Regularized Local Basis Function Approach to Identification of Nonstationary Processes

Artur Gańcza, *Student Member, IEEE*, Maciej Niedźwiecki, *Senior Member, IEEE*, and Marcin Ciołek, *Member, IEEE*

***Abstract*—The problem of identification of nonstationary stochastic processes (systems or signals) 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 regularized 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, the paper addresses and solves all major problems associated with implementation of the new class of estimators, such as optimization of the regularization matrix, adaptive selection of the number of basis functions and the width of the local analysis interval, and reduction of complexity of the computational algorithms.**

***Index Terms*—System identification, parameter estimation, time-varying systems, adaptive estimation.**

I. Introduction

LINEAR time-varying models of nonstationary processes (signals or systems) are the basis of many real-life applications in various disciplines such as telecommunications [1], biomedicine [2], [3], geophysics [4], [5] and control science [6], to name just a few. In all such cases accurate estimation/tracking of time-varying parameters of the underlying models is of primary importance. When process parameters vary slowly, their estimation can be carried out using the time-localized versions of classical identification algorithms. However, as the speed of parameter changes increases, such an unstructured local estimation approach fails unless some explicit models of parameter variation are incorporated into the identification process. Such models, often called hypermodels, can be either stochastic or deterministic. In the first case the parameter tracking task is usually reformulated as a problem of filtering/smoothing in the state space, and its solution is obtained using an appropriately designed Kalman filters/smoothers [7]–[10]. In the deterministic setup, parameter trajectories are approximated by linear combinations of some known functions of time, usually referred to as basis functions [11]–[18].

Quite recently a new estimation paradigm for identification of linear time-varying processes, based on the concept of preestimation, was proposed [19]. Preestimates are raw estimates of process parameters – unbiased but with a very large variability. For this reason, to obtain statistically meaningful results, preestimates must be further processed (denoised). The resulting two-stage identification procedure compares favorably with the existing solutions to the problem of parameter tracking as it offers, without compromising good tracking performance, significantly lower computational complexity and increased numerical robustness. The current paper aims to further improve, using the regularization technique, the solutions described in [18] and [19].

Regularization is a successful and long standing concept in machine learning and system identification. The idea is to add to the minimized cost function a term (regularizer) which reduces a norm of the solution. Regularization can serve different purposes. The L_2 regularization, such as ridge regression or Tikhonov regularization, was originally introduced as a way of solving ill-posed or numerically ill-conditioned inverse problems [20], [21]. In the context of least squares estimation, the difficulty arises when the inverted regression matrix is poorly conditioned (almost singular) [22], [23].

However, from the estimation perspective, regularization has more to offer [24], [25]. First, it can improve the bias-variance trade-off which decides upon accuracy of the identified model. More specifically, by choosing the regularization term in a judicious way, one can decrease the variance of the estimates at the expense of increasing their bias, while ensuring that the overall mean square error is reduced [26], [27]. Secondly, regularization allows one to include in the formulation of the identification problem some expected, or desired, properties of the solution. This way of looking at the estimation task was pioneered by Whittaker [28], [29]. Whittaker considered the problem of approximation of an unknown function based on noisy measurements and suggested to add to the minimized quadratic cost function the L_2 smoothness enhancing regularizer. Later on, this idea was extended to signals [8] and systems [30]. Quite recently a considerable attention was payed to the problem of estimation of an impulse response of a time-invariant linear system [31]. In this case regularization may be used to incorporate into the estimation process some prior knowledge about a "typical" shape of an impulse response of an asymptotically stable dynamical system ("exponentially decaying and smooth").

Manuscript received May 29, 2020; revised January 11, 2021; accepted February 11, 2021. Date of publication February 25, 2021; date of current version March 22, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. S. Kar. 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. *(Corresponding author: Artur Gancza.)*

The authors are with the Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control, Gdańsk University of Technology, 80-233 Gdańsk, Poland (e-mail: artgancz@student.pg.edu.pl; maciekn@eti.pg.edu.pl; marciole@pg.edu.pl).

Digital Object Identifier 10.1109/TSP.2021.3062168

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

Finally, in some applications the main purpose of regularization is to prevent the identified models from overfitting, i.e., from including in the model more parameters than can be justified by the data. To this end the L_1 regularization is more effective than the L_2 one [32], leading to such well-known techniques as LASSO (least absolute shrinkage and selection operator) in statistics [33], or basis pursuit in signal processing [34].

Whatever is the purpose of regularization, its success heavily depends on the choice of regularization hyperparameters, such as gains, coefficients of the regularization matrix etc. Optimization of these hyperparameters can be performed using techniques such as cross-validation, Stein's unbiased risk estimator (SURE) or stochastic embedding [9], [35], [36]. It seldom leads to closed-form solutions – a numerical search has to be used instead.

The paper presents a regularized version of the local basis function approach described in [18]. To the best of our knowledge this is the first attempt to incorporate the regularization technique into identification of nonstationary processes in the deterministic hypermodel setup. First, extending the result established in [31] for time-invariant FIR systems, we derive closed-form expressions allowing one to optimize the coefficients of a diagonal regularization matrix. Since the optimal settings depend on true (unknown) parameters of the identified system, direct application of this result is not possible. Instead, we propose a two-stage approach. At the first stage identification is carried out using the standard basis function approach proposed in [18]. Then, at the second identification stage, the regularized estimates are evaluated after replacing the unknown regularization hyperparameters with their estimates obtained at the first stage. In the second part of the paper, which is based on the concept of parameter preestimation [19], the computationally fast and numerically robust version of the regularized basis function scheme is proposed. It is shown that in spite of its simplicity, the fast regularized local basis function algorithm has very good parameter tracking capability, comparable with that of the original scheme. It is also shown that regularization improves accuracy of parameter estimates and that the proposed approach compares favorably with the state-of-the-art wavelet-based approach described in [16], [17].

All algorithms developed in this paper provide a sequence of local point estimates of process parameters, each corresponding to a particular time instant t, namely, the estimation 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 resulting estimates are noncausal (the estimation results depend on both "past" and "future" observations relative to t) they cannot be used in real-time applications such as adaptive prediction or adaptive control. However, many almost real-time applications exist, such as time-frequency signal analysis [37], channel equalization [1] or self-interference mitigation [38], that are not time-critical in the sense that the model-based decisions can be delayed by a certain number of sampling intervals. Since the well-tuned noncausal estimation algorithms provide a better bias-variance trade-off than their comparable causal counterparts, the achievable parameter tracking performance is usually considerably better.

II. Estimation Scheme

A. 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 n-dimensional vector of unknown time-varying parameters, $\boldsymbol{\varphi}(t) = [\varphi_1(t), \ldots, \varphi_n(t)]^{\mathrm{T}}$ denotes regression vector and $e(t)$ denotes white noise with variance σ_e^2. The two practically important cases include finite impulse response (FIR) systems where

$$\boldsymbol{\varphi}(t) = [u(t), \ldots, u(t-n+1)]^{\mathrm{T}} \tag{2}$$

and $\{u(t)\}$ denotes the observable input signal, and autoregressive processes where

$$\boldsymbol{\varphi}(t) = [y(t-1), \ldots, y(t-n)]^{\mathrm{T}}. \tag{3}$$

Further assumptions about the sequences $\{\boldsymbol{\varphi}(t)\}$, $\{\boldsymbol{\theta}(t)\}$ and $\{e(t)\}$ will be given later.

In this article we will apply the localized approach combined with the basis function method, which assumes that inside the local analysis interval, $T_k(t) = [t-k, t+k]$, centered at the time instant t, the true parameter trajectories can be modeled as a linear combination of m linearly independent basis functions

$$\mathcal{F}_{m|k} = \{f_{1|k}(i), \ldots, f_{m|k}(i), i \in I_k\}$$

namely

$$\mathcal{H}_{m|k} \ : \ \theta_j(t+i) = \sum_{l=1}^{m} b_{jl;m|k} f_{l|k}(i) = \boldsymbol{\beta}_{j;m|k}^{\mathrm{T}} \mathbf{f}_{m|k}(i)$$
$$j = 1, \ldots, n, \ i \in I_k = [-k, k], \tag{4}$$

where

$$\boldsymbol{\beta}_{j;m|k} = [b_{j1;m|k}, \ldots, b_{jm;m|k}]^{\mathrm{T}}$$
$$\mathbf{f}_{m|k}(i) = [f_{1|k}(i), \ldots, f_{m|k}(i)]^{\mathrm{T}}.$$

For a given set of basis functions, the hypermodel $\mathcal{H}_{m|k}$ has two user-dependent degrees of freedom: the number of basis functions m and the width of the local analysis window $K = 2k+1$. The problem of selection of m and k will be discussed later in Section IV.

Let

$$\boldsymbol{\alpha}_{m|k} = [\boldsymbol{\beta}_{1;m|k}^{\mathrm{T}}, \ldots, \boldsymbol{\beta}_{n;m|k}^{\mathrm{T}}]^{\mathrm{T}}$$

be the $nm \times 1$ vector made up of all basis function coefficients, and denote by

$$\boldsymbol{\psi}_{m|k}(t,i) = \boldsymbol{\varphi}(t+i) \otimes \mathbf{f}_{m|k}(i) \tag{5}$$

the $nm \times 1$ generalized regression vector, where $\otimes$ stands for the Kronecker product.

Using the expansion (4), the time varying model (1) of order n can be locally, i.e., inside the analysis interval $T_k(t)$, expressed

MOST WIEDZY Downloaded from mostwiedzy.pl

as a nm-dimensional time invariant model

$$y(t+i) = \boldsymbol{\psi}_{m|k}^{\mathrm{T}}(t,i)\boldsymbol{\alpha}_{m|k} + e(t+i), \quad i \in I_k. \tag{6}$$

Based on (4) and (6), the parameter vector $\boldsymbol{\theta}(t)$ can be written down in the form

$$\boldsymbol{\theta}(t)|_{\mathcal{H}_{m|k}} = \boldsymbol{\theta}_{m|k}(t) = \mathbf{F}_{m|k}\boldsymbol{\alpha}_{m|k}$$
$$\mathbf{F}_{m|k} = \mathbf{I}_n \otimes \mathbf{f}_{m|k}^{\mathrm{T}}(0). \tag{7}$$

B. Identification Procedure

Our identification scheme is based on the regularized weighted least squares (WLS) method

$$\widehat{\boldsymbol{\alpha}}_{m|k}^{\mathrm{RLBF}}(t) = \arg\min_{\boldsymbol{\alpha}_{m|k}} \times \left\{ \sum_{i=-k}^{k} w_k(i) \quad \left[y(t+i) - \boldsymbol{\psi}_{m|k}^{\mathrm{T}}(t,i)\boldsymbol{\alpha}_{m|k}\right]^2 + \boldsymbol{\alpha}_{m|k}^{\mathrm{T}}\boldsymbol{\Lambda}_{m|k}\boldsymbol{\alpha}_{m|k} \right\}, \tag{8}$$

where $\boldsymbol{\Lambda}_{m|k} > \mathbf{O}$ denotes the $nm \times nm$ positive definite regularization matrix and $\{w_k(i), \quad i \in I_k\}$, $w_k(0) = 1$, is a bell-shaped, nonnegative and symmetric window used for localization purposes. We will further assume that $w_k(i) = w^0(\frac{i}{k})$ is obtained by sampling the continuous-time window generating function $w^0(s)$, $s \in [-1, 1]$. One of the possible choices of $w^0(s)$, attractive from the computational viewpoint, is the cosinusoidal window

$$w^0(s) = \cos\frac{\pi s}{2}. \tag{9}$$

The resulting local estimate of $\boldsymbol{\alpha}_{m|k}$ will be referred to as regularized local basis function (RLBF) estimate. Selection of the form (structure) of the regularization matrix $\boldsymbol{\Lambda}_{m|k}$ and optimization of its coefficients will be discussed in Section III B.

Without any loss of generality we will assume that the basis $\mathcal{F}_{m|k}$ is w-orthonormal, namely

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

Furthermore, to allow asymptotic reasoning, we will assume that $\mathcal{F}_{m|k}$ is obtained by means of orthonormalization of the basis

$$\mathcal{G}_{m|k} = \{g_{l|k}(i),\ l = 1, \ldots, m,\ i \in I_k\},$$

where

$$g_{l|k}(i) = g_l^0\left(\frac{i}{k}\right)$$

and $g_l^0(s),\ s \in [-1, 1]$, $l = 1, \ldots, m$, denote continuous-time, square integrable basis generating functions defined on the interval $[-1, 1]$.

When no prior knowledge about the nature of parameter variation is available, a reasonable choice of the basis would be to select functions with high generalization capacity. For example one could use the basis made up of powers of time (Taylor series approximation)

$$g_l^0(s) = s^{l-1}, \quad l = 1, \ldots, m \tag{11}$$

or cosinusoidal functions (Fourier series approximation)

$$g_l^0(s) = \cos\left[\frac{\pi s(l-1)}{2}\right], \quad l = 1, \ldots, m. \tag{12}$$

It is straightforward to check that the solution to (8) is given by

$$\widehat{\boldsymbol{\alpha}}_{m|k}^{\mathrm{RLBF}}(t) = \mathbf{R}_{m|k}^{-1}(t)\mathbf{p}_{m|k}(t) \tag{13}$$

where

$$\mathbf{R}_{m|k}(t) = \mathbf{P}_{m|k}(t) + \boldsymbol{\Lambda}_{m|k}$$
$$\mathbf{P}_{m|k}(t) = \sum_{i=-k}^{k} w_k(i)\boldsymbol{\psi}_{m|k}(t,i)\boldsymbol{\psi}_{m|k}^{\mathrm{T}}(t,i)$$
$$\mathbf{p}_{m|k}(t) = \sum_{i=-k}^{k} w_k(i)y(t+i)\boldsymbol{\psi}_{m|k}(t,i). \tag{14}$$

According to (7) the local (pointwise) estimate of the process parameter vector $\boldsymbol{\theta}(t)$ can be obtained from

$$\widehat{\boldsymbol{\theta}}_{m|k}^{\mathrm{RLBF}}(t) = \mathbf{F}_{m|k}\widehat{\boldsymbol{\alpha}}_{m|k}^{\mathrm{RLBF}}(t). \tag{15}$$

The described procedure is called local because estimation is repeated for each position of the sliding analysis window $T_k(t)$.

The estimate (13) is a straightforward generalization of the local basis function (LBF) estimate introduced recently in [18]

$$\widehat{\boldsymbol{\alpha}}_{m|k}^{\mathrm{LBF}}(t) = \widehat{\boldsymbol{\alpha}}_{m|k}^{\mathrm{RLBF}}(t)\Big|_{\boldsymbol{\Lambda}_{m|k}=\mathbf{O}} = \mathbf{P}_{m|k}^{-1}(t)\mathbf{p}_{m|k}(t)$$
$$\widehat{\boldsymbol{\theta}}_{m|k}^{\mathrm{LBF}}(t) = \mathbf{F}_{m|k}\widehat{\boldsymbol{\alpha}}_{m|k}^{\mathrm{LBF}}(t) \tag{16}$$

From the qualitative point of view, the LBF and RLBF approaches can be regarded as an extension, to the process identification case, of the signal smoothing technique known as Savitzky-Golay filtering [39] (filtering successive subsets of adjacent data points with a low-degree polynomial by the method of least squares).

III. Properties of RLBF Estimators

Originally, regularization was introduced as a general treatment of ill-posed problems, which can also be the case for LBF estimators since the dimension of the generalized regression matrix $\mathbf{P}_{m|k}$ can be much larger than the number of process parameters and close to the number of analyzed data points. From the identification point of view, regularization techniques may prove useful to achieve better bias-variance trade-off. It is well known that the discrepancy between true process trajectories and their estimates, measured by the mean square estimation error, is the sum of the bias and variance components. Therefore, reduction of the variance (at the cost of increased bias), accomplished with a carefully calibrated regularization method, can improve the MSE measure of fit.

MOST WIEDZY Downloaded from mostwiedzy.pl

A. Deterministic Case

Within this subsection we will assume that $\{\boldsymbol{\varphi}(t)\}$ is a deterministic sequence and hence all expectations will be carried over $\{e(t)\}$. We will assume that

A1) $\{e(t)\}$, independent of $\{\boldsymbol{\varphi}(t)\}$, is a sequence of zero-mean independent and identically distributed random variables with variance σ_e^2.

Note that

$$\widehat{\boldsymbol{\alpha}}_{m|k}^{\mathrm{RLBF}}(t) - \boldsymbol{\alpha}_{m|k} = \mathbf{v}_{m|k}(t) + \mathbf{h}_{m|k}(t), \tag{17}$$

where

$$\begin{aligned}
\mathbf{v}_{m|k}(t) &= -\mathbf{R}_{m|k}^{-1}(t)\boldsymbol{\Lambda}_{m|k}\boldsymbol{\alpha}_{m|k} \\
\mathbf{h}_{m|k}(t) &= \mathbf{R}_{m|k}^{-1}(t)\mathbf{r}_{m|k}(t) \\
\mathbf{r}_{m|k}(t) &= \sum_{i=-k}^{k} w_k(i)\boldsymbol{\psi}_{m|k}(t,i)e(t+i).
\end{aligned} \tag{18}$$

Since $\mathrm{E}[\mathbf{r}_{m|k}(t)] = 0$, one obtains

$$\mathrm{E}\left[\widehat{\boldsymbol{\alpha}}_{m|k}^{\mathrm{RLBF}}(t) - \boldsymbol{\alpha}_{m|k}\right] = \mathbf{v}_{m|k}(t) \tag{19}$$

which means that, unlike the LBF estimator, the regularized LBF estimator is biased. Denote by $\boldsymbol{\Delta}_{m|k}(t)$ the mean square hyperparameter estimation error matrix. Using (17), one arrives at

$$\begin{aligned}
\boldsymbol{\Delta}_{m|k}(t) &= \mathrm{E}\left\{ \left[\widehat{\boldsymbol{\alpha}}_{m|k}^{\mathrm{RLBF}}(t) - \boldsymbol{\alpha}_{m|k}\right] \left[\widehat{\boldsymbol{\alpha}}_{m|k}^{\mathrm{RLBF}}(t) - \boldsymbol{\alpha}_{m|k}\right]^{\mathrm{T}} \right\} \\
&= \mathrm{E}[\mathbf{v}_{m|k}(t)\mathbf{v}_{m|k}^{\mathrm{T}}(t)] + \mathrm{E}[\mathbf{h}_{m|k}(t)\mathbf{v}_{m|k}^{\mathrm{T}}(t)] \\
&\quad + \mathrm{E}[\mathbf{v}_{m|k}(t)\mathbf{h}_{m|k}^{\mathrm{T}}(t)] + \mathrm{E}[\mathbf{h}_{m|k}(t)\mathbf{h}_{m|k}^{\mathrm{T}}(t)].
\end{aligned}$$

Then, using (18), one obtains

$$\mathrm{E}[\mathbf{v}_{m|k}(t)\mathbf{v}_{m|k}^{\mathrm{T}}(t)] = \mathbf{R}_{m|k}^{-1}(t)\boldsymbol{\Lambda}_{m|k}\boldsymbol{\alpha}_{m|k}\boldsymbol{\alpha}_{m|k}^{\mathrm{T}}\boldsymbol{\Lambda}_{m|k}\mathbf{R}_{m|k}^{-1}(t).$$

Under assumption (A1) it holds that

$$\mathrm{E}[\mathbf{h}_{m|k}(t)\mathbf{v}_{m|k}^{\mathrm{T}}(t)] = \mathrm{E}[\mathbf{v}_{m|k}(t)\mathbf{h}_{m|k}^{\mathrm{T}}(t)] = 0.$$

Finally, since $\{e(t)\}$ is white noise

$$\begin{aligned}
\mathrm{E}[\mathbf{h}_{m|k}(t)\mathbf{h}_{m|k}^{\mathrm{T}}(t)] &= \mathbf{R}_{m|k}^{-1}(t)\mathrm{E}\Bigg[\sum_{i=-k}^{k}\sum_{j=-k}^{k} w_k(i)w_k(j) \\
&\quad \times \boldsymbol{\psi}_{m|k}(t,i)\boldsymbol{\psi}_{m|k}^{\mathrm{T}}(t,j)e(t+i)e(t+j)\Bigg]\mathbf{R}_{m|k}^{-1}(t) \\
&= \mathbf{R}_{m|k}^{-1}(t)\sigma_e^2\sum_{i=-k}^{k} w_k^2(i)\boldsymbol{\psi}_{m|k}(t,i)\boldsymbol{\psi}_{m|k}^{\mathrm{T}}(t,i)\mathbf{R}_{m|k}^{-1}(t).
\end{aligned}$$

Combining all of the above, one arrives at

$$\begin{aligned}
&\boldsymbol{\Delta}_{m|k}(t) \\
&= \mathbf{R}_{m|k}^{-1}(t)\left[\mathbf{S}_{m|k}(t) + \boldsymbol{\Lambda}_{m|k}\boldsymbol{\alpha}_{m|k}\boldsymbol{\alpha}_{m|k}^{\mathrm{T}}\boldsymbol{\Lambda}_{m|k}\right]\mathbf{R}_{m|k}^{-1}(t),
\end{aligned} \tag{20}$$

where

$$\mathbf{S}_{m|k}(t) = \sigma_e^2\sum_{i=-k}^{k} w_k^2(i)\boldsymbol{\psi}_{m|k}(t,i)\boldsymbol{\psi}_{m|k}^{\mathrm{T}}(t,i). \tag{21}$$

The mean square parameter estimation error matrix $\boldsymbol{\Sigma}_{m|k}(t)$ takes the form

$$\begin{aligned}
&\boldsymbol{\Sigma}_{m|k}(t) \\
&= \mathrm{E}\left\{ \left[\widehat{\boldsymbol{\theta}}_{m|k}^{\mathrm{RLBF}}(t) - \boldsymbol{\theta}(t)\right] \left[\widehat{\boldsymbol{\theta}}_{m|k}^{\mathrm{RLBF}}(t) - \boldsymbol{\theta}(t)\right]^{\mathrm{T}} \right\} \\
&= \mathbf{F}_{m|k}\boldsymbol{\Delta}_{m|k}(t)\mathbf{F}_{m|k}^{\mathrm{T}}.
\end{aligned} \tag{22}$$

B. Stochastic Case

In this section we will assume that the sequence of regression vectors is stochastic and all expectations will be additionally carried out over $\{\boldsymbol{\varphi}(t)\}$. In addition to (A1), we will assume that

A2) $\{\boldsymbol{\varphi}(t)\}$ is a zero-mean wide sense stationary Gaussian sequence, persistently exciting of order at least n, with an exponentially decaying autocorrelation function.

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

Denote by $\mathrm{E}[\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)] = \boldsymbol{\Phi}_0 > 0$, $\forall t$, the positive definite covariance matrix of $\boldsymbol{\varphi}(t)$. It is possible to show that (see Appendix 1)

$$\begin{aligned}
&\lim_{k\to\infty}\mathbf{P}_{m|k}(t) = \mathrm{E}[\mathbf{P}_{m|k}(t)] \\
&= \mathrm{E}\Bigg\{\sum_{i=-k}^{k} w_k(i)[\boldsymbol{\varphi}(t+i)\boldsymbol{\varphi}^{\mathrm{T}}(t+i)]\otimes \\
&\qquad \otimes [\mathbf{f}_{m|k}(i)\mathbf{f}_{m|k}^{\mathrm{T}}(i)]\Bigg\} = \boldsymbol{\Phi}_0\otimes\mathbf{I}_m = \overline{\mathbf{P}}_{m|k},
\end{aligned} \tag{23}$$

where convergence takes place in the mean square sense.

Setting $\mathbf{P}_{m|k}(t) \cong \overline{\mathbf{P}}_{m|k}$, one arrives at

$$\begin{aligned}
\mathrm{E}[\boldsymbol{\Delta}_{m|k}(t)] &\cong \overline{\mathbf{R}}_{m|k}^{-1}[\overline{\mathbf{S}}_{m|k} + \boldsymbol{\Lambda}_{m|k}\boldsymbol{\alpha}_{m|k}\boldsymbol{\alpha}_{m|k}^{\mathrm{T}}\boldsymbol{\Lambda}_{m|k}]\overline{\mathbf{R}}_{m|k}^{-1} \\
&= \overline{\boldsymbol{\Delta}}_{m|k}
\end{aligned} \tag{24}$$

and

$$\mathrm{E}[\boldsymbol{\Sigma}_{m|k}(t)] \cong \mathbf{F}_{m|k}\overline{\boldsymbol{\Delta}}_{m|k}\mathbf{F}_{m|k}^{\mathrm{T}} = \overline{\boldsymbol{\Sigma}}_{m|k}, \tag{25}$$

where

$$\begin{aligned}
\overline{\mathbf{R}}_{m|k} &= \overline{\mathbf{P}}_{m|k} + \boldsymbol{\Lambda}_{m|k} \\
\overline{\mathbf{S}}_{m|k} &= \mathrm{E}[\mathbf{S}_{m|k}(t)] = \sigma_e^2\boldsymbol{\Phi}_0\otimes\mathbf{H}_{m|k} \\
\mathbf{H}_{m|k} &= \sum_{i=-k}^{k} w_k^2(i)\mathbf{f}_{m|k}(i)\mathbf{f}_{m|k}^{\mathrm{T}}(i).
\end{aligned} \tag{26}$$

We will optimize the RLBF scheme in the case where the regularization matrix $\boldsymbol{\Lambda}_{m|k}$ is diagonal

$$\begin{aligned}
\boldsymbol{\Lambda}_{m|k} &= \mathrm{diag}\{\boldsymbol{\Lambda}_{1;m|k},\ldots,\boldsymbol{\Lambda}_{n;m|k}\} \\
\boldsymbol{\Lambda}_{j;m|k} &= \mathrm{diag}\{\lambda_{j1;m|k},\ldots,\lambda_{jm;m|k}\}
\end{aligned}$$

and $\boldsymbol{\Phi}_0 = \sigma_\varphi^2\mathbf{I}_n$. The latter condition holds e.g. when (1) describes a FIR system excited by a white input sequence. Under

MOST WIEDZY Downloaded from mostwiedzy.pl

the constraints mentioned above

$$\overline{\mathbf{R}}_{m|k} = \sigma_\varphi^2 \mathbf{I}_{nm} + \mathbf{\Lambda}_{m|k}.$$

B. Approach 1

We will choose the elements of the diagonal regularization matrix so as to minimize the mean square hyperparameter estimation errors $\mathrm{E}\{[\widehat{b}^{\mathrm{RLBF}}_{jl;m|k}(t) - b_{jl;m|k}]^2\}$ which constitute the diagonal elements of the MSE matrix $\overline{\mathbf{\Delta}}_{m|k}$. Denote by $\overline{\boldsymbol{\delta}}_{m|k}$ the vector made up of diagonal elements of $\overline{\mathbf{\Delta}}_{m|k}$

$$\overline{\boldsymbol{\delta}}_{m|k} = [\delta_{11;m|k}, \ldots, \delta_{1\,m;m|k}, \ldots, \delta_{n1;m|k}, \ldots, \delta_{nm;m|k}]^{\mathrm{T}}$$

Straightforward calculations lead to

$$\begin{aligned}
\delta_{jl;m|k} &= \frac{s_{jl;m|k} + b^2_{jl;m|k}\lambda^2_{jl;m|k}}{[\sigma_\varphi^2 + \lambda_{jl;m|k}]^2} \\
s_{jl;m|k} &= \sigma_e^2 \sigma_\varphi^2 \eta_{l;m|k} \\
\eta_{l;m|k} &= \sum_{i=-k}^{k} w_k^2(i) f^2_{l|k}(i).
\end{aligned} \tag{27}$$

Since for any $a, b, c > 0$ it holds that

$$\arg\min_x \frac{a + bx^2}{(c+x)^2} = \frac{a}{bc}$$

the regularization coefficient which is optimal in the mean square error sense is given by

$$\begin{aligned}
\lambda^{\mathrm{opt}}_{jl;m|k} &= \frac{\sigma_e^2 \eta_{l;m|k}}{b^2_{jl;m|k}} \\
& j = 1, \ldots, n, \;\; l = 1, \ldots, m.
\end{aligned} \tag{28}$$

Since the optimal regularization coefficients minimize all diagonal elements of $\overline{\mathbf{\Delta}}_{m|k}$, they also minimize the L_2 norm of the estimation error: $\mathrm{E}[\| \widehat{\boldsymbol{\alpha}}^{\mathrm{RLBF}}_{m|k}(t) - \boldsymbol{\alpha}_{m|k} \|^2]$.

B. Approach 2

We will choose the elements of the diagonal regularization matrix so as to minimize the mean square parameter estimation errors $\mathrm{E}\{[\widehat{\theta}^{\mathrm{RLBF}}_{j;m|k}(t) - \theta_j(t)]^2\}$ which constitute the diagonal elements of the MSE matrix $\overline{\mathbf{\Sigma}}_{m|k}$. Note that minimization of these errors is the primary goal of identification.

Let

$$\overline{\sigma}_{j;m|k} = [\overline{\mathbf{\Sigma}}_{m|k}]_{jj} = \mathrm{E}\{[\widehat{\theta}^{\mathrm{RLBF}}_{j;m|k}(t) - \theta_j(t)]^2\}.$$

It is easy to check that

$$\begin{aligned}
\overline{\sigma}_{j;m|k} &= \mathbf{f}^{\mathrm{T}}_{m|k}(0)[\sigma_\varphi^2 \mathbf{I}_m + \mathbf{\Lambda}_{j;m|k}]^{-1} \\
&\quad \times [\sigma_e^2 \sigma_\varphi^2 \mathbf{H}_{m|k} + \mathbf{\Lambda}_{j;m|k} \boldsymbol{\beta}_{j;m|k} \boldsymbol{\beta}^{\mathrm{T}}_{j;m|k} \mathbf{\Lambda}_{j;m|k}] \\
&\quad \times [\sigma_\varphi^2 \mathbf{I}_m + \mathbf{\Lambda}_{j;m|k}]^{-1} \mathbf{f}_{m|k}(0) \\
&= \sigma_e^2 \sigma_\varphi^2 \mathbf{x}^{\mathrm{T}}_{j;m|k} \mathbf{H}_{m|k} \mathbf{x}_{j;m|k} \\
&\quad + [\mathbf{f}_{m|k}(0) - \sigma_\varphi^2 \mathbf{x}_{j;m|k}]^{\mathrm{T}} \mathbf{A}_{j;m|k} [\mathbf{f}_{m|k}(0) - \sigma_\varphi^2 \mathbf{x}_{j;m|k}],
\end{aligned} \tag{29}$$

where $\mathbf{x}_{j;m|k} = [x_{j1;m|k}, \ldots, x_{jm;m|k}]^{\mathrm{T}}$, $\mathbf{A}_{j;m|k} = \boldsymbol{\beta}_{j;m|k} \boldsymbol{\beta}^{\mathrm{T}}_{j;m|k}$ and

$$x_{jl;m|k} = \frac{f_{l|k}(0)}{\lambda_{jl;m|k} + \sigma_\varphi^2}. \tag{30}$$

Prior to minimizing the mean square parameter estimation error (29), we will introduce the row/column deletion shorthands. Consider any $m \times 1$ vector $\mathbf{c}$ and any $m \times m$ matrix $\mathbf{C}$. Let $\mathcal{I} \neq \emptyset$ and $\mathcal{J} \neq \emptyset$ be any subsets of $\Omega = \{1, \ldots, m\}$. By $\mathbf{c}^{<\mathcal{I}|}$ and $\mathbf{C}^{<\mathcal{I}|}$ we will denote the vector/matrix obtained after deleting from $\mathbf{c}$ and $\mathbf{C}$ all rows indicated by the set $\mathcal{I}$. Similarly, by $\mathbf{c}^{|\mathcal{J}>}$ and $\mathbf{C}^{|\mathcal{J}>}$ we will denote the vector/matrix obtained after deleting from $\mathbf{c}$ and $\mathbf{C}$ all columns indicated by the set $\mathcal{J}$. Finally, by $\mathbf{C}^{<\mathcal{I}|\mathcal{J}>} = (\mathbf{C}^{<\mathcal{I}|})^{|\mathcal{J}>} = (\mathbf{C}^{|\mathcal{J}>})^{<\mathcal{I}|}$ we will denote the result of deleting from $\mathbf{C}$ the indicated rows and columns.

Note that when $f_{l|k}(0) = 0$, the mean square error $\overline{\sigma}_{j;m|k}$ does not depend on $x_{jl;m|k}$, and hence also on $\lambda_{jl;m|k}$. Such a situation may take place, for example, for a polynomial basis. Actually, when the basis set corresponding to (11) is orthogonalized sequentially using the Gram-Schmidt procedure, it holds that $f_{l|k}(0) = 0$ for all even values of l. Therefore, to make the problem of minimization of $\overline{\sigma}_{j;m|k}$ well posed, we will eliminate from the vector $\mathbf{x}_{j;m|k}$ all null elements. Denote by $\mathcal{L}$ the subset of Ω which indicates position of such zero elements $\mathcal{L} = \{l : f_{l|k}(0) = 0\}$ and let

$$\widetilde{\mathbf{x}}_{j;m|k} = \mathbf{x}^{<\mathcal{L}|}_{j;m|k}, \;\; \widetilde{\mathbf{f}}_{m|k}(0) = \mathbf{f}^{<\mathcal{L}|}_{m|k}(0), \;\; \widetilde{\boldsymbol{\beta}}_{j;m|k} = \boldsymbol{\beta}^{<\mathcal{L}|}_{j;m|k}$$

$$\widetilde{\mathbf{A}}_{j;m|k} = \widetilde{\boldsymbol{\beta}}_{j;m|k} \widetilde{\boldsymbol{\beta}}^{\mathrm{T}}_{j;m|k} \;\; \widetilde{\mathbf{H}}_{m|k} = \mathbf{H}^{<\mathcal{L}|\mathcal{L}>}_{m|k}.$$

Observe, that (29) can be rewritten in the form

$$\begin{aligned}
\overline{\sigma}_{j;m|k} &= \sigma_e^2 \sigma_\varphi^2 \widetilde{\mathbf{x}}^{\mathrm{T}}_{j;m|k} \widetilde{\mathbf{H}}_{m|k} \widetilde{\mathbf{x}}_{j;m|k} \\
&\quad + [\widetilde{\mathbf{f}}_{m|k}(0) - \sigma_\varphi^2 \widetilde{\mathbf{x}}_{j;m|k}]^{\mathrm{T}} \widetilde{\mathbf{A}}_{j;m|k} [\widetilde{\mathbf{f}}_{m|k}(0) - \sigma_\varphi^2 \widetilde{\mathbf{x}}_{j;m|k}]
\end{aligned} \tag{31}$$

leading to

$$\begin{aligned}
\widetilde{\mathbf{x}}^{\mathrm{opt}}_{j;m|k} &= \arg\min_{\widetilde{\mathbf{x}}_{j;m|k}} \overline{\sigma}_{j;m|k} \\
&= [\sigma_e^2 \widetilde{\mathbf{H}}_{m|k} + \sigma_\varphi^2 \widetilde{\mathbf{A}}_{j;m|k}]^{-1} \widetilde{\mathbf{A}}_{j;m|k} \widetilde{\mathbf{f}}_{m|k}(0) \\
&= \frac{\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}_{j;m|k} \widetilde{\mathbf{f}}_{m|k}(0)}{\sigma_e^2 + \sigma_\varphi^2 \widetilde{\boldsymbol{\beta}}^{\mathrm{T}}_{j;m|k} \widetilde{\mathbf{H}}^{-1}_{m|k} \widetilde{\boldsymbol{\beta}}_{j;m|k}} \widetilde{\mathbf{H}}^{-1}_{m|k} \widetilde{\boldsymbol{\beta}}_{j;m|k},
\end{aligned} \tag{32}$$

where the last transition follows from the well known matrix inversion lemma [40]. According to (30),

$$\lambda^{\mathrm{opt}}_{jl;m|k} = \frac{f_{l|k}(0)}{\widetilde{x}^{\mathrm{opt}}_{jl;m|k}} - \sigma_\varphi^2. \tag{33}$$

The expression (33) was derived assuming that $l \in \overline{\mathcal{L}} = \Omega - \mathcal{L}$. Our analysis shows that all elements of the regularization matrix $\mathbf{\Lambda}_{m|k}$ which correspond to $l \in \mathcal{L}$ should have no, or at least negligible [since (25) is an approximate MSE expression] influence on the value of the mean square parameter estimation error. This means that the corresponding values of $\lambda_{jl;m|k}$ can be chosen arbitrarily, e.g. set to zero. However, to retain numerical robustness of the RLBF estimator, we recommend replacement

MOST WIEDZY Downloaded from mostwiedzy.pl

of all such "dummy" elements with the quantities (28) obtained by means of minimization of the hyperparameter MSE. This leads to

$$\lambda_{jl;m|k}^{\mathrm{opt}} = \begin{cases} \frac{f_{l|k}(0)}{\widetilde{x}_{jl;m|k}^{\mathrm{opt}}} - \sigma_\varphi^2 \text{ for } l \in \overline{\mathcal{L}} \\ \frac{\sigma_e^2 \eta_{l;m|k}}{b_{jl;m|k}^2} \qquad \text{for } l \in \mathcal{L} \end{cases}$$
$$j = 1, \ldots, n, \;\; l = 1, \ldots, m \tag{34}$$

Remark 1: The formulae (28) and (33) are generalizations of a similar result obtained in [31] for time invariant FIR systems. Note that for $m = 1$, $f_{1|k}(i) = 1$, $\forall i$, and for the rectangular window $w(i) = 1, i \in I_k$, (28) reduces down to the formula derived in [31].

B. Adaptive Regularization

As expected, the optimal regularization coefficients depend on the true (unknown) hyperparameters. Therefore, to obtain suboptimal but implementable adaptive estimation scheme, we propose to replace the unknown quantities which appear in (28) and (32)–(34), with their LBF estimates. Even though optimization of regularization coefficients was carried out under the assumption that regressors are orthogonal, the method seems to work pretty well in the general correlated covariates case.

The LBF estimates of the coefficients $b_{jl;m|k}, l = 1, \ldots, m$, can be extracted from the the j-th subvector $\widehat{\boldsymbol{\beta}}_{j;m|k}^{\mathrm{LBF}}(t)$ of the vector $\widehat{\boldsymbol{\alpha}}_{m|k}^{\mathrm{LBF}}(t)$.

The LBF estimate of $\rho = \sigma_e^2$ can be obtained from

$$\begin{aligned} \widehat{\rho}_{m|k}^{\mathrm{LBF}}(t) &= \frac{1}{L_k} \sum_{i=-k}^{k} w_k(i)[y(t+i) - \boldsymbol{\psi}_{m|k}^{\mathrm{T}}(t,i)\widehat{\boldsymbol{\alpha}}_{m|k}^{\mathrm{LBF}}(t)]^2 \\ &= \frac{1}{L_k}[c_k(t) - \mathbf{p}_{m|k}^{\mathrm{T}}(t)\widehat{\boldsymbol{\alpha}}_{m|k}^{\mathrm{LBF}}(t)], \end{aligned} \tag{35}$$

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

$$L_k = \sum_{i=-k}^{k} w_k(i) \tag{36}$$

denotes the effective width of the window $\{w_k(i)\}$.

Finally, the local, exponentially weighted estimate of σ_φ^2 can be obtained from

$$\begin{aligned} \widehat{\sigma}_\varphi^2(t) &= \frac{r_\varphi(t)}{l_t m} \\ r_\varphi(t) &= \sum_{i=0}^{t-1} \lambda_0^i ||\boldsymbol{\varphi}(t-i)||^2 \; = \lambda_0 r_\varphi(t-1) + ||\boldsymbol{\varphi}(t)||^2, \end{aligned} \tag{37}$$

where $\lambda_0, 0 < \lambda_0 < 1$, denotes the so-called forgetting constant and $l_t = \sum_{i=0}^{t-1} \lambda_0^i = \lambda_0 l_{t-1} + 1$ denotes the effective width of the exponential window.

The adaptive version of the RLBF estimator can be obtained by replacing the matrix $\boldsymbol{\Lambda}_{m|k}$ in (14) with its locally optimized version

$$\begin{aligned} \widehat{\boldsymbol{\Lambda}}_{m|k}(t) = \mathrm{diag}\{&\widehat{\lambda}_{11;m|k}(t), \ldots, \widehat{\lambda}_{1\,m;m|k}(t), \ldots, \\ &\widehat{\lambda}_{n1;m|k}(t), \ldots, \widehat{\lambda}_{nm;m|k}(t)\} \end{aligned} \tag{38}$$

where the quantities $\widehat{\lambda}_{jl;m|k}(t)$ are obtained from (28) or (32)-(34) after replacing σ_e^2, σ_φ^2 and $b_{jl;m|k}$ with $\widehat{\rho}_{m|k}^{\mathrm{LBF}}(t)$, $\widehat{\sigma}_\varphi^2(t)$ and $\widehat{b}_{jl;m|k}^{\mathrm{LBF}}(t)$, respectively.

IV. Adaptive Selection of the Number of Basis Functions and the Analysis Window Size

The number of basis functions (m) and the size of the local analysis window (k) are important design parameters. When m is large and/or when k is small, the bias of the LBF estimates decreases and their variability increases. The opposite effect is observed when m is small and/or k is large [18]. Therefore, to minimize the mean square parameter estimation error, which is the sum of its bias and variance components, one should find such values of k and m which guarantee good bias-variance compromise. Such a trade-off can be reached using the parallel estimation approach. In this framework several LBF algorithms, corresponding to different values of $m \in \mathcal{M}$ and $k \in \mathcal{K}$, are run simultaneously and compared. At each time instant only one of the competing estimates is selected, i.e., the estimated parameter trajectory has the form $\widehat{\boldsymbol{\theta}}_{\widehat{m}(t)|\widehat{k}(t)}^{\mathrm{LBF}}(t)$ where

$$\{\widehat{m}(t), \widehat{k}(t)\} = \underset{\substack{m \in \mathcal{M} \\ k \in \mathcal{K}}}{\arg\min}\, J_{m|k}(t) \tag{39}$$

and $J_{m|k}(t)$ denotes the local decision statistic. We will use for this purpose the localized version of the Akaike's final prediction error (FPE) measure. Denote by $\widetilde{\Omega}_k(t) = \{\widetilde{\boldsymbol{\varphi}}(t+i), \widetilde{e}(t+i), i \in I_k\}$ another realization of the input-output data, independent of the set $\{\Omega_k(t) = \{\boldsymbol{\varphi}(t+i), e(t+i), i \in I_k\}$ that was used for identification purposes. The quantity

$$\mathrm{E}_{\widetilde{\Omega}_k(t), \Omega_k(t)} \left[\widetilde{y}(t) - \widetilde{\boldsymbol{\varphi}}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}_{m|k}^{\mathrm{LBF}}(t)\right]^2 \tag{40}$$

which can be called, after Akaike [41], [42], the final prediction error, is a good measure of predictive capability of the LBF estimator. According to [18], as an estimate of (40), one can use the following statistic

$$\mathrm{FPE}_{m|k}(t) = \frac{1 + \frac{n}{N_{m|k}}}{1 - \frac{n}{M_{m|k}}}\, \widehat{\rho}_{m|k}^{\mathrm{LBF}}(t), \tag{41}$$

where

$$N_{m|k} = \left\{ \sum_{i=-k}^{k} [w_k(i)\mathbf{f}_{m|k}^{\mathrm{T}}(0)\mathbf{f}_{m|k}(i)]^2 \right\}^{-1}$$

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

The selection rule can be obtained by setting $J_{m|k}(t) = \mathrm{FPE}_{m|k}(t)$.

MOST WIEDZY Downloaded from mostwiedzy.pl

V. Computational Complexity of LBF and RLBF Algorithms

Since LBF and RLBF estimates are evaluated in the sliding window mode, i.e., computations are repeated for every new location t of the analysis window $T_k(t)$, the computational burden is high. It can be lowered if the applied window $w_k(i)$ and the vector of basis functions $\mathbf{f}_{m|k}(i)$ are recursively computable. As an example, consider the case where the window is cosinusoidal (9) and the basis is made up of powers of time (12). Note that the weight $w_k(i)$ can be expressed in the form $w_k(i) = \mathrm{Re}\{v_k(i)\}$, where $v_k(i) = e^{j\frac{\pi i}{2k}}$ denotes the recursively computable complex-valued window

$$v_k(i) = \gamma_k v_k(i-1), \quad \gamma_k = e^{j\frac{\pi}{2k}}. \tag{42}$$

Similarly, for the polynomial basis the vector $\mathbf{f}_{m|k}(i)$ is recursively computable

$$\mathbf{f}_{m|k}(i) = \mathbf{\Gamma}_{m|k}\mathbf{f}_{m|k}(i-1) \tag{43}$$

where $\mathbf{\Gamma}_{m|k}$ denotes the $m \times m$ transition matrix. Taking advantage of (42) and (43), one can easily derive recursive algorithms for computation of the $mn \times mn$ generalized regression matrix $\mathbf{P}_{m|k}(t)$ and the $m \times 1$ vector $\mathbf{p}_{m|k}(t)$ needed to evaluate LBF and RLBF estimates – for more details see [18]. In this way the computational cost of evaluation $\mathbf{P}_{m|k}(t)$ and $\mathbf{p}_{m|k}(t)$ can be lowered from $O(m^2n^2\,K)$ and $O(mnK)$ multiply-add operations per time update, to $O(m^2n^2)$ and $O(mn)$ operations, respectively, i.e., it becomes independent of the window width $K = 2k+1$.

Note also that for a fixed value of k the generalized regression matrices $\mathbf{P}_{m_1|k}(t), \ldots, \mathbf{P}_{m_{M-1}|k}(t)$ corresponding to different values of m: $m_1 < m_2 < \ldots < m_{M-1} < m_M$ are nested in (are submatrices of) the matrix $\mathbf{P}_{m_M|k}(t)$, i.e., there is no need to evaluate them separately. The same holds true for the vectors $\mathbf{p}_{m_1|k}(t), \ldots, \mathbf{p}_{m_M|k}(t)$.

Unfortunately, it is not possible to recursively update the matrices $\mathbf{P}^{-1}_{m|k}(t)$ and $\mathbf{R}^{-1}_{m|k}(t)$. The computational cost of inverting $\mathbf{P}_{m|k}(t)$ and $\mathbf{R}_{m|k}(t)$ is roughly equal to $O(m^3n^3)$ operations per time update (it can be slightly decreased by exploiting the special structure of both matrices/vectors) which in some applications may be prohibitive.

Additionally, when the number of hypermodel coefficients, equal to mn, becomes comparable with (is not significantly smaller than) the window width K, the matrix $\mathbf{P}_{m|k}(t)$, inverted at the first stage of identification, may become numerically ill-conditioned. In the next section we will derive fast estimation algorithms which, without compromising very good parameter tracking capabilities of the LBF/RLBF scheme, significantly reduce computational burden and are numerically robust.

VI. Fast Local Basis Function Estimators

A. Derivation

Derivation of the fast local basis function (fLBF) algorithm will be based on the following approximation valid for sufficiently large values of k (cf. (23))

$$\widehat{\boldsymbol{\alpha}}^{\mathrm{LBF}}_{m|k}(t) = \begin{bmatrix} \widehat{\boldsymbol{\beta}}^{\mathrm{LBF}}_{1;m|k}(t) \\ \vdots \\ \widehat{\boldsymbol{\beta}}^{\mathrm{LBF}}_{n;m|k}(t) \end{bmatrix} \cong \overline{\mathbf{P}}^{-1}_{m|k}\mathbf{p}_{m|k}(t)$$

$$= \sum_{i=-k}^{k} w_k(i)y(t+i)[\mathbf{\Phi}_0^{-1} \otimes \mathbf{I}_m][\boldsymbol{\varphi}(t+i) \otimes \mathbf{f}_{m|k}(i)]. \tag{44}$$

Note that

$$\widehat{\boldsymbol{\beta}}^{\mathrm{LBF}}_{j;m|k}(t) = \mathbf{D}_j^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}^{\mathrm{LBF}}_{m|k}(t), \tag{45}$$

where

$$\mathbf{D}_j = [\underbrace{\mathbf{O}_m, \ldots, \mathbf{O}_m}_{j-1}, \mathbf{I}_m, \underbrace{\mathbf{O}_m, \ldots, \mathbf{O}_m}_{n-j}]^{\mathrm{T}} = \mathbf{d}_j^{\mathrm{T}} \otimes \mathbf{I}_m$$

$$\mathbf{d}_j = [\underbrace{0, \ldots, 0}_{j-1}, 1, \underbrace{0, \ldots, 0}_{n-j}]^{\mathrm{T}}. \tag{46}$$

Combining (44) with (45) and (46), and using the identity $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$ which holds for Kronecker products, one arrives at

$$\widehat{\boldsymbol{\beta}}^{\mathrm{LBF}}_{j;m|k}(t) \cong \sum_{i=-k}^{k} w_k(i)\mathbf{f}_{m|k}(i)\mathbf{d}_j^{\mathrm{T}}\mathbf{\Phi}_0^{-1}\boldsymbol{\varphi}(t+i)y(t+i). \tag{47}$$

Our next step will be based on the concept of preestimation. Denote by $\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t)$ the exponentially weighted least squares (EWLS) estimate of the parameter vector $\boldsymbol{\theta}(t)$:

$$\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t) = \arg\min_{\boldsymbol{\theta}} \sum_{i=0}^{t-1} \lambda_0^i[y(t-i) - \boldsymbol{\varphi}^{\mathrm{T}}(t-i)\boldsymbol{\theta}]^2. \tag{48}$$

The preestimate of $\boldsymbol{\theta}(t)$ was defined in [19] in the form

$$\boldsymbol{\theta}^*(t) = l_t\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t) - \lambda_0 l_{t-1}\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1) \tag{49}$$

Note that for large values of t, when l_t reaches its constant steady state value $l_\infty = 1/(1-\lambda_0)$, the relationship (49) can be put in the following time invariant form

$$\boldsymbol{\theta}^*(t) = \frac{1}{1-\lambda_0}[\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t) - \lambda_0\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1)]. \tag{50}$$

and the preestimate can be interpreted as a result of inverse filtering of the EWLS estimate.

The preestimate, which can be written down in the form

$$\boldsymbol{\theta}^*(t) = \boldsymbol{\theta}(t) + \mathbf{z}(t) \tag{51}$$

is a raw estimate of $\boldsymbol{\theta}(t)$: it is approximately unbiased ($\mathrm{E}[\mathbf{z}(t)] \cong 0$) but has a very large variability. For this reason, to obtain meaningful estimation results, it should be further processed (postfiltered). As argued in [19], under typical operating conditions the following approximation holds true

$$\mathbf{z}(t) \cong \mathbf{\Phi}_0^{-1}\boldsymbol{\varphi}(t)e(t) = \mathbf{z}_0(t).$$

Note that under assumptions (A1)–(A2), $\{\mathbf{z}_0(t)\}$ is a zero-mean white noise with covariance matrix $\sigma_e^2\mathbf{\Phi}_0^{-1}$.

MOST WIEDZY Downloaded from mostwiedzy.pl

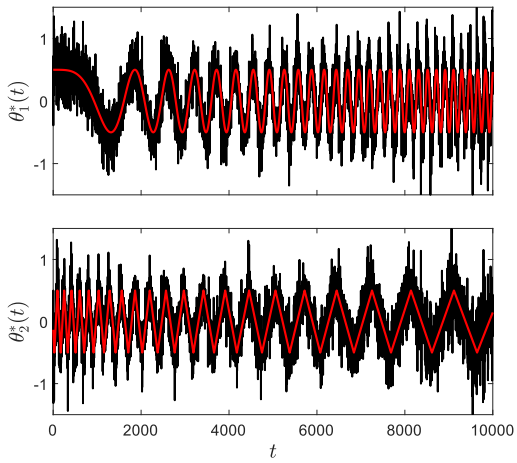

Fig. 1. Preestimated parameter trajectories of a nonstationary two-tap finite impulse response system (SNR=15 dB). Preestimates (black lines) are superimposed on true parameter trajectories (red lines).

To illustrate the preestimation technique, consider the problem of identification of a time-varying two-tap FIR system governed by

$$y(t) = \theta_1(t)u(t-1) + \theta_2(t)u(t-2) + e(t). \tag{52}$$

where $u(t)$ denotes the first order autoregressive input signal $u(t) = 0.8u(t-1) + v(t)$, $\mathrm{var}[v(t)] = 1$, and $\{v(t)\}$ denotes white noise independent of $\{e(t)\}$. The variance of the measurement noise was set to $\sigma_e^2 = 0.02$ and system parameters were modeled as a sinusoidal linear chirp $[\theta_1(t)]$, and an inverted triangular linear chirp $[\theta_2(t)]$, respectively. The preestimates of system parameters, obtained for $\lambda_0 = 0.9$ ($l_\infty = 10$), are shown in Fig. 1. Note that preestimation allows one to "X-ray" the structure of system parameter variation without making any assumptions about its functional form, degree of smoothness, etc. Under assumptions (A1)–(A3), for sufficiently large values of l_t it holds that $\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t) \cong \frac{1}{l_t}\boldsymbol{\Phi}_0^{-1}\sum_{i=0}^{t-1}\lambda_0^i\boldsymbol{\varphi}(t-i)y(t-i)$. Substituting this result into (49), one arrives at the following approximation

$$\boldsymbol{\theta}^*(t) \cong \boldsymbol{\Phi}_0^{-1}\boldsymbol{\varphi}(t)y(t). \tag{53}$$

Combining (53) with (47) and noting that $\mathbf{d}_j^{\mathrm{T}}\boldsymbol{\theta}^*(t) = \theta_j^*(t)$, where $\theta_j^*(t)$ is the preestimate of the j-th process coefficient $\theta_j(t)$, one obtains

$$\widehat{\boldsymbol{\beta}}_{j;m|k}^{\mathrm{fLBF}}(t) = \sum_{i=-k}^{k} w_k(i)\mathbf{f}_{m|k}(i)\theta_j^*(t+i) \cong \widehat{\boldsymbol{\beta}}_{j;m|k}^{\mathrm{LBF}}(t) \qquad j = 1,\ldots,n. \tag{54}$$

The fLBF estimates of process coefficients $\theta_j(t)$, can be obtained from

$$\widehat{\theta}_{j;m|k}^{\mathrm{fLBF}}(t) = \mathbf{f}_{m|k}^{\mathrm{T}}(0)\widehat{\boldsymbol{\beta}}_{j;m|k}^{\mathrm{fLBF}}(t) \qquad j = 1,\ldots,n. \tag{55}$$

Figure 2 shows comparison of the LBF and fLBF estimates of $\theta_1(t)$ and $\theta_2(t)$ for the system (52) in the case where $m = 10$ and

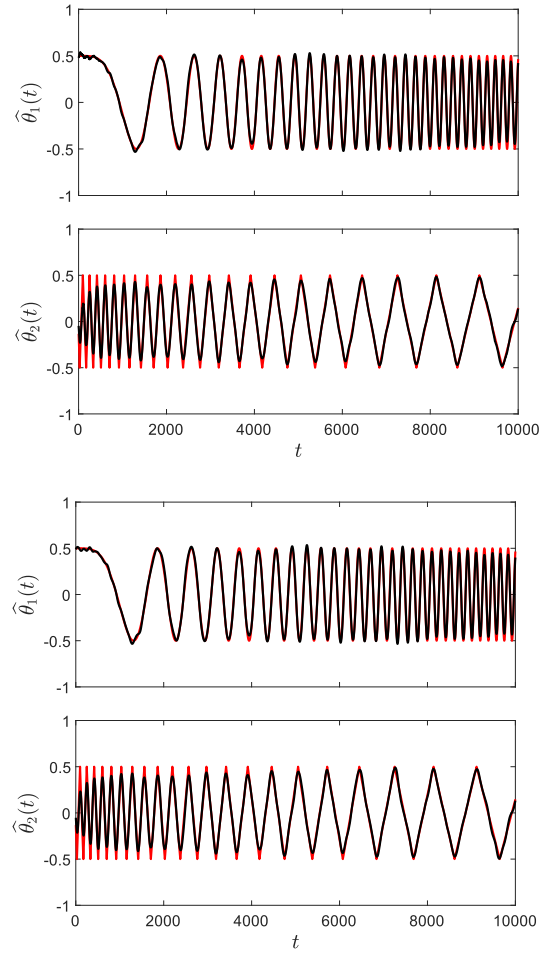

Fig. 2. Comparison of LBF estimates (two upper plots) and fLBF estimates (two lower plots) obtained for a nonstationary FIR system ($k = 250$, $m = 10$). Estimated trajectories (black lines) are superimposed on true trajectories (red lines).

$k = 250$. Note that the fLBF estimates (obtained by processing preestimates shown in Fig. 1) are almost identical with LBF estimates – in spite of the fact that the computational burden of the fLBF algorithm is much smaller than that of the LBF algorithm.

When the regression vector has the form (3), equation (1) describes a nonstationary autoregressive (AR) signal. Even though all computational algorithms carry on to this case, analysis of their properties along the lines presented above is not possible simply because the covariance matrix of $\boldsymbol{\varphi}(t)$ depends on $\boldsymbol{\theta}(t)$, and hence on time. It is therefore quite surprising that, in spite of the limitations mentioned above, the fLBF estimates evaluated for nonstationary AR signals are almost indistinguishable from the LBF estimates. This is illustrated in Fig. 3, which compares LBF and fLBF estimates obtained for a second order AR signal governed by

$$y(t) = \theta_1(t)y(t-1) + \theta_2(t)y(t-2) + e(t) \tag{56}$$

subject to the same chirp-like parameter changes as those considered in our FIR example (52). Moreover, the observed close

MOST WIEDZY Downloaded from mostwiedzy.pl

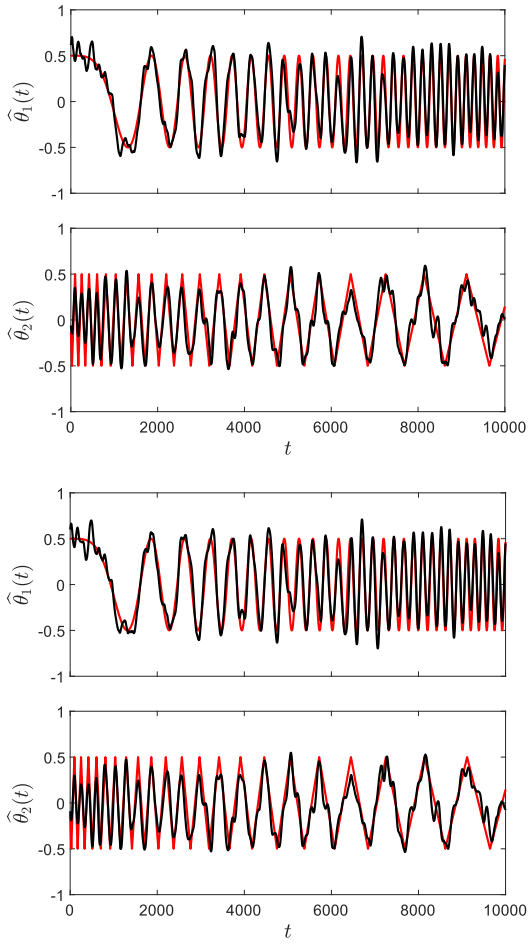

Fig. 3. Comparison of LBF estimates (two upper plots) and fLBF estimates (two lower plots) obtained for a nonstationary AR signal ($k = 250$, $m = 10$). Estimated trajectories (black lines) are superimposed on true trajectories (red lines).

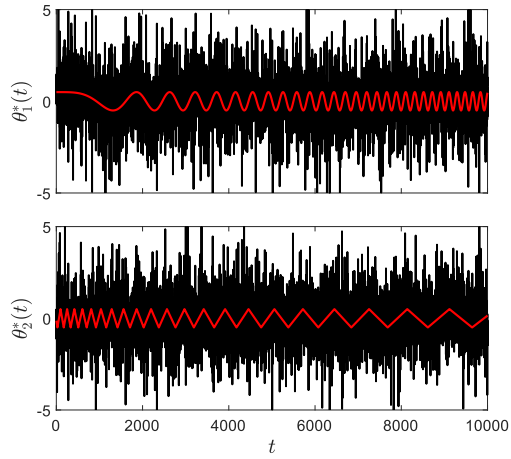

Fig. 4. Preestimated parameter trajectories of a nonstationary AR signal. Preestimates (black lines) are superimposed on true parameter trajectories (red lines).

resemblance occurs despite the fact that the preestimates, evaluated in the same way as before ($\lambda_0 = 0.9$), do not look promising since, unlike the FIR case, they hardly reveal the true nature of parameter variation – see Fig. 4.

B. Statistical Reinterpretation of the Fast LBF Estimator

According to (51), the preestimated trajectory $\{\theta_j^*(t)\}$ can be regarded as a noisy version of the true trajectory

$$\theta_j^*(t) = \theta_j(t) + z_j(t), \tag{57}$$

where $z_j(t)$ – the j-th component of $\mathbf{z}(t)$ – is a zero-mean white noise with large variance $\sigma_{z_j}^2$. Denoising of $\{\theta_j^*(t)\}$ can be carried out using the LBF approach. Assuming, at each time instant t, that $\{\theta_j(t)\}$ can be locally modeled as a linear combination of basis functions, i.e., adopting the hypermodel (4), the local estimate of $\boldsymbol{\beta}_{j;m|k}$ can be obtained in the form

$$\widetilde{\boldsymbol{\beta}}_{j;m|k}(t) = \arg\min_{\boldsymbol{\beta}_{j;m|k}} \sum_{i=-k}^{k} w_k(i)[\theta_j^*(t+i) - \mathbf{f}_{m|k}^{\mathrm{T}}(i)\boldsymbol{\beta}_{j;m|k}]^2. \tag{58}$$

It is straightforward to check that

$$\widetilde{\boldsymbol{\beta}}_{j;m|k}(t) = \widehat{\boldsymbol{\beta}}_{j;m|k}^{\mathrm{fLBF}}(t). \tag{59}$$

According to (59), the fLBF estimator, which was originally derived as an approximate solution to the LBF-based output fitting problem – minimization of the output modeling error $y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}(t)$:

$$\widehat{\boldsymbol{\alpha}}_{m|k}^{\mathrm{LBF}}(t) = \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{1;m|k}^{\mathrm{LBF}}(t) \\ \vdots \\ \widehat{\boldsymbol{\beta}}_{n;m|k}^{\mathrm{LBF}}(t) \end{bmatrix} = \arg\min_{\boldsymbol{\alpha}_{m|k}} \left\{ \sum_{i=-k}^{k} w_k(i) \left[y(t+i) - \boldsymbol{\psi}_{m|k}^{\mathrm{T}}(t,i)\boldsymbol{\alpha}_{m|k} \right]^2 \right\}, \tag{60}$$

can be also viewed as a solution to a direct parameter fitting problem – minimization of the parameter modeling error $\boldsymbol{\theta}^*(t) - \widehat{\boldsymbol{\theta}}(t)$. The second interpretation will be useful when designing mechanisms for adaptive tuning of parameters m and k.

C. Selection of Design Parameters

Adaptive selection of m and k can be done in the centralized or decentralized way. In the centralized approach the same values of m and k are adopted for all process coefficients $\theta_1(t), \ldots, \theta_n(t)$. Basically, the selection procedure does not differ from that described in Section IV. The only modification that must be introduced is replacement of the estimate $\widehat{\boldsymbol{\alpha}}_{m|k}^{\mathrm{LBF}}(t)$ in (35) with its approximate, computationally fast version $\widehat{\boldsymbol{\alpha}}_{m|k}^{\mathrm{fLBF}}(t)$. Note that in order to use (35), one has to update the vector $\mathbf{p}_{m|k}(t)$.

In the decentralized approach, parameters m and k are individually adjusted for each component of the parameter vector $\boldsymbol{\theta}(t)$. This can be useful when different parameters vary at different rates, as in the case illustrated in Fig. 1. A straightforward consequence of (58)–(59) is that the fLBF estimator $\widehat{\theta}_{j;m|k}^{\mathrm{fLBF}}(t) =$

MOST WIEDZY Downloaded from mostwiedzy.pl

$\mathbf{f}^{\mathrm{T}}(0)\widehat{\boldsymbol{\beta}}^{\mathrm{fLBF}}_{j;m|k}(t)$ can be regarded as a result of smoothing the preestimated parameter trajectory $\{\theta^*_j(t)\}$ using a generalized form of the Savitzky-Golay filter described in [43]. This also means that the best fitting values of m and k for the j-th system coefficient $\theta_j(t)$ can be selected using the following localized FPE decision rule derived in [43]:

$$\{\widehat{m}_j(t), \widehat{k}_j(t)\} = \arg\min_{\substack{m\in\mathcal{M}\\ k\in\mathcal{K}}} \mathrm{FPE}_{j;m|k}(t), \tag{61}$$

where

$$\mathrm{FPE}_{j;m|k}(t) = \frac{1+\frac{1}{N_{m|k}}}{1-\frac{1}{M_{m|k}}}\ \widehat{\rho}^{\mathrm{fLBF}}_{j;m|k}(t) \tag{62}$$

and $\widehat{\rho}^{\mathrm{fLBF}}_{j;m|k}(t)$ denotes the local estimate of the noise variance $\rho_j = \sigma^2_{z_j}$

$$\begin{aligned}\widehat{\rho}^{\mathrm{fLBF}}_{j;m|k}(t) &= \frac{1}{L_k}\sum_{i=-k}^{k} w_k(i)\left[\theta^*_j(t+i) - \mathbf{f}^{\mathrm{T}}_{m|k}(i)\widehat{\boldsymbol{\beta}}^{\mathrm{fLBF}}_{j;m|k}(t)\right]^2\\ &= \frac{1}{L_k}\sum_{i=-k}^{k} w_k(i)[\theta^*_j(t+i)]^2 - \frac{1}{L_k}\parallel \widehat{\boldsymbol{\beta}}^{\mathrm{fLBF}}_{j;m|k}(t)\parallel^2 .\end{aligned} \tag{63}$$

Note that the FPE statistic (62) has the same functional form as (41).

Remark 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 FPE criterion designed for finite memory causal estimation schemes [44].

D. Computational Complexity of the fLBF Algorithm

The computational cost of running the EWLS algorithm, used to obtain preestimates, is of order $O(n^2)$ per time update (since inverse of the associated regression matrix can be evaluated recursively). Note that the same sequence of preestimates can be used for different values of m and k.

Computation of $\widehat{\alpha}^{\mathrm{fLBF}}_{m|k}(t)$ is matrix-inversion-free and requires $O(mnK)$ operations per time update. It can be further lowered to $O(mn)$ operations if recursively computable basis and window functions are used. In the latter case the computational cost does not depend on the window size K.

When identification can be carried in the off-line mode (rather than with a constant delay, which is obligatory in almost real time applications), the convolutions (54) can be efficiently evaluated in the frequency domain using the FFT-based routine (for *all* basis and window functions) – the associated computational burden is of order $O(mn\log K)$.

Finally, we note that for a fixed value of k and any $m_1 < m_2$, the basis vector $\mathbf{f}_{m_1|k}(t)$ is nested in $\mathbf{f}_{m_2|k}(t)$. Hence, if different values of $m : m_1 < m_2 < \ldots < m_M$ are considered for parallel estimation purposes, all fLBF estimates $\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}_{m_i|k}(t), i < M$, are obtained in the course of evaluating $\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}_{m_M|k}(t)$. This means that there is almost no computational overhead when the analysis, carried out for a given window width, is extended from one hypermodel to M hypermodels.

VII. Fast Regularized Local Basis Function Estimators

Based on the reinterpretation of the fLBF scheme presented in the previous section, the fast regularized local basis function (fRLBF) estimator can be defined as follows

$$\begin{aligned}&\widehat{\boldsymbol{\beta}}^{\mathrm{fRLBF}}_{j;m|k}(t)\\ &= \arg\min_{\boldsymbol{\beta}_{j;m|k}}\left\{\sum_{i=-k}^{k} w_k(i)[\theta^*_j(t+i) - \mathbf{f}^{\mathrm{T}}_{m|k}(i)\boldsymbol{\beta}_{j;m|k}]^2\right.\\ &\qquad\left. + \boldsymbol{\beta}^{\mathrm{T}}_{j;m|k}\boldsymbol{\Lambda}_{j;m|k}\boldsymbol{\beta}_{j;m|k}\right\}\end{aligned} \tag{64}$$

It is straightforward to show that

$$\widehat{\boldsymbol{\beta}}^{\mathrm{fRLBF}}_{j;m|k}(t) = (\mathbf{I}_m + \boldsymbol{\Lambda}_{j;m|k})^{-1}\widehat{\boldsymbol{\beta}}^{\mathrm{fLBF}}_{j;m|k}(t) \tag{65}$$

which leads to

$$\begin{aligned}\widehat{b}^{\mathrm{fRLBF}}_{jl;m|k}(t) &= \frac{\widehat{b}^{\mathrm{fLBF}}_{jl;m|k}(t)}{1+\lambda_{jl;m|k}}\\ &l = 1,\ldots,m.\end{aligned} \tag{66}$$

Since the denominator in (66) is greater than 1, the fRLBF estimates can be obtained by "shrinking" the corresponding fLBF estimates. Note that while in the classical ridge regression [22] the regularization matrix has the form $\boldsymbol{\Lambda}_{j;m|k} = \lambda\mathbf{I}_m$, resulting in uniform shrinkage of fLBF estimates, the proposed procedure provides a more selective shrinkage, which can be considered a "soft thresholding" variant of subset regression.

In order to optimize the shrinkage procedure, note that the problem (64) is a simplified version of the problem (8) – it can be obtained from (8) by setting $n = 1$ and replacing $y(t)$, $\boldsymbol{\varphi}(t)$, σ^2_{φ}, $e(t)$, σ^2_e and $\boldsymbol{\psi}_{m|k}(t,i)$ with $\theta^*_j(t)$, 1, 1, $z_j(t)$, $\sigma^2_{z_j}$ and $\mathbf{f}_{m|k}(i)$, respectively. Optimization of the regularization matrix results in the formulas analogous to (28) and (32)-(34). In the first case, one obtains

$$\lambda^{\mathrm{opt}}_{jl;m|k} = \frac{\sigma^2_{z_j}\eta_{l;m|k}}{b^2_{jl;m|k}}\ . \tag{67}$$

In the second case, one arrives at

$$\lambda^{\mathrm{opt}}_{jl;m|k} = \frac{f_{l|k}(0)}{\widetilde{x}^{\mathrm{opt}}_{jl;m|k}} - 1,\ \ l\in\overline{\mathcal{L}} \tag{68}$$

where

$$\begin{aligned}\widetilde{\mathbf{x}}^{\mathrm{opt}}_{j;m|k} &= [\sigma^2_{z_j}\widetilde{\mathbf{H}}_{m|k} + \widetilde{\mathbf{A}}_{j;m|k}]^{-1}\widetilde{\mathbf{A}}_{j;m|k}\widetilde{\mathbf{f}}_{m|k}(0)\\ &= \frac{\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}_{j;m|k}\widetilde{\mathbf{f}}_{m|k}(0)}{\sigma^2_{z_j} + \widetilde{\boldsymbol{\beta}}^{\mathrm{T}}_{j;m|k}\widetilde{\mathbf{H}}^{-1}_{m|k}\widetilde{\boldsymbol{\beta}}_{j;m|k}}\ \widetilde{\mathbf{H}}^{-1}_{m|k}\widetilde{\boldsymbol{\beta}}_{j;m|k}.\end{aligned} \tag{69}$$

Similar to the RLBF scheme, the fRLBF algorithm can be obtained by replacing the regularization matrix in (65) with its locally optimized version

$$\widehat{\boldsymbol{\Lambda}}_{j;m|k}(t) = \mathrm{diag}\{\widehat{\lambda}_{j1;m|k}(t),\ldots,\widehat{\lambda}_{jm;m|k}(t)\}. \tag{70}$$

MOST WIEDZY Downloaded from mostwiedzy.pl

TABLE I
SUMMARY OF THE RLBF ALGORITHM

1) *Parallel estimation*
Use (16) and (35) to compute the LBF estimates $\widehat{\boldsymbol{\alpha}}^{\mathrm{LBF}}_{m|k}(t)$, $\widehat{\rho}^{\mathrm{LBF}}_{m|k}(t)$ for several choices of m and k.
2) *Selection*
Use (41) to select the best fitting LBF estimate $\widehat{\boldsymbol{\alpha}}^{\mathrm{LBF}}_{\widehat{m}(t)|\widehat{k}(t)}(t)$.
3) *Regularization*
Use (13) and (28) or (34) to compute the RLBF estimate $\widehat{\boldsymbol{\alpha}}^{\mathrm{RLBF}}_{\widehat{m}(t)|\widehat{k}(t)}(t)$.
4) *Output*
Use (15) to compute the RLBF estimate $\widehat{\theta}^{\mathrm{RLBF}}_{\widehat{m}(t)|\widehat{k}(t)}(t)$.

TABLE II
SUMMARY OF THE fRLBF ALGORITHM

1) *Preestimation*
Use (49) to compute preestimates of system coefficients $\theta^*_j(t), j = 1, \ldots, n$.
2) *Parallel estimation*
Use (54) and (63) to compute the fLBF estimates $\widehat{\boldsymbol{\beta}}^{\mathrm{fLBF}}_{j;m|k}(t)$, $\widehat{\rho}^{\mathrm{fLBF}}_{j;m|k}(t)$, $j = 1, \ldots, n$ for several choices of m and k.
3) *Selection*
Use (62) to select the best fitting fLBF estimates $\widehat{\boldsymbol{\beta}}^{\mathrm{fLBF}}_{j;\widehat{m}_j(t)|\widehat{k}_j(t)}(t)$, $j = 1, \ldots, n$.
4) *Regularization*
Use (66) and (67) or (68)-(69) to compute the fRLBF estimates $\widehat{\boldsymbol{\beta}}^{\mathrm{fRLBF}}_{j;\widehat{m}_j(t)|\widehat{k}_j(t)}(t)$, $j = 1, \ldots, n$.
5) *Output*
Use (71) to compute the fRLBF estimates $\widehat{\theta}^{\mathrm{fRLBF}}_{j;\widehat{m}_j(t)|\widehat{k}_j(t)}(t)$, $j = 1, \ldots, n$.

Finally, the fRLBF estimates of process coefficients $\theta_j(t)$, $j = 1, \ldots, n$, can be obtained from

$$\widehat{\theta}^{\mathrm{fRLBF}}_{j;m|k}(t) = \mathbf{f}^{\mathrm{T}}_{m|k}(0)\widehat{\boldsymbol{\beta}}^{\mathrm{fRLBF}}_{j;m|k}(t) = \sum_{l\in\overline{\mathcal{L}}} f_{l|k}(0)\widehat{b}^{\mathrm{fRLBF}}_{jl;m|k}(t). \tag{71}$$

VIII. SUMMARY OF THE PROPOSED IDENTIFICATION ALGORITHMS

The RLBF and fRLBF algorithms are summarized in Tables I and II, respectively.

All steps are repeated for consecutive values of t (sliding window approach). Estimation can be performed either in the off-line mode or in the almost real time mode. In the latter case the latency (estimation delay) is equal to $k_{\max}+1$ sampling intervals, where $k_{\max}$ is the largest adopted value of k.

Note that regularization is applied only to the best fitting LBF/fLBF estimates.

IX. COMPUTER SIMULATIONS

The first two simulation experiments were carried out for artificially generated data, obtained using the models described before: (52) in the case of a FIR system, and (56) in the case of an AR signal. In both cases the chirp-like parameter trajectories, shown in Figs. 1–4, were applied. To avoid boundary/transient issues, data generation was started 1000 time instants prior to $t = 1$ and was continued for 1000 time instants after $t = T_s$, where T_s denotes simulation time. Experiments were performed for three different speeds of parameter variation – fast ($T_s = 5000$), medium speed ($T_s = 10\,000$), and slow ($T_s = 20\,000$). The fast/slow parameter trajectories were obtained by downsampling/upsampling trajectories corresponding to medium speed changes by the factor of 2.

The measurement/driving noise variance was set to $\sigma_e^2 = 0.02$. For the FIR system (52) this corresponds to SNR, measured as $\mathrm{E}[(\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t))^2]/\sigma_e^2$, approximately equal to 15 dB. In the AR case, governed by (56), the analogously defined SNR is time-varying and does not depend on σ_e^2.

Identification was carried out using five algorithms: LBF, fLBF, RLBF, fRLBF and the multi-scale wavelet (MW) algorithm described in [16]. To compare the performance of the aforementioned methods, we used the MSE index averaged over time and over 100 independent realizations of the input and/or noise sequences. For algorithms based on the local basis function approach (polynomial basis, cosinusoidal window) the width of the analysis window was set to $K = 501$ ($k = 250$). The adaptive choice was confined to three estimators equipped with $m \in \{5, 10, 20\}$. In all cases regularization was carried out using Approach 1, after observing that the computationally more involved Approach 2 yields almost identical results.

Following recommendations given in [17], the multi-scale wavelet approach involved third, fourth and fifth order cardinal B-splines and the resolution level was set to 3. The best subset of 33 approximating wavelets was selected by the orthogonal least squares algorithm and the number of wavelets – by the well-known Akaike's information criterion (AIC). The approximation was carried out in intervals of length 501 using the overlap-add approach (50 % overlap, Hann synthesis window).

Based on the simulation results, gathered in Table III, several conclusions can be drawn. First, it can be seen that regularization provides results that are almost always better than those obtained for the corresponding algorithms without regularization. Secondly, for both models the performance of the LBF/RLBF algorithms and their fast versions fLBF/fRLBF is very similar (which is quite surprising in the autoregressive case). As a matter of fact, pretty often the fLBF/fRLBF methods yield results that are better than those provided by the original LBF/RLBF algorithms. Thirdly, the proposed adaptive approach yields results that are almost always better than results provided by the best algorithms with fixed settings. Finally, the proposed fRLBF scheme in its adaptive version provides results that are almost always better (often at least two times better in terms of MSE) than those yielded by the MW approach – see e.g. Fig. 5.

Our second simulation experiment aimed at checking what is the influence of the type of the applied functional basis on

MOST WIEDZY Downloaded from mostwiedzy.pl

TABLE III
MEAN SQUARE PARAMETER ESTIMATION ERRORS OBTAINED FOR DIFFERENT IDENTIFICATION ALGORITHMS EQUIPPED WITH DIFFERENT NUMBER OF BASIS FUNCTIONS (m) AND FOR AN ADAPTIVE CHOICE (A) OF THE NUMBER OF BASIS FUNCTIONS. MW DENOTES THE MULTI-SCALE WAVELET APPROACH. ALL AVERAGES WERE COMPUTED FOR 100 PROCESS REALIZATIONS, TWO DIFFERENT PROCESS MODELS (FIR AND AR), AND THREE SPEEDS OF PARAMETER VARIATION. THE BEST RESULTS IN EACH GROUP ARE SHOWN IN BOLDFACE

FIR

	Fast				Medium speed				Slow			
Method \ m	5	10	20	A	5	10	20	A	5	10	20	A
LBF	9.36E-02	4.85E-02	1.48E-03	1.48E-03	3.01E-02	1.62E-03	1.27E-03	1.12E-03	2.75E-03	7.52E-04	1.16E-03	7.66E-04
fLBF	8.94E-02	4.70E-02	2.44E-03	2.40E-03	2.87E-02	1.83E-03	1.55E-03	1.19E-03	2.67E-03	8.16E-04	1.24E-03	7.56E-04
RLBF	9.29E-02	4.80E-02	**1.21E-03**	1.21E-03	2.97E-02	1.57E-03	9.07E-04	**8.70E-04**	2.70E-03	6.23E-04	7.08E-04	**5.71E-04**
fRLBF	8.93E-02	4.67E-02	2.03E-03	2.21E-03	2.85E-02	1.79E-03	1.14E-03	1.09E-03	2.65E-03	7.08E-04	8.31E-04	6.57E-04
MW	1.64E-03				1.42E-03				1.25E-03			

AR

	Fast				Medium speed				Slow			
Method \ m	5	10	20	A	5	10	20	A	5	10	20	A
LBF	1.22E-01	7.53E-02	4.59E-02	3.58E-02	4.83E-02	2.42E-02	4.80E-02	2.39E-02	**1.90E-02**	2.67E-02	5.17E-02	2.09E-02
fLBF	1.04E-01	7.02E-02	4.59E-02	3.17E-02	4.44E-02	2.58E-02	4.67E-02	2.52E-02	2.28E-02	2.91E-02	5.09E-02	2.61E-02
RLBF	1.16E-01	6.86E-02	3.03E-02	**2.70E-02**	4.68E-02	**1.92E-02**	2.98E-02	1.94E-02	1.93E-02	2.19E-02	3.34E-02	1.97E-02
fRLBF	1.02E-01	6.49E-02	3.17E-02	2.71E-02	4.40E-02	2.21E-02	3.07E-02	2.25E-02	2.46E-02	2.62E-02	3.55E-02	2.57E-02
MW	6.56E-02				3.83E-02				4.13E-02			

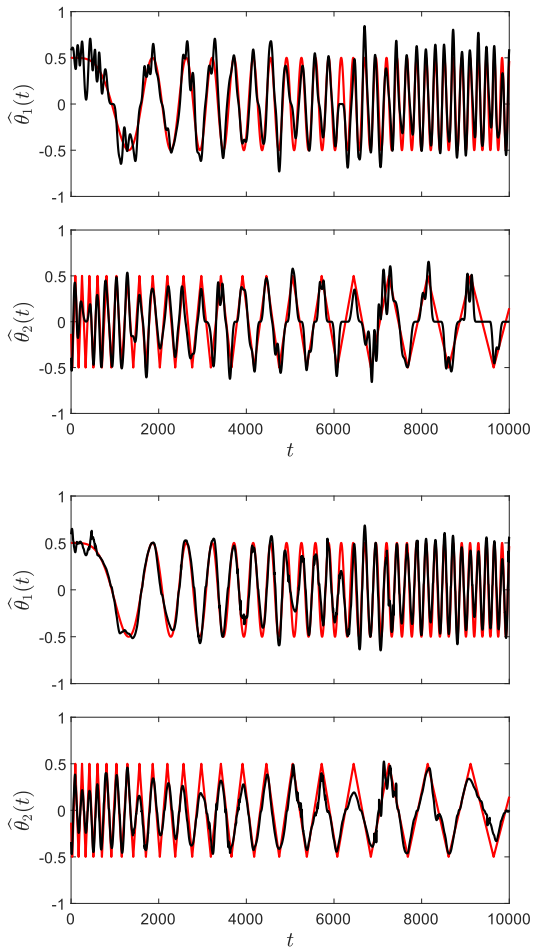

Fig. 5. Comparison of multi-scale wavelet estimates (two upper plots) and adaptive fLBF estimates (two lower plots) obtained for a nonstationary AR signal ($k = 250$, $m = 10$). Estimated trajectories (black lines) are superimposed on true trajectories (red lines).

the estimation results. The experiment was conducted for the polynomial basis (powers of time, $m = 5, 10, 20$), cosinusoidal basis ($m = 5, 10, 20$) and prolate spheroidal basis [45] (Slepian functions). In the latter case the basis set was enriched with a constant function and the corresponding time half bandwidth products were set to 2.5 ($m = 4 + 1$), 5 ($m = 9 + 1$) and 10 ($m = 19 + 1$), respectively.

The results obtained for order-adaptive versions of the LBF/RLBF and fLBF/fRLBF algorithms fixed n and k, $\widehat{m}(t) \in \{5, 10, 20\}$) are presented in Table IV. It is clear that if the basis is sufficiently "rich," no matter which one is chosen the obtained results are very similar and in all cases considerably better than those yielded by the MW approach.

Remark 3: It should be stressed that for some choices of the type and number of basis functions (e.g. polynomial basis and $m \geq 20$) the appropriate implementation of identification algorithms allows one to avoid excess numerical errors – see [46] for a more detailed discussion of this important issue. In our implementations we decided to use numerical procedures based on Cholesky decomposition of the inverted matrices. Slepian functions were generated using the MATLAB command `dpss`.

The third simulation experiment was focused on an interesting recent application – adaptive self-interference cancellation in full-duplex (FD) underwater acoustic (UWA) communication systems [38], [47]. FD UWA systems, designed to maximize the limited capacity of acoustic links, simultaneously transmit and receive data in the same frequency band. Due to the close spacing of the transmit and receive antennas, the far-end signal is strongly contaminated by the so-called self-interference introduced by the near-end transmitter. Self-interference is a multipath propagation effect caused, among others, by multiple reflections of the emitted signal from the water surface and/or the bottom.

MOST WIEDZY Downloaded from mostwiedzy.pl

TABLE IV
MEAN SQUARE PARAMETER ESTIMATION ERRORS OBTAINED FOR DIFFERENT ORDER-ADAPTIVE IDENTIFICATION ALGORITHMS EQUIPPED WITH DIFFERENT SETS OF BASIS FUNCTIONS: POLYNOMIAL (POWERS OF TIME), COSINUSOIDAL, AND SLEPIAN (PROLATE SPHEROIDAL). ALL AVERAGES WERE COMPUTED FOR 100 PROCESS REALIZATIONS, TWO DIFFERENT PROCESS MODELS (FIR AND AR), AND THREE SPEEDS OF PARAMETER VARIATION. THE BEST RESULTS IN EACH COLUMN ARE SHOWN IN BOLDFACE

FIR

	fast				medium				slow			
basis \method	fLBF	LBF	fRLBF	RLBF	fLBF	LBF	fRLBF	RLBF	fLBF	LBF	fRLBF	RLBF
polynomial	**2.40E-03**	**1.48E-03**	2.21E-03	**1.21E-03**	**1.19E-03**	**1.12E-03**	**1.09E-03**	**8.70E-04**	**7.56E-04**	**7.66E-04**	**6.57E-04**	**5.71E-04**
cosinusoidal	2.52E-03	1.83E-03	**2.13E-03**	1.35E-03	1.26E-03	1.24E-03	1.12E-03	9.41E-04	8.90E-04	9.24E-04	7.46E-04	6.64E-04
Slepian	2.72E-03	1.95E-03	2.51E-03	1.74E-03	1.53E-03	1.70E-03	1.38E-03	1.46E-03	1.23E-03	1.57E-03	1.01E-03	1.18E-03

AR

	fast				medium				slow			
basis \method	fLBF	LBF	fRLBF	RLBF	fLBF	LBF	fRLBF	RLBF	fLBF	LBF	fRLBF	RLBF
polynomial	**3.17E-02**	**3.58E-02**	2.71E-02	2.70E-02	**2.52E-02**	**2.39E-02**	**2.25E-02**	**1.94E-02**	**2.61E-02**	**2.09E-02**	**2.57E-02**	**1.97E-02**
cosinusoidal	3.34E-02	3.73E-02	**2.61E-02**	**2.65E-02**	2.84E-02	2.74E-02	2.34E-02	2.08E-02	2.99E-02	2.48E-02	2.70E-02	2.12E-02
Slepian	3.35E-02	3.76E-02	2.71E-02	2.83E-02	2.83E-02	2.69E-02	2.42E-02	2.16E-02	2.92E-02	3.08E-02	2.69E-02	2.47E-02

Following [47], the signal $y(t)$ measured by the receive antenna was modeled as the output of the 50-tap FIR filter

$$y(t) = \sum_{j=1}^{50} \theta_j(t) u(t-j+1) + e(t)$$

where $u(t)$ denotes the emitted (near-end) signal and $e(t)$ denotes a mixture of the far-end signal and the channel noise. For simplicity, all signals and parameters were assumed to be real-valued (extension to the complex-valued case is straightforward). Time variation of channel coefficients is caused by the transmitter/receiver motion and/or by the changes in the propagation medium. Note that in the case of FD UWA communication the underlying goal of identification is not channel inversion (since the signal $\{u(t)\}$ is known), but extraction of the signal $\{e(t)\}$ from $\{y(t)\}$. This can be easily achieved if the instantaneous impulse response of the channel $\{\theta_j(t), j = 1, \ldots, 50\}$ is known or can be estimated with sufficient accuracy.

An interesting feature of this application is that it allows one to work with a decision delay, which means that estimation of channel parameters can be based not only on past signal samples but also on a certain number of "future" (with respect to the moment of interest) ones. Hence, channel identification can be carried out using noncausal estimation algorithms, such as the ones described in this paper.

In accordance with [47], channel coefficients were modeled as lowpass signals

$$\theta_j(t) = c_j \zeta_j(t), \quad j = 1, \ldots, 50$$

where $\{\zeta_j(t)\}, j = 1, \ldots, 50$, denote mutually independent unity-variance signals obtained by passing white noise through a lowpass filter with cutoff frequency $f_{\max} = 0.001$. Under 1 kHz sampling the corresponding bandwidth of channel coefficient variation is 1 Hz, which can be regarded as fast changes in the UWA case. The mutually independent random scaling coefficients $c_j, j = 1, \ldots, 50$ were normally distributed $c_j \sim \mathcal{N}(0, \sigma_j^2)$, with decreasing variance profile $\sigma_j^2 = (0.69)^{j-1}$ which reflects the decaying power delay profile caused by the spreading and absorbtion loss; under such settings the ratio between the variance of the latest arrival ($j = 50$) and that of the first arrival ($j = 1$) is equal to 80 dB. Fig. 6 shows typical trajectories of selected impulse response coefficients (for a fixed pattern of scaling coefficients) and Fig. 7 displays a typical realization of the instantaneous impulse response (for fixed t).

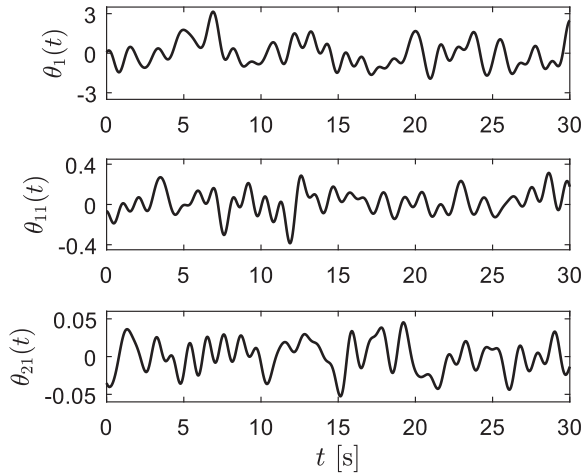

Fig. 6. Trajectories of the first, eleventh and twenty first parameter of the simulated underwater acoustic channel.

The generated input signal was white binary $u(t) = \pm 1$ and the signal $e(t)$ was zero-mean white Gaussian with variance σ_e^2 equal to 0.3, 0.03, 0.003 and 0.0003, which corresponds to the signal-to-noise ratio equal to 10 dB, 20 dB, 30 dB and 40 dB, respectively. Identification was carried out using the LBF, fLBF, RLBF and fRLBF algorithms (polynomial basis, rectangular window) and their order-adaptive versions. The estimation design parameters were set to $k = 250$, $m \in \{1, 3, 5\}$ and $\lambda_0 = 0.96$.

Performance was evaluated in terms of the normalized root mean squared error measure of fit used in [24]

$$\mathrm{FIT}(t) = 100 \left(1 - \left[\frac{\sum_{j=1}^{50} |\theta_j(t) - \widehat{\theta}_j(t)|^2}{\sum_{j=1}^{50} |\theta_j(t) - \bar{\theta}(t)|^2} \right]^{1/2} \right) \quad (72)$$

MOST WIEDZY Downloaded from mostwiedzy.pl

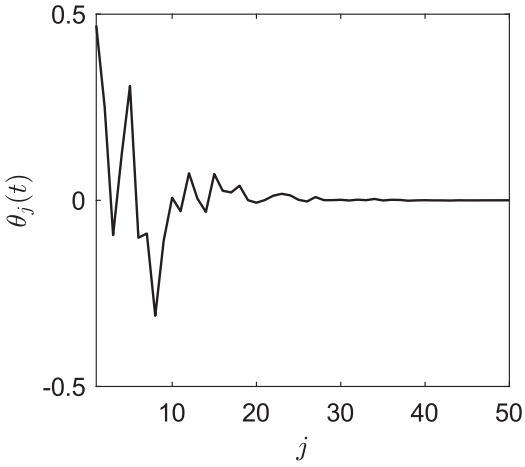

Fig. 7. A snapshot of the time-varying impulse response of the simulated underwater communication channel.

TABLE V
FIT[%] SCORES OBTAINED FOR 4 SIGNAL-TO-NOISE RATIOS (SNR) AND DIFFERENT NUMBER OF BASIS FUNCTIONS (m) FOR THE ALGORITHMS DESCRIBED IN THE TEXT AND FOR THEIR ORDER-ADAPTIVE VERSIONS (A)

SNR	m	LBF	RLBF	fLBF	fRLBF
10 dB	A	76.0	80.6	71.4	75.7
	1	66.8	69.8	71.2	72.4
	3	76.6	81.1	74.0	78.1
	5	68.5	74.4	68.2	74.3
20 dB	A	91.2	92.6	88.6	89.6
	1	70.2	72.2	76.4	76.5
	3	92.2	93.3	89.0	90.1
	5	90.0	91.7	87.2	89.1
30 dB	A	96.9	97.3	92.1	92.3
	1	70.6	72.5	77.0	77.0
	3	96.5	96.9	92.0	92.3
	5	96.9	97.3	91.3	92.1
40 dB	A	99.0	99.1	92.5	92.6
	1	70.7	72.5	77.1	77.0
	3	97.4	97.6	92.4	92.6
	5	99.0	99.1	91.8	92.4

where $\bar{\theta}(t) = \frac{1}{50}\sum_{j=1}^{50}\theta_j(t)$. The maximum value of FIT, equal to 100, corresponds to the perfect match between the true and estimated impulse response. The final scores were obtained by combined time averaging (10 000 time steps) and ensemble averaging (20 randomly drawn sets of scaling coefficients).

According to the results summarized in in Table V, regularization improves channel identification scores in all cases considered. As expected, the largest performance improvements are observed for small values of SNR. Figure 8 shows FIT scores obtained for the order-adaptive algorithms (SNR=10 dB) for all 20 realizations of scaling coefficients. Note that regularized algorithms RLBF/fRLBF yield *consistently* better results than their not regularized versions LBF/fLBF, i.e., better not only in the mean sense but also for *every* process realization.

The times needed to execute a single identification step (single time update) using a computer equipped with the Intel Core i7 2.2 GHz processor (4 cores) were equal to 3.6 ms, 0.24 ms, 4.5 ms and 0.26 ms for the order-adaptive versions of the algorithms LBF, fLBF, RLBF and fRLBF, respectively.

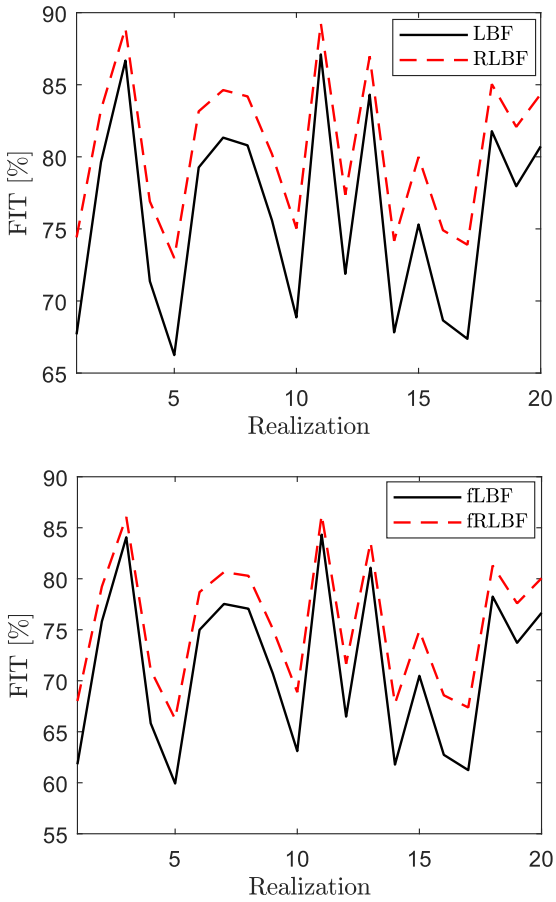

Fig. 8. FIT scores obtained for all 20 process realizations.

X. CONCLUSION

It was shown that performance of the local basis function estimators of linear time-varying systems/signals can be improved using the regularization technique. The proposed identification scheme is two-stage. At the first stage estimation is carried out using the standard basis function approach. At the second stage the regularized estimates are evaluated after replacing the unknown optimal regularization hyperparameters with their estimates obtained at the first stage. Finally, the computationally fast and numerically robust version of the regularized estimator is derived. It is shown that the resulting fast regularized local basis function algorithm, in spite of its simplicity, has very good parameter tracking capabilities, comparable with those of the original scheme. If some prior knowledge on the evolution of process parameters is available, e.g. if the correlation matrix $\mathrm{E}[\boldsymbol{\theta}(t)\boldsymbol{\theta}^{\mathrm{T}}(t)]$ is constant and known, as in some underwater communication applications, regularization can be carried out in a more sophisticated way. Such an application-oriented approach seems to be a promising direction for future research.

MOST WIEDZY Downloaded from mostwiedzy.pl

APPENDIX I
[OUTLINE OF DERIVATION OF (23)]

Let

$$\mathbf{Q}_{m|k}(t) = \mathbf{P}_{m|k}(t) - \mathrm{E}[\mathbf{P}_{m|k}(t)] =$$

$$\sum_{i=-k}^{k} w_k(i)[\boldsymbol{\varphi}(t+i)\boldsymbol{\varphi}^{\mathrm{T}}(t+i)] \otimes [\mathbf{f}_{m|k}(i)\mathbf{f}_{m|k}^{\mathrm{T}}(i)] - \boldsymbol{\Phi}_0 \otimes \mathbf{I}_m$$

$$= \sum_{i=-k}^{k} w_k(i)[\boldsymbol{\varphi}(t+i)\boldsymbol{\varphi}^{\mathrm{T}}(t+i) - \boldsymbol{\Phi}_0] \otimes [\mathbf{f}_{m|k}(i)\mathbf{f}_{m|k}^{\mathrm{T}}(i)].$$

Consider any element of the matrix $\mathbf{Q}_{m|k}(t)$, namely

$$q_{m|k}(t) = [\mathbf{Q}_{m|k}(t)]_{n(j_1-1)+k_1, n(j_2-1)+k_2}$$

$$= \sum_{i=-k}^{k} w_k(i) f_{j_1|k}(i) f_{j_2|k}(i) \times$$

$$\times [u(t+i-k_1+1)u(t+i-k_2+1) - [r_u(k_1-k_2)],$$

where $1 \le k_1, k_2 \le n, 1 \le j_1, j_2 \le m$ and $\{r_u(\tau)\}$ denotes the autocorrelation function of $\{u(t)\}$.

According to the assumption (A2) there exist constants $c_1 > 0$ and $\zeta \in (0,1)$ such that $|r_u(\tau)| \le c_1\zeta^{|\tau|}, \ \forall \tau$. Since for any zero-mean jointly Gaussian variables $x_1, \ldots, x_4$ it holds that [40]

$$\mathrm{E}[x_1x_2x_3x_4] = \mathrm{E}[x_1x_2]\mathrm{E}[x_3x_4] + \mathrm{E}[x_1x_3]\mathrm{E}[x_2x_4] + \mathrm{E}[x_1x_4]\mathrm{E}[x_2x_3],$$

one obtains

$$\mathrm{E}[q_{m|k}^2(t)]$$

$$= \sum_{i_1=-k}^{k} \sum_{i_2=-k}^{k} w_k(i_1)w_k(i_2) f_{j_1|k}(i_1) f_{j_2|k}(i_1) f_{j_1|k}(i_2) f_{j_2|k}(i_2)$$

$$\times [r_u^2(i_1-i_2) + r_u(i_1-i_2+k_1-k_2) r_u(i_1-i_2+k_2-k_1)].$$

It can be shown that if the basis and window functions are obtained by sampling their continuous-time prototypes, there exist constants $c_2 > 0$ and $k_0 \ge 1$ such that for every $k \ge k_0$ [19]

$$\left|f_{j|k}(i)\right| \le \frac{c_2}{\sqrt{k}}, \quad \forall i, 1 \le j \le m.$$

Using this upper bound, one arrives at

$$\mathrm{E}[q_{m|k}^2(t)] \le \frac{2c_1^2c_2^4}{k^2} \sum_{i_1=-k}^{k} \sum_{i_2=-k}^{k} \zeta^{2|i_1-i_2|} = O(1/k), \ \forall k \ge k_0,$$

which means that $q_{m|k}(t)$ converges to zero in the mean square sense (and hence also in probability) as k grows to infinity.

REFERENCES

[1] M. K. Tsatsanis and G. B. Giannakis, "Modelling and equalization of rapidly fading channels," *Int. J. Adaptive Control Signal Process.*, vol. 310, pp. 515–522, 1996.

[2] S. G. Fabri, K. P. Camilleri, and T. Cassar, "Parametric modelling of EEG data for the identification of mental tasks," *Biomed. Eng. Trends Electron., Commun. & Softw.*, pp. 367–386, 2011.

[3] T. Wada, M. Jinnouchi, and Y. Matsumura, "Application of autoregressive modelling for the analysis of clinical and other biological data," *Ann. Inst. Statist. Math.*, vol. 40, no. 2, pp. 211–227, 1998.

[4] P. Lesage, F. Glangeaud, and J. Mars, "Applications of autoregressive models and time–frequency analysis to the study of volcanic tremor and long-period events," *J. Volcanology Geothermal Res.*, vol. 114, pp. 391–417, 2002.

[5] C. Li and R. L. Nowack, "Application of autoregressive extrapolation to seismic tomography," *Bull. Seismological Soc. Amer.*, vol. 94, no. 4, pp. 1456–1466, 2004.

[6] K. Aström, and B. Wittenmark, *Adaptive Control*. New York, NY, USA: Dover Books, 2008.

[7] J. P. Norton, "Optimal smoothing in the identification of linear time-varying systems," *Proc. IEEE*, vol. 122, no. 6, pp. 663–668, 1975.

[8] G. Kitagawa and W. Gersch, "A smoothness priors time-varying AR coefficient modeling of nonstationary covariance time series," *IEEE Trans. Autom. Control*, vol. AC-30, no. 1, pp. 48–56, Jan. 1985.

[9] G. Kitagawa and W. Gersch, *Smoothness Priors Analysis of Time Series*. Berlin, Germany: Springer-Verlag, 1996.

[10] M. Niedźwiecki, "Locally adaptive cooperative Kalman smoothing and its application to identification of nonstationary stochastic systems," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 48–59, Jan. 2012.

[11] 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.

[12] Y. Grenier, "Time-dependent ARMA modeling of nonstationary signals," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-31, no. 4, pp. 899–911, Aug. 1981.

[13] M. Niedźwiecki, "Functional series modeling approach to identification of nonstationary stochastic systems," *IEEE Trans. Autom. Control*, vol. 33, no. 10, pp. 955–961, Oct. 1988.

[14] J. M. Liporace, "Linear estimation of nonstationary signals," *J. Acoustical Soc. Amer.*, vol. 58, pp. 1288–1295, 1975.

[15] M. K. Tsatsanis and G. B. Giannakis, "Time-varying system identification and model reduction using wavelets," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3512–3523, Dec. 1994.

[16] H. L. Wei, J. J. Liu, and S. A. Billings, "Identification of time-varying systems using multi-resolution wavelet models," *Int. J. Syst. Sci.*, vol. 33, pp. 1217–1228, 2002.

[17] H. L. Wei, S. A. Billings, and J. J. Liu, "Time-varying parametric modelling and time-dependent spectral characterisation with application to EEG signals using multiwavelets," *Int. J. Modelling, Identification Control*, vol. 9, pp. 215–224, 2010.

[18] M. Niedźwiecki, and M. Ciołek, "Generalized savitzky-golay filters for identification of nonstationary systems," *Automatica*, vol. 108, 2019, Art. no. 108522.

[19] M. Niedźwiecki, M. Ciołek, and A. Gańcza, "A new look at the statistical identification of nonstationary systems," *Automatica*, vol. 118, 2020, Art. no. 109037.

[20] D. L. Phillips, "A technique for the numerical solution of certain integral equations of the first kind," *J. ACM*, vol. 9, pp. 84–97, 1962.

[21] A. Tikhonov, and V. Arsenin, *Solutions of Ill-Posed Problems*. Washington, DC, USA: Winston/Wiley, 1977.

[22] A. E. Hoerl, and W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, pp. 55–67, 1970.

[23] L. Elden, "Algorithms for the regularization of ill-conditioned least squares problems," *BIT Numer. Math.*, vol. 17, pp. 134–145, 1977.

[24] L. Ljung, T. Chen, "What can regularization offer for estimation of dynamical systems?," in *Proc. 11th IFAC Workshop Adapt. Learn. Control Signal Process.*, Caen, France, pp. 1–8, 2013.

[25] L. Ljung, T. Chen, and B. Mu, "A shift in paradigm for system identification," *Int. J. Control*, vol. 93, pp. 173–180, 2020.

[26] B. Efron, "Biased versus unbiased estimation," *Adv. Math.*, vol. 16, pp. 259–277, 1975.

[27] Y. C. Eldar, "Uniformly improving the Cramer-Rao bound and maximum-likelihood estimation," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 2943–2956, Aug. 2006.

[28] E. T. Whittaker, "On a new method of graduation," in *Proc. Edinborough Math. Assoc.*, vol. 78, pp. 81–89, 1923.

[29] E. T. Whittaker, and G. Robinson, *Calculus of Observations: A Treatise on Numerical Mathematics*. Glasgow, U.K.: Blackie and Son, 1924.

[30] W. Gersch, and G. Kitagawa, "Smoothness priors transfer function estimation," *Automatica*, vol. 25, pp. 603–608, 1989.

MOST WIEDZY Downloaded from mostwiedzy.pl

[31] T. Chen, H. Ohlsson and L. Ljung, "On the estimation of transfer functons, regularizations and Gaussian processes—Revisited," *Automatica*, vol. 48, pp. 1525–1535, 2012.
[32] A. Y. Ng, "Feature selection, L_1 vs. L_2 regularization, and rotational invariance," in *Proc. 21st Int. Conf. Mach. Learn.*, pp. 78–85, 2004.
[33] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. R. Statist. Soc. B*, vol. 58, pp. 267–288, 1996.
[34] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Berlin, Germany: Springer, 2013.
[35] B. Mu, T. Chen and L. Ljung, "Asymptotic properties of generalized cross-validation estimators for regularized system identification," *IFAC-PapersOnLine*, vol. 51, pp. 203–208, 2018.
[36] B. Mu, T. Chen and L. Ljung, "On asymptotic properties of hyperparameter estimators for kernel-based regularization metohds," *Automatica*, vol. 94, pp. 381–395, 2018.
[37] Y. Li, H. L. Wei, S. A. Billings and P. Sarrigiannis, "Time-varying model identification for time-frequency feature extraction from EEG data," *J. Neurosci. Methods*, vol. 196, pp. 151–158, 2011.
[38] L. Shen, B. Henson, Y. Zakharov and P. Mitchell, "Digital self-interference cancellation for underwater acoustic systems," *IEEE Trans. Circuits Syst. II: Exp. Briefs*, vol. 67, no. 1, pp. 192–196, Jan. 2020.
[39] R. W. Schafer, "What is a savitzky-golay filter?," *IEEE Signal Process. Mag.*, vol. 28, no. 4, pp. 111–117, Jul. 2011.
[40] T. Söderström and P. Stoica, System Identification. Englewood Cliffs, NJ, USA: Prentice-Hall, 1988.
[41] H. Akaike, "Fitting autoregressive models for prediction," *Ann. Inst. Statistical Math.*, vol. 21, pp. 243–247, 1969.
[42] H. Akaike, "Statistical predictor identification," *Ann. Inst. Statist. Math.*, vol. 22, pp. 203–217, 1970.
[43] M. Niedźwiecki and M. Ciołek, "Fully adaptive savitzky-golay type smoothers," in *Proc. 27th Eur. Signal Process. Conf.*, A Coruna, Spain, 2019, pp. 1–5.
[44] M. Niedźwiecki and M. Ciołek, "Akaike's final prediction error criterion revisited," in *Proc. 40th IEEE Int. Conf. Telecommun. Signal Process.*, Barcelona, Spain, 2017, pp. 237–242.
[45] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications*. Cambridge, U.K.: Cambridge Univ. Press, 1993.
[46] S. Chen, S. A. Billings and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *Int. J. Control*, vol. 50, pp. 1873–1896, 1989.
[47] L. Shen, B. Henson, Y. Zakharov, N. Morozs, and P. Mitchell, "Adaptive filtering for full-duplex UWA systems with time-varying self-interference channel," *IEEE Access*, vol. 8, pp. 187590–187604, 2020.

Artur Gańcza (Student Member, IEEE) received the M.Sc. degree in 2019 from the Gdańsk University of Technology, Gdańsk, Poland, where he is currently working toward the Ph.D. degree with the Department of Automatic Control, Faculty of Electronics, Telecommunications and Informatics. His research interests include speech recognition, system identification, and adaptive signal processing.

Maciej Niedźwiecki (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees from the Technical University of Gdańsk, Gdańsk, Poland, in 1977 and 1981, respectively, and the Dr.Hab. (D.Sc.) degree from the Technical University of Warsaw, Warsaw, Poland, in 1991. From 1986 to 1989, he spent three years as a Research Fellow with the Department of Systems Engineering, Australian National University, Canberra, ACT, Australia . From 1990 to 1993, he was the Vice Chairman of Technical Committee on Theory of the International Federation of Automatic Control (IFAC). He is currently a Professor and the Head of the Department of Automatic Control, Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, Gdańsk, Poland.

He is the author of the book *Identification of Time-varying Processes* (Wiley, 2000). His 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.

Marcin Ciołek (Member, IEEE) 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 an Adjunct Professor with the Department of Automatic Control, Faculty of Electronics, Telecommunications and Informatics, GUT. His research interests include speech, music, and biomedical signal processing.

MOST WIEDZY Downloaded from mostwiedzy.pl