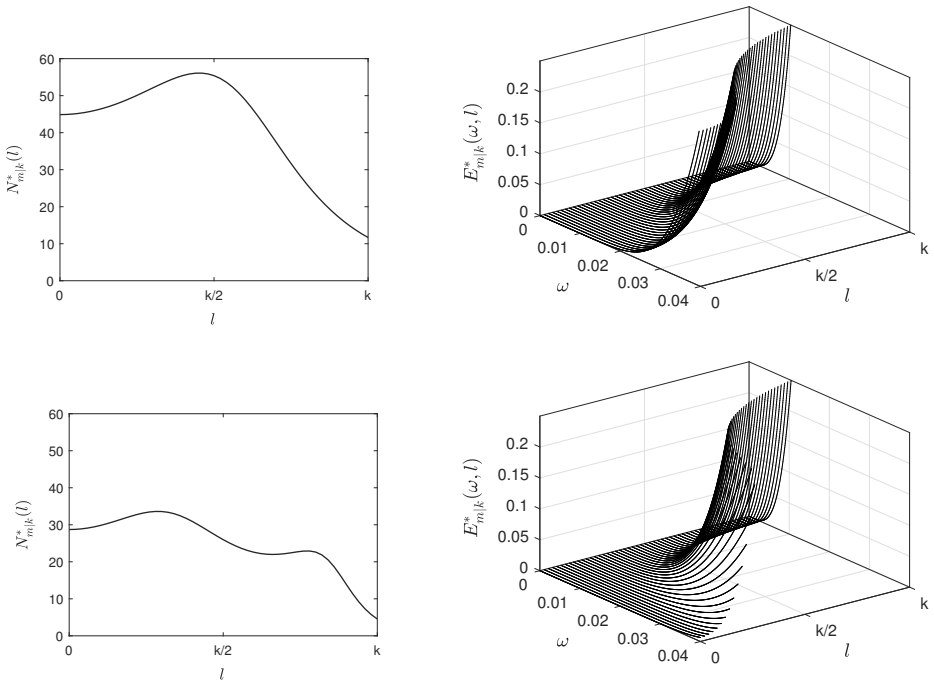

Fig. 3. Dependence of the equivalent number of observations $N^*_{m|k}(l)$ and parameter matching characteristic $E^*_{m|k}(\omega, l)$ on the location l within the rectangular window of width $l_k = 2k + 1 = 101$: a) polynomial basis of order $m = 3$ (two upper figures), b) polynomial basis of order $m = 5$ (two lower figures). Since the plots are symmetric with respect to $l = 0$ they are shown only for $l \in [0, k]$.

component increases. Similarly, for a fixed window size increasing the number of basis functions allows one to reduce the bias component but, at the same time, the variance component increases. It is therefore clear that the values of m and k should be chosen so as to trade-off the bias and variance components of the mean squared matching error. 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 procedures that allow for adaptive scheduling of m and k.

The proposed solution is based on parallel estimation. We will consider 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. At each time instant only one of the competing algorithms is selected, i.e., the estimated parameter and variance trajectories have the form $\widehat{\boldsymbol{\theta}}_{\widehat{m}(t)|\widehat{k}(t)}(t)$ and $\widehat{\rho}_{\widehat{m}(t)|\widehat{k}(t)}(t)$ where

$$\{\widehat{m}(t), \widehat{k}(t)\} = \arg \min_{\substack{m \in \mathcal{M} \\ k \in \mathcal{K}}} J_{m|k}(t) \tag{33}$$

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

4.1 Final prediction error based approach

Denote by $\Omega'_k(t) = \{\boldsymbol{\varphi}'(t+i), e'(t+i), i \in I_k\}$ another realization of the input-output data, independent of the set $\Omega_k(t) = \{\boldsymbol{\varphi}(t+i), e(t+i), i \in I_k\}$ used for identification purposes. Following (Akaike, 1970), as a local performance measure we can use the following quantity

$$\delta_{m|k}(t) = \mathrm{E}_{\Omega_k(t), \Omega'_k(t)} \left\{ \left[y'(t) - [\boldsymbol{\varphi}'(t)]^{\mathrm{T}} \widetilde{\boldsymbol{\theta}}_{m|k}(t) \right]^2 \right\} \tag{34}$$

called by Akaike final prediction error (FPE). Final prediction error quantifies prediction accuracy when the model is verified using an independent data set.

Under assumptions (A1)-(A3) it holds that (see Appendix)

$$\delta_{m|k}(t) = \rho \left(1 + \frac{n}{N_{m|k}} \right) \tag{35}$$

$$\mathrm{E}[\widehat{\rho}_{m|k}(t)] \cong \rho \left(1 - \frac{n}{M_{m|k}} \right) \tag{36}$$

where

$$M_{m|k} = \frac{L_k}{\sum_{i=-k}^{k} w_k^2(i) \widetilde{\mathbf{f}}^{\mathrm{T}}_{m|k}(i) \widetilde{\mathbf{f}}_{m|k}(i)} \,. \tag{37}$$

Combining (35) and (36), one arrives at the following estimate of the final prediction error

$$\widehat{\delta}_{m|k}(t) = \frac{1 + \frac{n}{N_{m|k}}}{1 - \frac{n}{M_{m|k}}} \, \widehat{\rho}_{m|k}(t) \tag{38}$$

The local measure of fit minimized in (33) can be set to

$$J_{m|k}(t) = \widehat{\delta}_{m|k}(t). \tag{39}$$

4.2 Cross-validation approach

Denote by $\widehat{\boldsymbol{\theta}}^{\circ}_{m|k}(t)$ the holey LBF estimator of $\boldsymbol{\theta}(t)$, i.e., the one that eliminates from the estimation process the measurement $y(t)$ collected at the instant t

$$\widehat{\boldsymbol{\theta}}^{\circ}_{m|k}(t) = \mathbf{F}_{m|k}\widehat{\boldsymbol{\alpha}}^{\circ}_{m|k}(t) \tag{40}$$

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}^{\circ}_{m|k}(t) &= \arg\min_{\boldsymbol{\alpha}_{m|k}} \sum_{\substack{i=-k \\ i\neq 0}}^{k} [y(t+i) - \boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,i)\boldsymbol{\alpha}_{m|k}]^2 \\ &= [\mathbf{P}^{\circ}_{m|k}(t)]^{-1}\mathbf{p}^{\circ}_{m|k}(t)\end{aligned} \tag{41}$$

where, since $w_k(0) = 1$,

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

The interpolation error associated with (40) takes the form $\varepsilon^{\circ}_{m|k}(t) = y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}^{\circ}_{m|k}(t) = y(t) - \boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,0)\widehat{\boldsymbol{\alpha}}^{\circ}_{m|k}(t)$. Selection of m and k can be based on minimization of the localized sum of squared interpolation errors, i.e., one can adopt

$$J_{m|k}(t) = \sum_{l=-L}^{L} [\varepsilon^{\circ}_{m|k}(t+l)]^2 \tag{42}$$

where L (typically chosen from the interval [20,30]) determines the size of the local decision window.

The computational complexity of the cross-validation (CV) approach described above can be significantly reduced by noting that

$$\varepsilon^{\circ}_{m|k}(t) = \frac{\varepsilon_{m|k}(t)}{1 - q_{m|k}(t)} \tag{43}$$

where $\varepsilon_{m|k}(t)$ denotes residual error $\varepsilon_{m|k}(t) = y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}_{m|k}(t) = y(t) - \boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,0)\widehat{\boldsymbol{\alpha}}_{m|k}(t)$, and

$$q_{m|k}(t) = \boldsymbol{\psi}^{\mathrm{T}}_{m|k}(t,0)\mathbf{P}^{-1}_{m|k}(t)\boldsymbol{\psi}_{m|k}(t,0). \tag{44}$$

According to (43), which can be easily derived using the well-known matrix inversion lemma (Sayed, 2003), the cross-validation statistic (42) can be evaluated without actually implementing the holey estimation scheme.

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

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

5 Simulation results

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

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

excited by the autoregressive input signal $u(t) = 0.8u(t-1) + v(t)$, $\mathrm{var}[v(t)] = 1$, where $\{v(t)\}$ denotes white noise independent of $\{e(t)\}$. System parameters were modeled as cosinusoidal linear chirps with instantaneous angular frequencies increasing from zero to the prescribed values over the simulation interval T (see Fig. 4)

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

Three simulation scenarios were checked, corresponding to three speeds of parameter variation (SoV): slow, medium and fast – the corresponding settings are shown in Table 1. Finally, two signal to noise ratios (SNR = $\mathrm{E}\{[\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t)]^2\}/\sigma_e^2$) were considered: 25 dB ($\sigma_e^2 = 0.01$) and 15 dB ($\sigma_e^2 = 0.1$).

Table 2 shows comparison of the mean squared parameter estimation errors obtained for 9 LBF estimators (polynomial basis, cosinusoidal window) corresponding to different choices of design parameters k (50, 100, 200) and m (1, 3, 5), and for the proposed adaptive estimation schemes with FPE-based and CV-based model selection. Note that the CV criterion yields slightly (but consistently) better results than the FPE criterion. In all cases discussed above the proposed adaptive schemes yield either better or comparable results to those provided by the best LBF algorithms with fixed settings. The advantages of the adaptive tuning strategies are particularly visible in the presence of fast parameter changes. Fig. 4 shows comparison of the true parameter trajectories with those obtained for a fixed choice of design parameters ($k = 200$, $m = 3$, SNR=15 dB) and for adaptively selected design parameters (CV).

6 Conclusion

The problem of identification of a nonstationary stochastic process was considered and solved using the time-localized variant of the basis function approach. The proposed local basis function (LBF) estimators generate a sequence of point estimates of system parameters assuming that parameter trajectories can be locally approximated by linear

Table 2
Mean squared parameter estimation errors obtained for 9 LBF estimators corresponding to different choices of design parameters k (50, 100, 200) and m (1, 3, 5), and for the proposed adaptive estimation schemes with FPE-based and CV-based model selection. The averages were computed for 100 process realizations, 3 speeds of parameter variation (SoV) and 2 signal-to-noise ratios (SNR).

SNR	SoV	slow			medium			fast		
	$k\backslash m$	1	3	5	1	3	5	1	3	5
15 dB	50	6.0E-03	5.5E-03	8.6E-03	2.8E-02	5.7E-03	8.6E-03	1.9E-01	1.6E-02	8.9E-03
	100	2.2E-02	2.7E-03	4.0E-03	1.7E-01	1.1E-02	4.3E-03	5.9E-01	2.5E-01	5.4E-02
	200	1.6E-01	9.1E-03	2.1E-03	5.8E-01	2.4E-01	4.9E-02	8.0E-01	6.6E-01	5.1E-01
	FPE		4.2E-03			5.3E-03			7.1E-03	
	CV		3.7E-03			4.7E-03			6.9E-03	
25 dB	50	3.7E-03	5.5E-04	8.6E-04	2.5E-02	7.1E-04	8.6E-04	1.9E-01	1.1E-02	1.1E-03
	100	2.1E-02	3.7E-04	4.1E-04	1.7E-01	9.0E-03	5.8E-04	5.9E-01	2.5E-01	5.1E-02
	200	1.6E-01	7.9E-03	3.4E-04	5.8E-01	2.4E-01	4.7E-02	8.0E-01	6.6E-01	5.1E-01
	FPE		4.5E-04			6.1E-04			9.6E-04	
	CV		4.0E-04			5.5E-04			9.4E-04	

‘

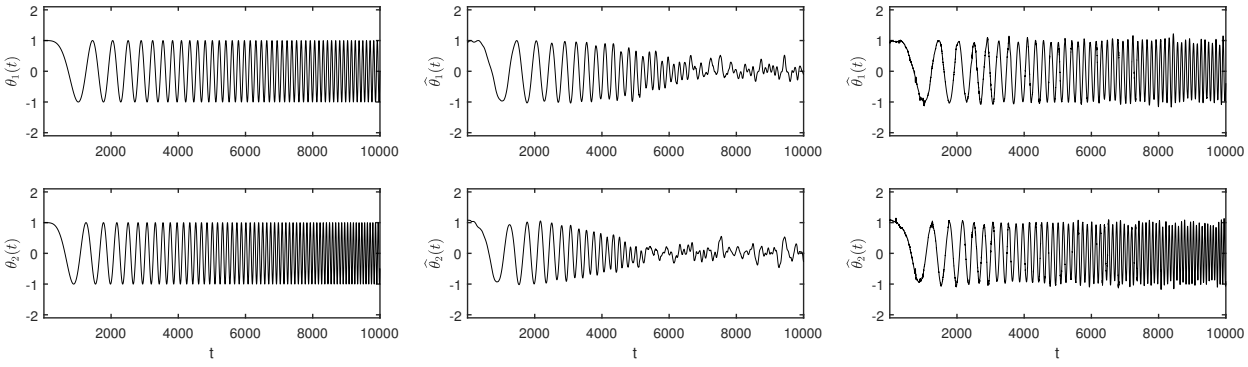

Fig. 4. Evolution of system parameters (a) and their estimates obtained for a fixed choice of design parameters (b) and for adaptively selected design parameters (c).

combinations of a certain number of known functions of time (basis functions). It was shown that two important design 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 two selection criteria are based on the modified Akaike's final prediction error statistic and on leave-one-out cross-validation approach, respectively. The resulting adaptive algorithms are computationally attractive (recursively computable) and yield 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

Akaike H. (1970). Statistical predictor identification. *Ann. Inst. Statist. Math.*, (22), 203–217.

Billings, S. A. (2013). *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*. Wiley, 2013.

Dahlhaus, R. (2012). Locally stationary processes. *Handbook Statist.*, (25), 1–37.

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.

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

Liu, W., Principe, J.C. & Haykin, S. (2011) *Kernel Adaptive Filtering: A Comprehensive Introduction*. Wiley, 2011.

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

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. (1988). Functional series modeling approach to

identification of nonstationary stochastic systems. *IEEE Trans. Automat. Contr.*, (33), 955–961.

Niedźwiecki M. & Gackowski S. (2013). New approach to noncausal identification of nonstationary FIR systems subject to both smooth and abrupt parameter changes. *IEEE Trans. Automat. Contr.*, (58), 1847–1853.

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

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.

Sayed, A. H. (2003). *Fundamentals of Adaptive Filtering*. Wiley, 2003.

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

Tsatsanis, M. K., & Giannakis, G. B. (1996) *Modeling and equalization of rapidly fading channels*. 10, 159-176.

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 (35) and (36)]

In order to derive (35), note that $y'(t) = [\boldsymbol{\varphi}'(t)]^{\mathrm{T}}\boldsymbol{\theta}(t) + e'(t)$, $\mathrm{E}\{[e'(t)]^2\} = \rho$. Using these relationships and exploiting mutual independence of $\Omega_k(t)$ and $\Omega'_k(t)$, one obtains under assumptions (A1)-(A3)

$$\begin{aligned}
&\delta_{m|k}(t) = \mathrm{E}_{\Omega_k(t),\Omega'_k(t)}\left\{\left[e'(t) - [\boldsymbol{\varphi}'(t)]^{\mathrm{T}}\Delta\widetilde{\boldsymbol{\theta}}_{m|k}(t)\right]^2\right\} \\
&= \rho + \mathrm{E}_{\Omega_k(t),\Omega'_k(t)}\left\{[\boldsymbol{\varphi}'(t)]^{\mathrm{T}}\Delta\widetilde{\boldsymbol{\theta}}_{m|k}(t)\Delta\widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{m|k}(t)\boldsymbol{\varphi}'(t)\right\} = \rho \\
&+ \mathrm{tr}\left\{\mathrm{E}_{\Omega'_k(t)}\left[\boldsymbol{\varphi}'(t)[\boldsymbol{\varphi}'(t)]^{\mathrm{T}}\right]\mathrm{E}_{\Omega_k(t)}\left[\Delta\widetilde{\boldsymbol{\theta}}_{m|k}(t)\Delta\widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{m|k}(t)\right]\right\} \\
&\cong \rho\left(1 + \frac{1}{N_{m|k}}\mathrm{tr}\{\mathbf{I}_n\}\right) = \rho\left(1 + \frac{n}{N_{m|k}}\right)
\end{aligned}$$

where the third transition follows from (19).

To arrive at (36) note that

$$\begin{aligned}
&\widehat{\rho}_{m|k}(t) = \widetilde{\rho}_{m|k}(t) \\
&= \frac{1}{L_k}\sum_{i=-k}^{k} w_k(i)\left[e(t+i) - \widetilde{\boldsymbol{\psi}}^{\mathrm{T}}_{m|k}(t,i)\Delta\widetilde{\boldsymbol{\alpha}}_{m|k}(t)\right]^2 \\
&= \frac{1}{L_k}z_k(t) - \frac{1}{L_k}\boldsymbol{\xi}^{\mathrm{T}}_{m|k}(t)\widetilde{\mathbf{P}}^{-1}_{m|k}(t)\boldsymbol{\xi}_{m|k}(t)
\end{aligned}$$

where $z_k(t) = \sum_{i=-k}^{k} w_k(i)e^2(t+i)$. Under assumptions (A1)-(A3) it holds that $\mathrm{E}[z_k(t)] = L_k\rho$ and

$$\begin{aligned}
&\mathrm{E}\left[\boldsymbol{\xi}^{\mathrm{T}}_{m|k}(t)\widetilde{\mathbf{P}}^{-1}_{m|k}(t)\boldsymbol{\xi}_{m|k}(t)\right] \cong \mathrm{E}\left[\boldsymbol{\xi}^{\mathrm{T}}_{m|k}(t)\bar{\mathbf{P}}^{-1}_{m}\boldsymbol{\xi}_{m|k}(t)\right] \\
&= \mathrm{tr}\left\{\bar{\mathbf{P}}^{-1}_{m}\mathrm{E}\left[\boldsymbol{\xi}_{m|k}(t)\boldsymbol{\xi}^{\mathrm{T}}_{m|k}(t)\right]\right\} \\
&= \rho\ \mathrm{tr}\left\{[\boldsymbol{\Phi}^{-1}\otimes\mathbf{I}_m][\boldsymbol{\Phi}\otimes\mathbf{W}_{m|k}]\right\} = \rho\ \mathrm{tr}\left\{\mathbf{I}_n\otimes\mathbf{W}_{m|k}\right\} \\
&= n\rho\sum_{i=-k}^{k} w^2_k(i)\widetilde{\mathbf{f}}^{\mathrm{T}}_{m|k}(i)\widetilde{\mathbf{f}}_{m|k}(i)
\end{aligned}$$

where the last transition follows from (**??**) and from the fact that $\mathrm{tr}\{\mathbf{A}\otimes\mathbf{B}\} = \mathrm{tr}\{\mathbf{A}\}\mathrm{tr}\{\mathbf{B}\}$. Combining all earlier results, one obtains (36).

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.