Automatica 133 (2021) 109865

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

Application of regularized Savitzky–Golay filters to identification of time-varying systems☆

Maciej Jan Niedźwiecki *, Marcin Ciołek, Artur Gańcza, Piotr Kaczmarek

Faculty of Electronics, Telecommunications and Informatics, Department of Automatic Control, Gdańsk University of Technology, Narutowicza 11/12, 80-233 Gdańsk, Poland

ARTICLE INFO

Article history:
Received 12 October 2020
Received in revised form 11 May 2021
Accepted 1 July 2021
Available online 23 August 2021

Keywords:
Savitzky–Golay filters
Identification of time-varying systems
Basis functions
Empirical Bayes optimization

ABSTRACT

Savitzky–Golay (SG) filtering is a classical signal smoothing technique based on the local least squares approximation of the analyzed signal by a linear combination of known functions of time (originally — powers of time, which corresponds to polynomial approximation). It is shown that the regularized version of the SG algorithm can be successfully applied to identification of time-varying finite impulse response (FIR) systems. Such a solution is possible owing to the recently proposed preestimation technique, which converts the problem of identification of a time-varying FIR system into the problem of smoothing of the appropriately generated preestimates of system parameters. The resulting fast regularized local basis function estimators, optimized using the empirical Bayes approach, have very good parameter tracking capabilities, favorably comparing with the state-of-the-art in terms of accuracy, computational complexity and numerical robustness.

1. Introduction

Savitzky–Golay (SG) filtering, i.e., filtering successive subsets of adjacent data points with a low degree polynomial by the method of least squares, is a classical smoothing technique (Savitzky & Golay, 1964; Schafer, 2011). Due to their analytical and computational simplicity, and good smoothing capabilities, SG filters have been extensively used in such research areas as spectroscopy (Li, Deng, Li, & Yu, 2015; Turton, 1992), voltammetry (Jakubowska & Kubiak, 2004) and biomedical signal processing (Acharya, Rani, Agarwal, & Singh, 2016; Goel, Kaur, & Tomar, 2016; Hargittai, 2005), among many others.

The idea behind SG filtering was recently extended, in a generalized form, to identification of time-varying systems (Niedźwiecki & Ciołek, 2019b). In the follow-up paper (Niedźwiecki, Ciołek, & Gańcza, 2020a) it was shown that almost identical results can be obtained by smoothing, using the SG filter, the sequence of parameter preestimates obtained by "inverse filtering" the estimates yielded by the short-memory exponentially weighted least squares algorithm. The resulting identification procedure has very good parameter tracking capabilities, favorably comparing with the state-of-the art multi-wavelet estimation scheme proposed in Wei, Liu, and Billings (2002) - see Niedźwiecki, Gańcza, and Ciołek (2020b). Additionally, it is computationally simple and numerically robust.

The paper aims to show that identification results can be further improved if preestimates are postfiltered using the "regularized" version of the SG smoother. Regularization is a well-established technique in machine learning and system identification. Although originally introduced as a way of solving ill-posed or numerically ill-conditioned inverse problems (Phillips, 1962; Tikhonov & Arsenin, 1977), it has more to offer (Ljung & Chen, 2013; Ljung, Chen, & Mu, 2020). First, when the regularization term, added to the minimized cost function, is tuned to the experimental data, the regularized solution can improve the estimation bias–variance trade-off which decides upon accuracy of the identified model. Secondly, regularization allows one to include in the formulation of the identification problem some expected, or desired, properties of the solution, such as smoothness, rate of decay etc. Both aspects of the problem are discussed in the paper. Regularization hyperparameters are tuned using the empirical Bayes approach (Akaike, 1980; Berger, 1985; Good, 1965). Simulation evidence confirms good properties of the proposed identification algorithms. The paper seems to be the first attempt to incorporate the regularization technique into identification of time-varying systems.

☆ This work was partially supported by the National Science Center, Poland under the agreement UMO-2018/29/B/ST7/00325. Computer simulations were carried out at the Academic Computer Centre in Gdańsk. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Juan I. Yuz under the direction of Editor Torsten Söderström.

* Corresponding author.
E-mail addresses: maciekn@eti.pg.edu.pl (M.J. Niedźwiecki), marcin.ciolek@pg.edu.pl (M. Ciołek), artgancz@student.pg.edu.pl (A. Gańcza), piokaczm@pg.edu.pl (P. Kaczmarek).

https://doi.org/10.1016/j.automatica.2021.109865

2. Identification of time-varying systems

2.1. Local basis function estimators

In this section we will consider the problem of identification of a time-varying FIR system governed by

$$y(t)=\sum_{j=1}^{n}\theta_j(t)u(t-j+1)+e(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t)+e(t) \tag{1}$$

where $y(t)$ denotes system output, $\boldsymbol{\varphi}(t)=[u(t),\ldots,u(t-n+1)]^{\mathrm{T}}$ denotes regression vector made up of past values of the observable input signal $u(t)$, and $e(t)$ denotes white measurement noise. Finally, $\boldsymbol{\theta}(t)=[\theta_1(t),\ldots,\theta_n(t)]^{\mathrm{T}}$ denotes the vector of time-varying system parameters. Note that the sequence $\theta_1(t),\ldots,\theta_n(t)$ can be interpreted as a time-varying impulse response of the system (1).

One of the challenging recent applications, which requires identification of a time-varying FIR system, is tracking and equalization of underwater acoustic (UWA) channels. Due to the multipath propagation of the transmitted signal, caused, among others, by its multiple reflections from the water surface and the bottom, impulse response of the UWA channel usually involves many coefficients (taps). Their time variation is the result of Doppler effect due to the transmitter/receiver and water motion (Stojanovic & Preisig, 2009). Therefore, to secure reliable UWA communication, on-line channel estimation is necessary (Kochańska, 2015; Li & Preisig, 2007). Additionally, since some tracking delay is acceptable, channel identification can be carried out using noncausal estimation algorithms, such as the ones considered in this paper. Another successful application of time-varying system identification techniques is the self-interference mitigation in full-duplex UWA systems (Qiao, Gan, Liu, Ma, & Sun, 2018; Shen, Henson, Zakharov, & Mitchell, 2020).

As recently shown in Niedźwiecki and Ciołek (2019b), estimation of fast time-varying parameters in (1) can be efficiently carried out using the local basis function (LBF) approach. Denote by $T_t=[t-k,t+k]$ the local analysis interval of width $K=2k+1$, centered at t, and let $\mathcal{F}=\{f_1(i),\ldots,f_m(i)\}, i\in I_k=[-k,k]$, be the set of m linearly independent basis functions which will be used to approximate (locally) the parameter trajectories. For convenience, but without any loss of generality, we will assume that basis functions are orthonormal, i.e., $\sum_{i=-k}^{k}\mathbf{f}(i)\mathbf{f}^{\mathrm{T}}(i)=\mathbf{I}_m$, where $\mathbf{f}(i)=[f_1(i),\ldots,f_m(i)]^{\mathrm{T}}$ and $\mathbf{I}_m$ denotes the $m\times m$ identity matrix. The typical choices of basis functions prior to orthonormalization are powers of time (local Taylor approximation) or cosine functions (local Fourier approximation).

The idea behind the LBF approach is to repeatedly, for every position of the sliding analysis window T_t, apply the functional series approximation

$$\begin{aligned}\theta_j(t+i)&=\sum_{l=1}^{m}f_l(i)a_{jl}=\mathbf{f}^{\mathrm{T}}(i)\boldsymbol{\alpha}_j\\ &i\in I_k,\quad j=1,\ldots,n\\ \boldsymbol{\alpha}_j&=[a_{j1},\ldots,a_{jm}]^{\mathrm{T}}\end{aligned} \tag{2}$$

which in the vector form can be written down as

$$\begin{aligned}\boldsymbol{\theta}(t+i)&=\mathbf{F}(i)\boldsymbol{\alpha},\quad i\in I_k\\ \boldsymbol{\alpha}&=[\boldsymbol{\alpha}_1^{\mathrm{T}},\ldots,\boldsymbol{\alpha}_n^{\mathrm{T}}]^{\mathrm{T}}\end{aligned} \tag{3}$$

where the $n\times mn$ matrix $\mathbf{F}(i)$ is given by

$$\mathbf{F}(i)=\mathbf{I}_n\otimes\mathbf{f}^{\mathrm{T}}(i) \tag{4}$$

and $\otimes$ denotes the Kronecker product of the corresponding vectors/matrices. The parameter estimates are obtained using the method of least squares

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}^{\mathrm{LBF}}(t)&=\arg\min_{\boldsymbol{\alpha}}\sum_{i=-k}^{k}[y(t+i)-\boldsymbol{\psi}^{\mathrm{T}}(t,i)\boldsymbol{\alpha}]^2\\ \widehat{\boldsymbol{\theta}}^{\mathrm{LBF}}(t)&=\mathbf{F}_0\widehat{\boldsymbol{\alpha}}^{\mathrm{LBF}}(t)\end{aligned} \tag{5}$$

where $\mathbf{F}_0=\mathbf{F}(0)=\mathbf{I}_n\otimes\mathbf{f}_0^{\mathrm{T}}$, $\mathbf{f}_0=\mathbf{f}(0)$ and $\boldsymbol{\psi}(t,i)=\mathbf{F}^{\mathrm{T}}(i)\boldsymbol{\varphi}(t+i)=\boldsymbol{\varphi}(t+i)\otimes\mathbf{f}(i)$ denotes the generalized regression vector. The procedure is repeated for consecutive values of t, i.e., the estimation is carried out in the sliding window mode.

Due to its conceptual similarity to the SG scheme, the resulting estimation algorithm was called in Niedźwiecki and Ciołek (2019b) the generalized Savitzky–Golay filter. The generalized SG algorithm has superb parameter tracking capabilities, even if parameter changes are fast, but is computationally very demanding, as it requires inversion of the $mn\times mn$ generalized regression matrix $\sum_{i=-k}^{k}\boldsymbol{\psi}(t,i)\boldsymbol{\psi}^{\mathrm{T}}(t,i)$ every time step t.

2.2. Preestimation technique

As demonstrated in the follow-up paper (Niedźwiecki et al., 2020a), the computational complexity of LBF estimators can be significantly reduced, without compromising their tracking properties, using the preestimation technique. Following (Niedźwiecki et al., 2020a), we will assume that:

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

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

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

We note that assumptions (A1)–(A3) are fulfilled in typical channel identification applications — see e.g. Shen et al. (2020). The preestimated parameter trajectory $\{\boldsymbol{\theta}^*(t)\}$ can be obtained by "inverse filtering" the estimates yielded by the short-memory exponentially weighted least squares estimates

$$\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t)=\arg\min_{\boldsymbol{\theta}}\sum_{i=0}^{t-1}\lambda_0^i[y(t-i)-\boldsymbol{\varphi}^{\mathrm{T}}(t-i)\boldsymbol{\theta}]^2 \tag{6}$$

namely

$$\boldsymbol{\theta}^*(t)=L_t\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t)-\lambda_0L_{t-1}\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1) \tag{7}$$

where λ_0, $0<\lambda_0<1$, denotes the so-called forgetting constant and $L_t=\sum_{i=0}^{t-1}\lambda_0^i=\lambda_0L_{t-1}+1$, $L_0=1$, denotes the effective width of the exponential window. Note that for large values of t, when the effective window width reaches its steady state value $L_\infty=1/(1-\lambda_0)$, the formula (7) can be replaced with

$$\boldsymbol{\theta}^*(t)=\frac{1}{1-\lambda_0}\left[\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t)-\lambda_0\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1)\right] \tag{8}$$

According to Niedźwiecki et al. (2020a), when the system obeys assumptions (A1)–(A3), the preestimates defined in this way are approximately unbiased, i.e.,

$$\boldsymbol{\theta}^*(t)=\boldsymbol{\theta}(t)+\mathbf{z}(t) \tag{9}$$

where $\mathbf{z}(t)=[z_1(t),\ldots,z_n(t)]^{\mathrm{T}}$ denotes (approximately) a zero-mean white noise with a large covariance matrix. The best preestimation results can be obtained for small values of L_∞. The rule of thumb, which works pretty well in practice, is to choose $\lambda_0=\max\{0.9,1-\frac{2}{n}\}$. When $\lambda_0=1-\frac{2}{n}$, the equivalent width of the exponential window $N_\infty=(1+\lambda_0)/(1-\lambda_0)\cong 2/(1-\lambda_0)$, different from its effective width L_∞ (Niedźwiecki, 2000), is approximately equal to the number of estimated coefficients n.

2.3. Fast local basis function estimators

Since preestimates have a very large variability, they must be further processed to obtain statistically reliable parameter estimates. This can be done using the basis function approach. The resulting estimates, called in Niedźwiecki et al. (2020a) the fast local basis function (fLBF) estimates, have the form

$$\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t) = \arg\min_{\boldsymbol{\alpha}} \sum_{i=-k}^{k} \|\boldsymbol{\theta}^*(t+i) - \mathbf{F}(i)\boldsymbol{\alpha}\|^2$$
$$= \sum_{i=-k}^{k} \boldsymbol{\theta}^*(t+i) \otimes \mathbf{f}(i)$$
$$\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t) = \mathbf{F}_0 \widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t) = \sum_{i=-k}^{k} \mathbf{f}_0^{\mathrm{T}} \mathbf{f}(i) \boldsymbol{\theta}^*(t+i) \tag{10}$$

where $\mathbf{f}_0^{\mathrm{T}}\mathbf{f}(i), i \in I_k$ is the impulse response of a postprocessing FIR filter. As shown in Niedźwiecki et al. (2020a), it holds that $\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t) \cong \widehat{\boldsymbol{\theta}}^{\mathrm{LBF}}(t)$. As a matter of fact, in most cases the estimated parameter trajectories obtained using the fLBF approach are almost indistinguishable from those yielded by the – computationally much more involved – LBF approach.

According to (10), the estimate $\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t)$ is nothing but the result of smoothing the preestimated parameter trajectory $\{\boldsymbol{\theta}^*(t)\}$ using the Savitzky–Golay filter.

2.4. Prior-aware fast regularized local basis function estimators

Whenever some prior knowledge of the estimated impulse response is available, it may be beneficial to expand the measure of fit by a regularization term $\mu\|\boldsymbol{\theta}(t)\|^2_{\mathbf{R}} = \mu\boldsymbol{\theta}^{\mathrm{T}}(t)\mathbf{R}\boldsymbol{\theta}(t)$, where $\mu > 0$ denotes regularization gain and $\mathbf{R} \geq 0$ is the $n \times n$ positive semidefinite regularization matrix chosen so as to promote some known/desired features of the solution, such as its smoothness, rate of the decay etc. Selection of the adequate regularization matrix for identification of time-invariant systems has attracted a great deal of attention in recent years (Pillonetto, Dinuzzo, Chen, De Nicolao, & Ljung, 2014). A more detailed comment on this problem will be provided in the next section. For the time being we will assume that $\mathbf{R} = \mathbf{D}^{\mathrm{T}}\mathbf{D}$, where $\mathbf{D}$ is a $l \times n$, $l \leq n$ matrix. When $l < n$ the regularization matrix is singular. Without any loss of generality we will assume that $\mathbf{D}\mathbf{D}^{\mathrm{T}} > 0$.

When $\boldsymbol{\theta}(t) = \mathbf{F}_0\boldsymbol{\alpha}$, it holds that $\|\boldsymbol{\theta}(t)\|^2_{\mathbf{R}} = \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{B}\boldsymbol{\alpha} = \|\boldsymbol{\alpha}\|^2_{\mathbf{B}^{\mathrm{T}}\mathbf{B}}$, where $\mathbf{B} = \mathbf{D}\mathbf{F}_0 = \mathbf{D} \otimes \mathbf{f}_0^{\mathrm{T}}$ is the $l \times mn$ matrix. Note that, unlike the classical regularization approach, the regularization term penalizes the norm of $\boldsymbol{\theta}(t)$, rather than the norm of $\boldsymbol{\alpha}$ which is not of our primary interest. Due to singularity of the matrix $\mathbf{f}_0\mathbf{f}_0^{\mathrm{T}}$, the regularization matrix $\mathbf{B}^{\mathrm{T}}\mathbf{B}$ is also singular (even if $\mathbf{R}$ is nonsingular), which is another nonstandard feature.

Fast regularized local basis function (fRLBF) estimators can be defined in the following way

$$\widehat{\boldsymbol{\alpha}}^{\mathrm{fRLBF}}(t) =$$
$$= \arg\min_{\boldsymbol{\alpha}} \left\{ \sum_{i=-k}^{k} \|\boldsymbol{\theta}^*(t+i) - \mathbf{F}(i)\boldsymbol{\alpha}\|^2 + \mu\|\boldsymbol{\alpha}\|^2_{\mathbf{B}^{\mathrm{T}}\mathbf{B}} \right\}$$
$$= [\mathbf{I}_{mn} + \mu\mathbf{B}^{\mathrm{T}}\mathbf{B}]^{-1}\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t)$$
$$\widehat{\boldsymbol{\theta}}^{\mathrm{fRLBF}}(t) = \mathbf{F}_0\widehat{\boldsymbol{\alpha}}^{\mathrm{fRLBF}}(t) \tag{11}$$

where $\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t)$ is given by (10). We will prove that

Lemma 1. *The fRLBF estimate given by (11) can be expressed in the form*

$$\widehat{\boldsymbol{\theta}}^{\mathrm{fRLBF}}(t) = [\mathbf{I}_n + \mu\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{R}]^{-1}\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t). \tag{12}$$

Proof. Using the Woodbury matrix identity (Söderström & Stoica, 1988)

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}[\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B}]^{-1}\mathbf{D}\mathbf{A}^{-1}$$

(assuming that all inverses exist), one obtains

$$[\mathbf{I}_{mn} + \mu\mathbf{B}^{\mathrm{T}}\mathbf{B}]^{-1} = \mathbf{I}_{mn} - \mu\mathbf{B}^{\mathrm{T}}[\mathbf{I}_l + \mu\mathbf{B}\mathbf{B}^{\mathrm{T}}]^{-1}\mathbf{B} \tag{13}$$

Using the well-known property of the Kronecker product: $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A}\mathbf{C}) \otimes (\mathbf{B}\mathbf{D})$ (provided that all dimensions match), one obtains

$$\mathbf{B}\mathbf{B}^{\mathrm{T}} = (\mathbf{D} \otimes \mathbf{f}_0^{\mathrm{T}})(\mathbf{D}^{\mathrm{T}} \otimes \mathbf{f}_0) = \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{D}\mathbf{D}^{\mathrm{T}}. \tag{14}$$

Hence

$$[\mathbf{I}_{mn} + \mu\mathbf{B}^{\mathrm{T}}\mathbf{B}]^{-1}$$
$$= \mathbf{I}_{mn} - \mu[\mathbf{D}^{\mathrm{T}} \otimes \mathbf{f}_0][\mathbf{I}_l + \mu\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{D}\mathbf{D}^{\mathrm{T}}]^{-1}[\mathbf{D} \otimes \mathbf{f}_0^{\mathrm{T}}] \tag{15}$$
$$= \mathbf{I}_{mn} - \left\{\mu\mathbf{D}^{\mathrm{T}}[\mathbf{I}_l + \mu\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{D}\mathbf{D}^{\mathrm{T}}]^{-1}\mathbf{D}\right\} \otimes [\mathbf{f}_0\mathbf{f}_0^{\mathrm{T}}].$$

Observe that

$$[\mathbf{I}_n + \mu\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{D}^{\mathrm{T}}\mathbf{D}]^{-1}$$
$$= \mathbf{I}_n - \mu\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{D}^{\mathrm{T}}[\mathbf{I}_l + \mu\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{D}\mathbf{D}^{\mathrm{T}}]^{-1}\mathbf{D} \tag{16}$$

Combining (15) and (16), and noting that $\mathbf{D}^{\mathrm{T}}\mathbf{D} = \mathbf{R}$, one arrives at

$$[\mathbf{I}_{mn} + \mu\mathbf{B}^{\mathrm{T}}\mathbf{B}]^{-1} = \mathbf{I}_{mn} - \mathbf{I}_n \otimes \left[\frac{\mathbf{f}_0\mathbf{f}_0^{\mathrm{T}}}{\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}\right]$$
$$+ [\mathbf{I}_n + \mu\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{R}]^{-1} \otimes \left[\frac{\mathbf{f}_0\mathbf{f}_0^{\mathrm{T}}}{\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}\right] \tag{17}$$

and

$$\widehat{\boldsymbol{\alpha}}^{\mathrm{fRLBF}}(t) = \widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t) - \widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t) \otimes \left[\frac{\mathbf{f}_0}{\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}\right]$$
$$+ \left\{ [\mathbf{I}_n + \mu\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{R}]^{-1}\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t) \right\} \otimes \left[\frac{\mathbf{f}_0}{\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}\right]. \tag{18}$$

Finally, combining $\widehat{\boldsymbol{\theta}}^{\mathrm{fRLBF}}(t) = (\mathbf{I}_n \otimes \mathbf{f}_0^{\mathrm{T}})\widehat{\boldsymbol{\alpha}}^{\mathrm{fRLBF}}(t)$ with (18), one arrives at (12). ■

According to Lemma 1, the fRLBF estimate can be obtained without evaluating $\widehat{\boldsymbol{\alpha}}^{\mathrm{fRLBF}}(t)$ by "shrinking" the ordinary fLBF estimate.

Evaluation of the fRLBF estimates can be easily performed in the matrix-inversion-free way. Actually, suppose that $l = n$, i.e., the regularization matrix $\mathbf{R}$ is nonsingular. As a symmetric positive definite matrix, $\mathbf{R}$ admits the factorization $\mathbf{R} = \mathbf{Q}\boldsymbol{\Lambda}_n\mathbf{Q}^{\mathrm{T}}$ where $\boldsymbol{\Lambda}_n = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}$ is a diagonal matrix made up of the eigenvalues of $\mathbf{R}$, and $\mathbf{Q}$, such that $\mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{Q}\mathbf{Q}^{\mathrm{T}} = \mathbf{I}_n$, is an orthogonal matrix made up of its normalized eigenvectors.

Note that

$$[\mathbf{I}_n + \mu\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{R}]^{-1} = [\mathbf{Q}(\mathbf{I}_n + \mu\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\boldsymbol{\Lambda}_n)\mathbf{Q}^{\mathrm{T}}]^{-1}$$
$$= \mathbf{Q}(\mathbf{I}_n + \mu\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\boldsymbol{\Lambda}_n)^{-1}\mathbf{Q}^{\mathrm{T}} \tag{19}$$

and

$$(\mathbf{I}_n + \mu\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\boldsymbol{\Lambda}_n)^{-1}$$
$$= \mathrm{diag}\left\{\frac{1}{1 + \mu\lambda_1\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}, \ldots, \frac{1}{1 + \mu\lambda_n\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}\right\}.$$

When $l < n$, i.e., the regularization matrix is singular, one can use the identity (16) and apply the same technique to invert the $l \times l$ matrix $\mathbf{I}_l + \mu\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{D}\mathbf{D}^{\mathrm{T}}$.

Finally, note that the fRLBF estimate (18) can be also expressed in a matrix-inversion-free form

$$\widehat{\boldsymbol{\alpha}}^{\mathrm{fRLBF}}(t) = \widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t) + [\widehat{\boldsymbol{\theta}}^{\mathrm{fRLBF}}(t) - \widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t)] \otimes \left[\frac{\mathbf{f}_0}{\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}\right] . \tag{20}$$

2.5. Optimization

Optimization of μ will be carried out using the type II maximum likelihood method (Akaike, 1980; Good, 1965), recently more frequently referred to as the empirical Bayes approach (Carlin & Louis, 1996; Chen, Ohlsson, & Ljung, 2012). Observe that minimization of the regularized measure of fit is equivalent to maximization of the quantity

$$\exp\left\{-\frac{1}{2\sigma_z^2}\sum_{i=-k}^{k}\|\boldsymbol{\theta}^*(t+i) - \mathbf{F}(i)\boldsymbol{\alpha}\|^2\right\} \times \exp\left\{-\frac{\mu}{2\sigma_z^2}\|\boldsymbol{\alpha}\|^2_{\mathbf{B}^{\mathrm{T}}\mathbf{B}}\right\} \tag{21}$$

where the first term can be attributed to the conditional (normal) data distribution $p(\Theta(t)|\boldsymbol{\alpha}, \sigma_z^2)$, $\Theta(t) = \{\boldsymbol{\theta}^*(t+i), i \in I_k\}$, and the second term – to the prior distribution of $\boldsymbol{\alpha}$: $\pi(\boldsymbol{\alpha}|\mu, \sigma_z^2)$. This leads to the following

Lemma 2. *Under Gaussian probability distributions of $\Theta(t)$ (conditional) and $\boldsymbol{\alpha}$ (prior) induced by (21), the best-local value of the regularization gain μ can be obtained from*

$$\widehat{\mu}(t) = \arg\min_{\mu}\big\{M\log\delta(t,\mu) - l\log\mu + \log|\mathbf{I}_{mn} + \mu\mathbf{B}^{\mathrm{T}}\mathbf{B}|\big\}. \tag{22}$$

where $|\cdot|$ denotes the determinant of a matrix, $M = kn - mn + l$, and

$$\delta(t,\mu) = \sum_{i=-k}^{k}\|\boldsymbol{\theta}^*(t+i)\|^2 - [\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t)]^{\mathrm{T}}[\mathbf{I}_{mn} + \mu\mathbf{B}^{\mathrm{T}}\mathbf{B}]^{-1}\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t). \tag{23}$$

Derivation

The conditional Gaussian data distribution associated with the first term in (21) has the form

$$p(\Theta(t)|\boldsymbol{\alpha}, \sigma_z^2) = \frac{1}{(2\pi\sigma_z^2)^{Kn/2}}\exp\left\{-\frac{1}{2\sigma_z^2}\|\boldsymbol{\alpha}\|^2 + \frac{1}{\sigma_z^2}\boldsymbol{\alpha}^{\mathrm{T}}\sum_{i=-k}^{k}\mathbf{F}^{\mathrm{T}}(i)\boldsymbol{\theta}^*(t+i) - \frac{1}{2\sigma_z^2}\sum_{i=-k}^{k}\|\boldsymbol{\theta}^*(t+i)\|^2\right\} \tag{24}$$

Such a form of the likelihood corresponds to the following assumptions specifying the *local* properties of the identified system

(A4) System parameters $\{\boldsymbol{\theta}(t+i), i \in I_k\}$ obey (3), i.e., they can be locally expressed as linear combinations of basis functions.

(A5) $\{\mathbf{z}(t+i), i \in I_k\}$ is a sequence of zero-mean uncorrelated and normally distributed random vectors with covariance matrix $\sigma_z^2\mathbf{I}_n$.

According to (A4), it holds that $\boldsymbol{\theta}^*(t+i) - \mathbf{F}(t)\boldsymbol{\alpha} = \mathbf{z}(t+i)$, $i \in I_k$. Assumption (A5) is fulfilled in typical communication systems, where both the input (transmitted) and noise sequences (both determining $\mathbf{z}(t)$ - see Niedźwiecki et al., 2020a) are white.

The prior distribution of $\boldsymbol{\alpha}$ associated with the second term in (21) is singular Gaussian (supported on a subspace of $\mathbb{R}^{mn}$) of the form Rao (2009)

$$f(\boldsymbol{\alpha}) = \frac{1}{\sqrt{|2\pi\boldsymbol{\Sigma}|_+}}e^{-\frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Sigma}^+\boldsymbol{\alpha}}$$

where $\boldsymbol{\Sigma}^+$ denotes the Moore–Penrose pseudoinverse of a positive semidefinite matrix $\boldsymbol{\Sigma}$, and $|\boldsymbol{\Sigma}|_+$ denotes its pseudodeterminant – the product of all nonzero eigenvalues of $\boldsymbol{\Sigma}$. In the case considered $\boldsymbol{\Sigma}^+ = \frac{\mu}{\sigma_z^2}\mathbf{B}^{\mathrm{T}}\mathbf{B}$. Note that $\mathbf{B}^{\mathrm{T}}\mathbf{B} = (\mathbf{D}^{\mathrm{T}}\otimes\mathbf{f}_0)(\mathbf{D}\otimes\mathbf{f}_0^{\mathrm{T}}) = (\mathbf{D}^{\mathrm{T}}\mathbf{D})\otimes(\mathbf{f}_0\mathbf{f}_0^{\mathrm{T}}) = \mathbf{R}\otimes(\mathbf{f}_0\mathbf{f}_0^{\mathrm{T}})$, and $|\mathbf{R}|_+ = |\mathbf{D}^{\mathrm{T}}\mathbf{D}|_+ = |\mathbf{D}\mathbf{D}^{\mathrm{T}}|$ where $|\cdot|$ denotes determinant of a square matrix. Using the identity $|\mathbf{A}\otimes\mathbf{B}|_+ = |\mathbf{A}|_+^{r(\mathbf{B})}|\mathbf{B}|_+^{r(\mathbf{A})}$, where $r(\cdot)$ denotes the rank of the corresponding matrix, the prior density associated with the second factor of (21) can be expressed in the form

$$\pi(\boldsymbol{\alpha}|\mu,\sigma_z^2) = \frac{(\mu\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0)^{l/2}|\mathbf{D}\mathbf{D}^{\mathrm{T}}|^{1/2}}{(2\pi\sigma_z^2)^{l/2}} \times \exp\left\{-\frac{1}{2\sigma_z^2}\boldsymbol{\alpha}^{\mathrm{T}}\left[\mathbf{R}\otimes\mu[\mathbf{f}_0\mathbf{f}_0^{\mathrm{T}}]\right]\boldsymbol{\alpha}\right\} . \tag{25}$$

Using the formula

$$\int_{\mathbb{R}^p}\exp\left\{-\frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} - \mathbf{x}^{\mathrm{T}}\mathbf{b} - c\right\}d\mathbf{x} = \frac{(2\pi)^{p/2}}{|\mathbf{A}|^{1/2}}\exp\left\{\frac{1}{2}\mathbf{b}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{b} - c\right\}$$

valid for a $p\times p$ positive definite matrix $\mathbf{A}$ (Gupta, 1963), the likelihood for the unknown parameters μ and σ_z^2 can be expressed in the form

$$L(\mu,\sigma_z^2) = \int_{\mathbb{R}^{mn}} p(\Theta(t)|\boldsymbol{\alpha},\sigma_z^2)\pi(\boldsymbol{\alpha}|\mu,\sigma_z^2)d\boldsymbol{\alpha} = \frac{(\mu\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0)^{l/2}|\mathbf{D}\mathbf{D}^{\mathrm{T}}|^{1/2}}{(2\pi\sigma_z^2)^{M/2}|\mathbf{I}_{mn}+\mu\mathbf{B}^{\mathrm{T}}\mathbf{B}|^{1/2}}\exp\left\{-\frac{\delta(t,\mu)}{2\sigma_z^2}\right\} \tag{26}$$

where $\delta(t,\mu)$ denotes the residual sum of squares given by (23).

It is convenient to work with -2 log likelihood. According to (26)

$$-2\log L(\mu,\sigma_z^2) = \mathrm{const} + M\log\sigma_z^2 - l\log\mu + \log|\mathbf{I}_{mn}+\mu\mathbf{B}^{\mathrm{T}}\mathbf{B}| + \frac{\delta(t,\mu)}{\sigma_z^2} . \tag{27}$$

Setting to zero the derivative of (27) with respect to σ_z^2, one arrives at the following maximum likelihood estimate of σ_z^2: $\widehat{\sigma}_z^2(t) = \delta(t,\mu)/M$. Finally, the optimal-local value of the regularization gain can be obtained by minimizing $-2\log L(\mu, \widehat{\sigma}_z^2(t))$, which leads to (22).

Minimization of the empirical Bayes measure of fit can be carried out numerically or, approximately, using a grid search over a set of predefined values, i.e., by restricting the search of μ in (22) to the set $\mathcal{M} = \{\mu_1, \ldots, \mu_L\}$.

Using (20) and the identity $\mathbf{F}_0^{\mathrm{T}}\boldsymbol{\theta} = (\mathbf{I}_n\otimes\mathbf{f}_0)(\boldsymbol{\theta}\otimes 1) = \boldsymbol{\theta}\otimes\mathbf{f}_0$, the residual sum of squares (23) can be rewritten in a computationally more convenient form

$$\delta(t,\mu) = \sum_{i=-k}^{k}\|\boldsymbol{\theta}^*(t+i)\|^2 - \|\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t)\|^2 + \frac{1}{\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}\|\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t)\|^2 - \frac{1}{\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}[\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t)]^{\mathrm{T}}[\mathbf{I}_n + \mu\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{R}]^{-1}\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t) \tag{28}$$

Note also that using the identity $|\mathbf{I}_i + \mathbf{P}\mathbf{Q}| = |\mathbf{I}_j + \mathbf{Q}\mathbf{P}|$ (Lütkepohl, 2005), where $\mathbf{P}$ and $\mathbf{Q}$ are matrices of dimensions $i\times j$ and $j\times i$, respectively, the last term of (22) can be written down in

the form

$$\begin{aligned}&\log|\mathbf{I}_{mn}+\mu\mathbf{B}^{\mathrm{T}}\mathbf{B}|=\log|\mathbf{I}_l+\mu\mathbf{B}\mathbf{B}^{\mathrm{T}}|\\&=\log|\mathbf{I}_l+\mu\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{D}\mathbf{D}^{\mathrm{T}}|=\log\prod_{i=1}^{l}(1+\mu\lambda_i\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0)\\&=\sum_{i=1}^{l}\log(1+\mu\lambda_i\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0)\end{aligned}\tag{29}$$

where $\lambda_i, i=1,\ldots,l$, denote eigenvalues of the matrix $\mathbf{D}\mathbf{D}^{\mathrm{T}}$, i.e., nonzero eigenvalues of the matrix $\mathbf{R}$.

3. Selection of regularization matrix

Depending on the available prior knowledge and/or desired properties of the estimated impulse response, one can shape the regularization matrix in different ways. So far it was assumed that the requlariztion matrix is fixed and hence is not subject to optimization. In practice, to gain greater flexibility, the matrix $\mathbf{R}$ is often parameterized in terms of some prior-related hyperparameter vector $\boldsymbol{\beta}$. In such a case the matrices $\mathbf{R}=\mathbf{D}^{\mathrm{T}}\mathbf{D}$ and $\mathbf{B}$ are functions of $\boldsymbol{\beta}$ and can be written down as $\mathbf{R}(\boldsymbol{\beta})=\mathbf{D}^{\mathrm{T}}(\boldsymbol{\beta})\mathbf{D}(\boldsymbol{\beta})$ and $\mathbf{B}(\boldsymbol{\beta})$, respectively. Consequently, the minimization carried out in (22) can be extended, at the cost of additional computational complexity, to μ *and* $\boldsymbol{\beta}$.

3.1. Time domain smoothness priors

Historically, this seems to be the oldest approach, originally proposed by Whittaker (Whittaker, 1923), (Whittaker & Robinson, 1924), and later exploited or reinvented by Akaike (1980), Congdon (2014), Gersch and Kitagawa (1989), Hunt (1973), Kitagawa and Gersch (1985, 1996), Phillips (1962), among many others.

Denote by g the one step lag advance: $g\theta_j(t)=\theta_{j+1}(t)$. As a local measure of smoothness of the impulse response (for a fixed value of t), one can adopt the pth order difference

$$\nabla^p\theta_j(t)=(1-g)^p\theta_j(t)=\sum_{i=0}^{p}c_i\theta_{j+i}(t)\tag{30}$$

where $c_i=(-1)^i\binom{p}{i}, i=0,\ldots,p$. Assuming that $\theta_j(t)=0$ for $j>n$, which is in the case considered reasonable, the $n\times n$ smoothness-enhancing regularization matrix can be adopted in the form $\mathbf{R}(p)=\mathbf{D}^{\mathrm{T}}(p)\mathbf{D}(p)$, where

$$\mathbf{D}(p)=\begin{bmatrix}c_0&c_1&\ldots&c_p&0&\ldots&0&0\\0&c_0&\ldots&c_{p-1}&c_p&\ldots&0&0\\\vdots&&\ddots&&&&&\\0&0&&c_0&c_1&\ldots&c_{p-1}&c_p\\0&0&&0&c_0&\ldots&c_{p-2}&c_{p-1}\\\vdots&\vdots&&&&\ddots&&\\0&0&&0&0&\ldots&c_0&c_1\\0&0&&0&0&\ldots&0&c_0\end{bmatrix}\tag{31}$$

Note that in such a case $\boldsymbol{\theta}^{\mathrm{T}}(t)\mathbf{R}(p)\boldsymbol{\theta}(t)=\|\mathbf{x}(t)\|^2$ where $\mathbf{x}(t)=\mathbf{D}(p)\boldsymbol{\theta}(t)=[\nabla^p\theta_1(t),\ldots,\nabla^p\theta_n(t)]^{\mathrm{T}}$, i.e., the regularization term quantifies the degree of smoothness of the sequence of impulse response coefficients $\theta_1(t),\ldots,\theta_n(t)$ under zero terminal conditions. The typical choices are $p=1$, which promotes local constancy of the impulse response, $p=2$, which promotes its local linearity, and $p=3$, which promotes its local convexity.

3.2. Frequency domain smoothness priors

This approach, originally proposed by Kitagawa and Gersch (1985), Gersch and Kitagawa (1989) for identification of time-invariant systems, can be easily adapted to the time-varying case. Denote by

$$H(\omega,t)=\sum_{i=1}^{n}\theta_i(t)e^{-j\omega i},\quad j=\sqrt{-1}\tag{32}$$

the frozen frequency response of the system (1). Note that

$$\int_{-\pi}^{\pi}\left|\frac{d^rH(\omega,t)}{d\omega^r}\right|^2d\omega=\sum_{i=1}^{n}i^{2r}\theta_i^2(t)\tag{33}$$

Adopting (33) as a measure of smoothness of the frequency response, one can select the $n\times n$ regularization matrix in the form $\mathbf{R}(r)=\mathbf{D}(r)\mathbf{D}(r)$ where

$$\mathbf{D}(r)=\mathrm{diag}\{1,2^r,3^r,\ldots,n^r\}.\tag{34}$$

3.3. Exponential stability priors

Assuming that the identified system is exponentially stable, one has the right to expect that the coefficients of its impulse response should smoothly decay to zero at an exponential rate. There may be also other physical reasons for such an expectation. For example, in the UWA channel case, the decaying power profile is due to the spreading and absorption loss (Stojanovic & Preisig, 2009). Exponential stability of the identified system has been the main intuition behind designing many regularization matrices, also referred to as kernels, proposed recently (Chen et al., 2012; Pillonetto et al., 2014). One of the examples is the first-order stable spline kernel, known also as tuned correlation kernel, of the form

$$\mathbf{R}^{-1}(\gamma)=\begin{bmatrix}1&\gamma&\ldots&\gamma^{n-1}\\\gamma&\gamma&\ldots&\gamma^{n-1}\\\vdots&&&\vdots\\\gamma^{n-1}&\gamma^{n-1}&\ldots&\gamma^{n-1}\end{bmatrix},\quad 0<\gamma<1\tag{35}$$

which, as shown in Marconato, Schoukens, and Schoukens (2016), combines the second order smoothness constraints with the exponential decay requirement – see the same reference for an interesting extension of this concept.

4. Computational aspects

Since evaluation of parameter estimates is repeated every time instant t, reduction of the computational load of the identification algorithms is of primary importance. From this point of view the proposed approach, summarized below, is very attractive. First, the cost of evaluation of the preestimates is low and equal to $O(n^2)$ flops per time update (or even to $O(n)$ flops if the so-called fast EWLS algorithms are used). Secondly, for selected functional bases, such as Legendre or Fourier, the fLBF estimates $\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t)$ and $\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t)$ can be updated recursively at the cost of $O(m^2n)$ flops (note that this cost does not depend on the width of the analysis window K) (Niedźwiecki & Ciołek, 2019a). When the fLBF algorithm is operated in the batch (off-line) mode and the convolutions (10) are computed using the FFT-based procedure, the total cost of evaluating the estimates $\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t)$ and $\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t)$ at N points $t=1,\ldots,N$, is equal to $O(mnN\log N)$ flops (for any K and any set of basis functions), i.e., $O(mn\log N)$ flops per time step. To obtain regularized fLBF estimates $\widehat{\boldsymbol{\theta}}^{\mathrm{fRLBF}}(t)$ and $\widehat{\boldsymbol{\alpha}}^{\mathrm{fRLBF}}(t)$ using (12) and (20), one needs additional $O(n^2)$ and $O(mn)$ flops per time update, respectively. Finally, if optimization is carried out via grid

search, the matrices $[\mathbf{I}_n + \mu \mathbf{f}_0^{\mathrm{T}} \mathbf{f}_0 \mathbf{R}(\boldsymbol{\beta})]^{-1}$ and $\mathbf{D}(\boldsymbol{\beta})$, which are data independent, can be precomputed for the considered values of μ and $\boldsymbol{\beta}$, and saved in the computer memory. In such a case the cost of optimization using (22) and exploiting (28) and (29), is of order $O(Ln^2)$ where L is the number of grid points.

Summary of the fRLBF algorithm (batch mode)

Input/output data: $y(t), \boldsymbol{\varphi}(t), t = 1, \ldots, N$

Settings:

- $n = \dim[\boldsymbol{\varphi}(\cdot)]$ – number of estimated parameters
- m – number of basis functions
- $K = 2k + 1$ – width of the sliding analysis window $T_k(t) = [t - k, t + k]$
- $\mathcal{F} = \{f_1(i), \ldots, f_m(i)\}$, $i \in I_k = [-k, k]$ – set of linearly independent orthonormal basis functions
- $\mathcal{M} = \{\mu_1, \ldots, \mu_L\}$ – set of grid points used for optimization of the regularization gain
- $\lambda_0 = \max\{0.9, 1 - 2/n\}$ – forgetting constant

Parameter estimation:

1. Compute EWLS estimates $\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t), t = 1, \ldots, N$, according to (6), using the recursive EWLS algorithm.
2. Compute preestimates $\boldsymbol{\theta}^*(t), t = k + 1, \ldots, N - k$ using (7).
3. Compute fLBF estimates $\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t), t = k+1, \ldots, N-k$, using (10).
4. Find the best fitting regularization gains $\widehat{\mu}(t) \in \mathcal{M}, t = k + 1, \ldots, N - k$, using (22).
5. Compute fRLBF estimates $\widehat{\boldsymbol{\theta}}^{\mathrm{fRLBF}}(t), t = k + 1, \ldots, N - k$, using (12) with $\mu = \widehat{\mu}(t)$.

Notice: The values of m and k can be selected adaptively using the parallel estimation technique.
When operated in a sliding window mode, the fRLBF algorithm is suitable for almost real-time applications, i.e., applications that tolerate a constant decision (estimation) delay of k sampling intervals.

5. Selection of the number of basis functions and the analysis window size

It is known that m (the number of basis functions) and k (half-width of the analysis interval) are important design parameters which should be chosen so as to trade-off the bias and variance components of the mean squared parameter estimation error. Increasing m and/or decreasing k results in decreasing the estimation bias at the cost of increasing its variance. Decreasing m and/or increasing k has the opposite effect (Niedźwiecki et al., 2020a). For this reason, to guarantee satisfactory estimation results, the values of m and k should be locally adjusted to the rate and mode of parameter variation. The problem can be solved using the parallel estimation technique. In this framework, not one but several identification algorithms, equipped with different settings, are run simultaneously yielding the estimates $\widehat{\boldsymbol{\alpha}}_{m|k}^{\mathrm{fLBF}}(t)$, $\widehat{\boldsymbol{\theta}}_{m|k}^{\mathrm{fLBF}}(t)$, $m \in \mathcal{M}$, $k \in \mathcal{K}$. At each time instant only one of the competing algorithms is chosen, i.e., the parameter estimates are obtained in the form

$$\widehat{\boldsymbol{\alpha}}_{\widehat{m}(t)|\widehat{k}(t)}^{\mathrm{fLBF}}(t), \quad \widehat{\boldsymbol{\theta}}_{\widehat{m}(t)|\widehat{k}(t)}^{\mathrm{fLBF}}(t) \tag{36}$$

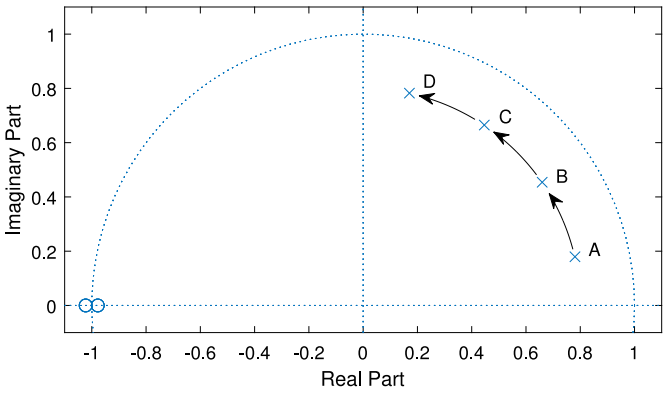

Fig. 1. Location of "zeros" (○) and "poles" (×) of the identified time-varying IIR system.

where

$$\{\widehat{m}(t), \widehat{k}(t)\} = \arg\min_{\substack{m \in \mathcal{M} \\ k \in \mathcal{K}}} J_{m|k}(t) \tag{37}$$

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

The cross-validation selection rule proposed in Niedźwiecki et al. (2020a) is based on minimization of the localized sum of squared leave-one-out output interpolation errors. Alternatively, the best-local choice of m and k can be made using the suitably modified Akaike's final prediction error (FPE) criterion (Niedźwiecki & Ciołek, 2019a). Once the best-fitting fLBF estimates are chosen, one can apply to (36) the adaptive regularization procedure described above.

6. Simulation results

To make simulations more realistic, the second-order IIR (infinite impulse response) system governed by

$$y(t) = y_0(t) + e(t), \quad y_0(t) = G_0(t, q^{-1})u(t-1)$$
$$G_0(t, q^{-1}) = \frac{b_0 + b_1 q^{-1} + b_2 q^{-2}}{1 + a_1(t) q^{-1} + a_2(t) q^{-2}} \tag{38}$$

where q^{-1} denotes the one-step time delay, was used to generate data which were next approximated by the 50-tap FIR filter. The transfer function of the identified system had two time-invariant real "zeros" located at -1.02 and 0.98 ($b_0 = b_2 = 0.02008$, $b_1 = 0.04017$) and two time-varying complex-conjugate "poles" slowly moving, with a constant speed, along the trajectory parallel to the unit circle, from their initial position A ($0.78 \pm j0.18$) to the terminal position D ($0.18 \pm j0.78$) - see Fig. 1. The distance from A to D was covered in $T = 3000$ time steps. At the point A the "frozen" transfer function $G_0(1000, q^{-1})$ describes the second-order Butterworth filter analyzed in Ljung et al. (2020). The system was excited by the first order autoregressive signal $u(t) = 0.8u(t-1)+v(t)$, $\mathrm{var}[v(t)] = 1$, where $\{v(t)\}$ denotes white noise independent of $\{e(t)\}$. The variance of the measurement noise was set to $\sigma_e^2 = 0.0025$. At the checkpoints A, B, C and D the average signal-to-noise ratio (SNR) was equal to: 28 dB, 20 dB, 12 dB and 9 dB, respectively.

To provide a clear picture of the improvement yielded by regularization, simulations were carried out for fixed values of $k = 100$ and $m = 3$. The basis set, prior to orthonormalization, was made up of powers of time: $g_1(i) = 1, g_2(i) = i, g_3(i) = i^2, i \in I_k$. It is worth noticing that parameter trajectories of the simulated IIR system *cannot* be expressed as linear combinations of the adopted basis functions. To avoid initialization (transient) effects, until $t = 1000$ simulation was carried out for a system with

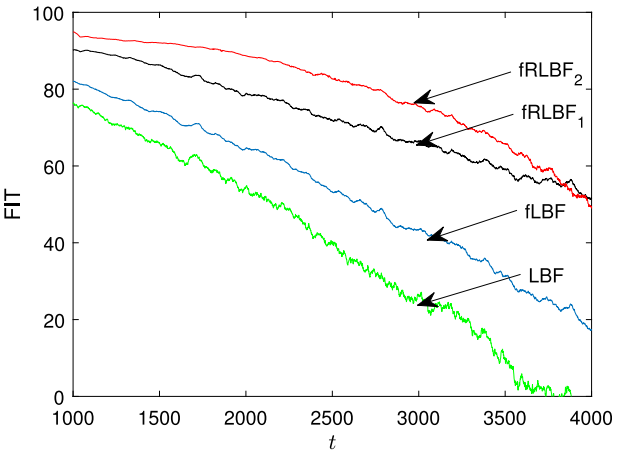

Fig. 2. Average FIT scores for the LBF algorithm, fLBF algorithm, and for two regularized versions of the fLBF algorithm: fRLBF_1 and fRLBF_2.

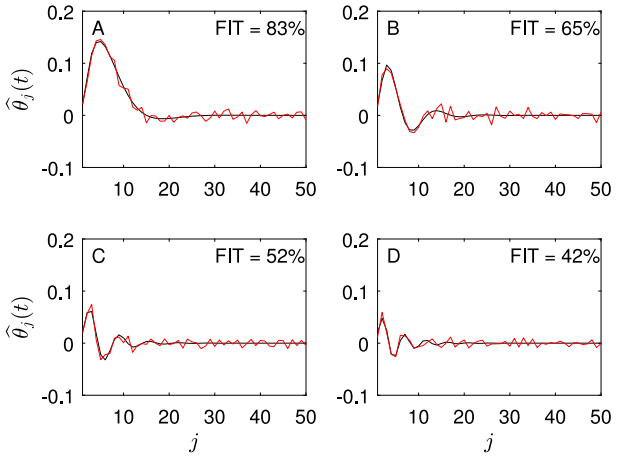

Fig. 3. Four snapshots of the true (black line) and estimated (red line) impulse response of the identified system – the fLBF approach. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

constant a-parameters set to $a_1(1) = \cdots = a_1(1000) = -1.561$ and $a_2(1) = \cdots = a_2(1000) = 0.6414$. After $t = 4000$ system parameters were again set to constant (terminal) values $a_1(4000)$ and $a_2(4000)$. Evaluation of identification results was started at the instant $t = 1001$, once system parameters begun to vary with time, and stopped at the instant $t = 4000$. At each time instant the first 50 coefficients of the true time-varying impulse response were calculated and checked against their estimates obtained for the FIR model (note: the true impulse response of a time-varying IIR system differs from its "frozen" impulse response, i.e., impulse response of a time-invariant system with parameters fixed at $\boldsymbol{\theta}(t)$). The forgetting constant used to generate preestimates was set to the recommended value $\lambda_0 = 1 - \frac{2}{50} = 0.96$.

Estimation accuracy was evaluated using the following normalized root mean squared error measure of fit (Ljung & Chen, 2013)

$$w(t) = 100\left(1 - \left[\frac{\sum_{j=1}^{50} |\theta_j(t) - \widehat{\theta}_j(t)|^2}{\sum_{j=1}^{50} |\theta_j(t) - \bar{\theta}(t)|^2}\right]^{1/2}\right) \tag{39}$$

where $\bar{\theta}(t) = \frac{1}{50}\sum_{j=1}^{50} \theta_j(t)$. The maximum value of $w(t)$, equal to 100, corresponds to the perfect match between the true and estimated impulse response. The final scores, further referred to as FIT (%) were obtained by ensemble averaging (100 independent realizations of $\{e(t)\}$ and $\{v(t)\}$) of the instantaneous measures.

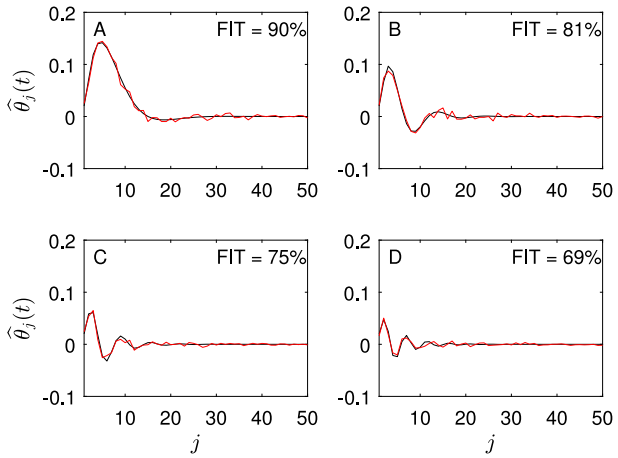

Fig. 4. Four snapshots of the true (black line) and estimated (red line) impulse response of the identified system – the fRLBF_1 approach. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Two variants of regularization were checked: (a) with an exponential stability prior matrix (35) and $\gamma \in \{0.9, 0.92, 0.94, 0.96, 0.98\}$ – referred to as fRLBF_1; (b) with a smoothness priors regularization matrix (31) and $p = 3$ – referred to as fRLBF_2. In both cases optimization of μ was carried out numerically using (22) in the interval [0,100] with a step 0.1. Fig. 2 shows the average value of FIT in the evaluation interval [1001,4000]. All FIT scores decrease with time because of declining SNR. Note that, compared to the standard (not regularized) fLBF algorithm, the fRLBF_1 algorithm yields improvement of 8%–33%. When the fRLBF_2 algorithm is used, the improvement is even more significant and ranges between 12% and 33%. Finally, note poor performance of the LBF algorithm. This is not a surprise since, in the case considered, the LBF algorithm estimates at each time step $mn = 150$ hyperparameters from $K = 2k+1 = 201$ input/output measurements, which is against the good practice in system identification called the principle of parsimony (Niedźwiecki, 2000) (the asymptotic equivalence of LBF and fLBF estimators, mentioned in Section 3.3, holds true provided that $mn \ll K$). Finally, we note that the situation does not change if k and m are chosen in an adaptive manner set out in Section 5 - the regularized algorithms continue to perform better than the not regularized ones.

To give additional insight into the identification process, four snapshots of the true and estimated impulse responses, evaluated at the points A, B, C and D, respectively, were depicted in Fig. 3 (fLBF estimates), Fig. 4 (fRLBF_1 estimates) and Fig. 5 (fRLBF_2 estimates). Note that in each case regularization yields both quantitative and qualitative improvement over the standard fLBF solution.

The times needed to execute a single identification step (single time update) using computer equipped with the Intel Core i7 1.8 GHz processor (4 cores) were equal to 0.3 ms, 20 ms and 4 ms for the algorithms fLBF, fRLBF_1 and fRLBF_2, respectively. For the LBF algorithm the analogous execution time was equal to 4 ms. MATLAB codes used to generate simulation results are available under the link: https://eti.pg.edu.pl/katedra-systemow-automatyki/fRLBF.

Remark. It was noticed that the optimization rule (22) may occasionally select very large values of μ, which results in temporary degradation of the model quality. This negative effect, which

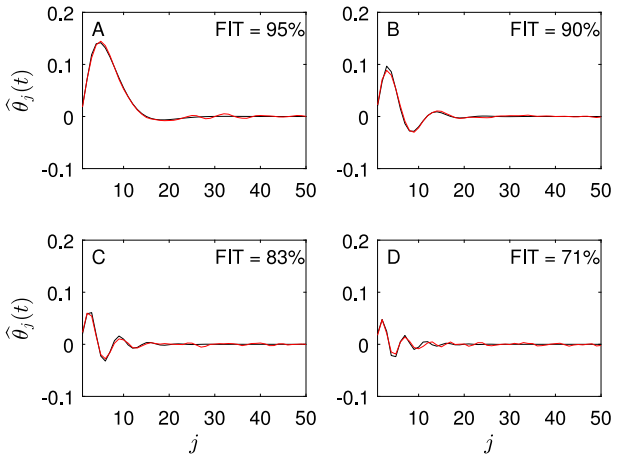

Fig. 5. Four snapshots of the true (black line) and estimated (red line) impulse response of the identified system — the fRLBF_2 approach. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

occurs mostly for small signal-to-noise ratios (SNR$\leq$ 10 dB) can be eliminated, or at least significantly reduced, by imposing an upper bound on μ, like in our simulation examples above. It should be stressed that the effect described above is *not* caused by the fact that the preestimation model (9) is only an approximation — the same behavior can be observed in the "ideal" case, where the signal $\boldsymbol{\theta}^*(t)$ is artificially generated by adding spatially uncorrelated white Gaussian noise to $\boldsymbol{\theta}(t)$. For this reason it seems to be worth further investigation.

7. Conclusion

It was shown that identification of time-varying FIR systems can be effectively carried out by means of smoothing the appropriately generated sequence of parameter preestimates. As a smoothing filter one can use the regularized Savitzky–Golay algorithm, locally optimized using the empirical Bayes approach. The resulting identification algorithm outperforms the currently available solutions in terms of estimation (tracking) accuracy, computational complexity and numerical robustness.

References

Acharya, D., Rani, A., Agarwal, S., & Singh, V. (2016). Application of adaptive Savitzky-Golay filter for EEG signal processing. *Perspectives in Science*, (8), 677–679.

Akaike, H. (1980). Likelihood and the Bayes procedure. In J. M. Bernardo, M. H. De Groot, D. V Lindley, & A. F. M. Smith (Eds.), *Bayesian statistics* (pp. 143–166). Valencia, Spain: University Press.

Berger, J. (1985). *Statistical decision theory and bayesian analysis*. New York: Springer.

Carlin, B. P., & Louis, T. A. (1996). *Bayes and empirical bayes methods for data analysis*. London: Chapman & Hall.

Chen, T., Ohlsson, H., & Ljung, L. (2012). On the estimation of transfer functions, regularizations and Gaussian processes - Revisited. *Automatica*, (48), 1525–1535.

Congdon, P. (2014). *Applied bayesian modelling*. Wiley.

Gersch, W., & Kitagawa, G. (1989). Smoothness priors transfer function estimation. *Automatica*, (25), 603–608.

Goel, S., Kaur, G., & Tomar, P. (2016). ECG signal processing by Savitzky Golay least square polynomial filter. *International Journal of Advances in Electronics and Computer Science*, (3), 20–22.

Good, I. J. (1965). *The estimation of probabilities*. Cambridge, Mass: MIT Press.

Gupta, S. S. (1963). Probability integrals of multivariate normal and multivariate t^1. *The Annals of Mathematical Statistics*, (34), 792–828.

Hargittai, S. (2005). Savitzky-Golay least-squares polynomial filters in ECG signal processing. *Proceedings of Computers in Cardiology*, (512), 763–766.

Hunt, B. R. (1973). The application of constrained least-squares estimation to image restoration by digital computer. *IEEE Transactions on Computers*, (22), 805–812.

Jakubowska, M., & Kubiak, W. W. (2004). Adaptive-degree polynomial filter for voltammetric signals. *Analytica Chimica*, (512), 241–250.

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

Kitagawa, G., & Gersch, W. (1996). Smoothness priors analysis of time series. In *Lecture notes in statistisc*, (No. 116), Springer.

Kochańska, I. (2015). Adaptive identification of time-varying impulse response of underwater acoustic communication channel. *Hydroacoustics*, (18), 87–94.

Li, J., Deng, H., Li, P., & Yu, B. (2015). Real-time infrared detection based on adaptive Savitzky-Golay algorithm. *Applied Physics B*, (120), 207–216.

Li, W., & Preisig, J. C. (2007). Estimation of rapidly time-varying sparse channels. *IEEE J. Oceanic Eng.*, (32), 927–939.

Ljung, L., & Chen, T. (2013). What can regularization offer for estimation of dynamical systems? In *Proc. of the 11th IFAC workshop on adaptation and learning in control and signal processing* (pp. 1–8). Caen, France.

Ljung, L., Chen, T., & Mu, B. (2020). A shift in paradigm for system identification. *International Journal of Control*, (93), 173–180.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Berlin: Springer-Verlag.

Marconato, A., Schoukens, M., & Schoukens, J. (2016). Filter-based regularisation for impulse response modelling. *IET Control Theory & Applications*, (112), 194–204, see also arXiv:1610.07353.

Niedźwiecki, M. (2000). *Identification of time-varying processes*. New York: Wiley.

Niedźwiecki, M., & Ciołek, M. (2019a). Fully adaptive Savitzky-Golay type smoothers. In *Proc. 27th European signal processing conference* (p. 5). A Coruna, Spain.

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

Niedźwiecki, M., Ciołek, M., & Gańcza, A. (2020a). A new look at the statistical identification of nonstationary systems. *Automatica*, (118), Article 109037, 1–9.

Niedźwiecki, M., Gańcza, A., & Ciołek, M. (2020b). On the preestimation technique and its application to identification of nonstationary systems. In *Proc. of the 59th IEEE conference on decision and control, Jeju Island, Republic of Korea* (pp. 282–293).

Phillips, D. L. (1962). A technique for the numerical solution of certain integral equations of the first kind. *Journal of the ACM*, (9), 84–97.

Pillonetto, G., Dinuzzo, F., Chen, T., De Nicolao, G., & Ljung, L. (2014). Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica*, (50), 657–682.

Qiao, G., Gan, S., Liu, S., Ma, L., & Sun, Z. (2018). Digital self-interference cancellation for asynchronous in-band full-duplex underwater acoustic communication. *Sensors*, (18), 1700–1716.

Rao, C. R. (2009). *Linear statistical inference and its applications*. Wiley.

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data with simplified least squares procedures. *Annals of Chemistry*, (36), 1627–1639.

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

Shen, L., Henson, B., Zakharov, Y., & Mitchell, P. (2020). Digital self-interference cancellation for underwater acoustic systems. *IEEE Access*, (8), Article 187590-187604.

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

Stojanovic, M., & Preisig, J. (2009). Underwater acoustic communication channels: Propagation models and statistical characterization. *IEEE Communications Magazine*, (47), 84–89.

Tikhonov, A., & Arsenin, V. (1977). *Solutions of ill-posed problems*. Winston/Wiley.

Turton, B. C. H. (1992). A novel variant of the Savitzky-Golay filter for spectroscopic applications. *Measurement Science Technology*, (3), 858–863.

Wei, H. L., Liu, J. J., & Billings, S. A. (2002). Identification of time-varying systems using multi-resolution wavelet models. *International Journal of Systems Science*, (33), 1217–1228.

Whittaker, E. T. (1923). On a new method of graduation. *Proceedings of Edinborough Mathematical Association*, (78), 81–89.

Whittaker, E. T., & Robinson, G. (1924). *Calculus of observations: A treatise on numerical mathematics*. Blackie and Son.

Maciej Jan Niedźwiecki received the M.Sc. and Ph.D. degrees from the Technical University of Gdańsk, Gdańsk, Poland and the Dr.Hab. (D.Sc.) degree from the Technical University of Warsaw, Warsaw, Poland, in 1977, 1981 and 1991, respectively. He spent three years as a Research Fellow with the Department of Systems Engineering, Australian National University, 1986–1989. In 1990–1993 he served as a Vice Chairman of Technical Committee on Theory of the International Federation of Automatic Control (IFAC). He is the author of the book *Identification of Time-varying Processes* (Wiley, 2000). His main areas of research interests include system identification, statistical signal processing and adaptive systems.

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

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

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

Piotr Kaczmarek received the M.Sc. and Ph.D. degrees in automatic control from the Gdańsk University of Technology, Gdańsk, Poland, in 2000 and 2008, respectively. He has also graduated from the European Master Degree Course in Control and Management of Lean Manufacturing in Network Systems conducted in co-operation between Gdańsk University of Technology (Poland), Catholic University of Louvain (Belgium), and University of Karlsruhe (Germany). Since 2008, he has been working as an Adjunct Professor in the Department of Automatic Control, Faculty of Electronics, Telecommunications and Informatics, GUT. His interests include system identification and adaptive filtering as well as optimization of production techniques.