Postprint of: Gańcza A., Niedźwiecki M., Regularized identification of fast time-varying systems - comparison of two regularization strategies, 60th IEEE Conference on Decision and Control (CDC), 2021, pp. 864-871, DOI: 10.1109/CDC45484.2021.9683773

Regularized identification of fast time-varying systems - comparison of two regularization strategies

Artur Gańcza and Maciej Niedźwiecki

***Abstract*— The problem of identification of a time-varying FIR system is considered and solved using the local basis function approach. It is shown that the estimation (tracking) results can be improved by means of regularization. Two variants of regularization are proposed and compared: the classical L^2 (ridge) regularization and a new, reweighted L^2 one. It is shown that the new approach can outperform the classical one and is computationally attractive.**

I. Introduction

When parameters of the identified dynamic system vary slowly with time, their estimation can be carried out successfully using the time-localized versions of the classical estimation methods, such as least squares or maximum likelihood [1] - [3]. Such an approach is based on the (implicit) assumption that the analyzed system is "locally stationary", i.e., that its parameters can be regarded as constant in sufficiently long time intervals [4], [5].

When system parameters vary at a fast rate, like in the case of some rapidly fading telecommunication channels, the simple solution described above may fail to provide estimates of sufficient accuracy [6], [7]. The well-known way out of difficulty is via incorporation in the system description an explicit model of parameter time-variation. Such a model, often referred to as a hypermodel, can be deterministic or stochastic. In the first case, parameter trajectories are approximated by linear combinations of known functions of time, called basis functions (BF). This allows one to convert the problem of estimation of time-varying parameters into a simpler problem of estimation of constant hyperparameters – coefficients characterizing the adopted hypermodel. Such a problem can be easily solved using the classical estimation methods [8] - [15].

In the case of stochastic hypermodels, the problem of identification of a time-varying system can be regarded as a problem of filtration/smoothing in the state space. It can be solved using Kalman filtering techniques [16] - [19].

In the majority of studies devoted to the BF approach, which is also a theme of the current work, basis functions are used to generate interval estimates of parameter trajectories. Recently a new class of identification algorithms was described, which combines the BF approach with the local estimation technique [20], [21]. The proposed local basis function (LBF) and fast local basis function (fLBF) estimators provide a sequence of point estimates of system parameters corresponding to different locations of a sliding analysis window of a fixed width. As shown in [20], such a point approach yields more accurate estimates than the interval one, favorably comparing with the state-of-the-art multi-wavelet estimation scheme proposed in [14] - see [23].

The current contribution aims to show that accuracy of LBF/fLBF estimators can be further increased by means of regularization. Regularization is a well established technique in estimation and machine learning. In system identification, regularization, achieved by adding to the minimized cost function a term penalizing the norm of the solution, allows one to improve the estimation bias-variance trade-off which decides upon accuracy of the identified model [24], [25]. So far most of the work performed in this area was restricted to identification of time-invariant systems. We will show that similar advantages can be reached when regularization is incorporated in LBF-based identification of time-varying systems. Two variants of regularization will be proposed and compared: the classical L^2 regularization, often referred to as ridge regression [26], and the reweighted L^2 regularization, which can be regarded as the "first order approximation" of the L^1 regularization, known also as LASSO (least absolute shrinkage and selection operator) [27]. It will be shown that in both cases optimization of regularization gains can be carried out in a computationally efficient way using the leave-one-out cross-validation approach.

All algorithms presented in this paper are noncausal, i.e., they yield parameter estimates that depend on both "past" and "future" input/output data (relative to the time instant of interest) so they cannot be used in real time applications such as prediction or control. However, they can be applied whenever the model-based decisions can be delayed by a certain number of sampling intervals. Channel equalization and self-interference mitigation in full-duplex communication systems are examples of such feasible almost real time applications [28].

II. Local basis function and fast local basis function estimators

Consider the problem of identification of a time-varying complex-valued finite impulse response (FIR) system governed by

$$\begin{aligned} y(t) &= \sum_{j=1}^{n} \theta_j^*(t)u(t-j+1) + e(t) \\ &= \boldsymbol{\theta}^{\mathbf{H}}(t)\boldsymbol{\varphi}(t) + e(t) \end{aligned} \tag{1}$$

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

where $t = \ldots, -1, 0, 1, \ldots$ denotes discrete (normalized) time, $y(t)$ denotes the complex-valued output signal, $\boldsymbol{\varphi}(t) = [u(t), \ldots, u(t-n+1)]^{\mathrm{T}}$ denotes regression vector made up of past samples of the complex-valued input signal $u(t)$, $\boldsymbol{\theta}(t) = [\theta_1(t), \ldots, \theta_n(t)]^{\mathrm{T}}$ is the vector of time-varying system coefficients, and $\{e(t)\}$ denotes measurement noise. The symbol $*$ stands for complex conjugate and H – complex conjugate transpose (Hermitian transpose). We will assume that

(A1) $\{e(t)\}$, independent of $\{u(t)\}$, is a sequence of zero-mean independent and identically distributed random variables.

(A2) $\{\boldsymbol{\theta}(t)\}$ is a uniformly bounded sequence, independent of $\{u(t)\}$ and $\{e(t)\}$.

The LBF/fLBF identification technique is based on the assumption that in the local analysis interval $T(t) = [t-k, t+k]$ of length $K = 2k+1$, centered at t, system parameters can be expressed as linear combinations of a certain number of linearly independent real-valued functions of time $f_1(i), \ldots, f_m(i), i \in I_k = [-k, k]$, further referred to as basis functions, namely

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

where $\mathbf{f}(i) = [f_1(i), \ldots, f_m(i)]^{\mathrm{T}}$.

In agreement with the local estimation paradigm, estimation of parameter trajectories, based on the hypermodel (2), is carried out independently for each localization of the analysis interval $T(t)$, i.e., it is performed in the sliding window mode. Note that even though system hyperparameters a_{jl} are assumed to be constant in the interval $[t-k, t+k]$, their values are allowed to change along with the position of the analysis window. For this reason they are written down as functions of t.

Denote by $w(i), i \in I_k, w(0) = 1$, a symmetric, nonnegative, bell-shaped window which will be used to put more emphasis on data gathered at instants close to t than on instants far from t. For convenience, but without any loss of generality, we will assume that the adopted basis functions are w-orthonormal, namely

$$\sum_{i=-k}^{k} w(i)\mathbf{f}(i)\mathbf{f}^{\mathrm{T}}(i) = \mathbf{I}_m \,. \tag{3}$$

where $\mathbf{I}_m$ denotes the $m \times m$ identity matrix. The most common choices of basis functions prior to normalization are powers of time (Taylor series approximation) and cosine functions (Fourier series approximation).

A. LBF approach

Note that the hypermodel (2) can be expressed in a more compact form

$$\begin{aligned}\boldsymbol{\theta}(t+i) &= \mathbf{F}(i)\boldsymbol{\alpha}(t), \quad i \in I_k \\ \boldsymbol{\alpha}(t) &= [\boldsymbol{\alpha}_1^{\mathrm{T}}(t), \ldots, \boldsymbol{\alpha}_n^{\mathrm{T}}(t)]^{\mathrm{T}}\end{aligned} \tag{4}$$

where $\mathbf{F}(i) = \mathbf{I}_n \otimes \mathbf{f}^{\mathrm{T}}(i)$ and $\otimes$ denotes the Kronecker product of the corresponding vectors/matrices. Using (4), the system equation (1) can be written down in the form

$$y(t+i) = \boldsymbol{\alpha}^{\mathbf{H}}(t)\boldsymbol{\psi}(t,i) + e(t+i), \quad i \in I_k \tag{5}$$

where $\boldsymbol{\psi}(t,i) = \boldsymbol{\varphi}(t+i) \otimes \mathbf{f}(i)$ denotes the generalized regression vector.

The LBF estimator has the form [20]

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}^{\mathrm{LBF}}(t) &= \arg\min_{\boldsymbol{\alpha}} \sum_{i=-k}^{k} w(i)|y(t+i) - \boldsymbol{\alpha}^{\mathbf{H}}\boldsymbol{\psi}(t,i)|^2 \\ &= \mathbf{P}^{-1}(t)\mathbf{p}(t) \\ \widehat{\boldsymbol{\theta}}^{\mathrm{LBF}}(t) &= \mathbf{F}_0\widehat{\boldsymbol{\alpha}}^{\mathrm{LBF}}(t)\end{aligned} \tag{6}$$

where

$$\begin{aligned}\mathbf{P}(t) &= \sum_{i=-k}^{k} w(i)\boldsymbol{\psi}(t,i)\boldsymbol{\psi}^{\mathbf{H}}(t,i) \\ \mathbf{p}(t) &= \sum_{i=-k}^{k} w(i)y^*(t+i)\boldsymbol{\psi}(t,i)\end{aligned} \tag{7}$$

and $\mathbf{F}_0 = \mathbf{F}(0) = \mathbf{I}_n \otimes \mathbf{f}_0^{\mathrm{T}},\ \mathbf{f}_0 = \mathbf{f}(0)$.

For recursively computable basis and window functions the $mn \times mn$ generalized regression matrix $\mathbf{P}(t)$ and the $mn \times 1$ vector $\mathbf{p}(t)$ can be evaluated in a recursive way [20]. However, since the matrix $\mathbf{P}(t)$ must be inverted every time instant t, the computational burden associated with LBF estimators may be substantial.

B. fLBF approach

As shown in [21], [23], under assumptions (A1), (A2) and

(A3) $\{u(t)\}$ is a zero-mean wide sense stationary Gaussian sequence with an exponentially decaying autocorrelation function

the LBF estimates $\widehat{\boldsymbol{\alpha}}^{\mathrm{LBF}}(t)$ and $\widehat{\boldsymbol{\theta}}^{\mathrm{LBF}}(t)$ can be approximated by the following computationally fast formulas

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t) &= \left[\left[\widehat{\boldsymbol{\alpha}}_1^{\mathrm{fLBF}}(t)\right]^{\mathrm{T}}, \ldots, \left[\widehat{\boldsymbol{\alpha}}_n^{\mathrm{fLBF}}(t)\right]^{\mathrm{T}}\right]^{\mathrm{T}} \\ \widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t) &= \left[\widehat{\theta}_1^{\mathrm{fLBF}}(t), \ldots, \widehat{\theta}_n^{\mathrm{fLBF}}(t)\right]^{\mathrm{T}} \\ \widehat{\boldsymbol{\alpha}}_j^{\mathrm{fLBF}}(t) &= \arg\min_{\boldsymbol{\alpha}_j} \sum_{i=-k}^{k} w(i)|\widetilde{\theta}_j(t) - \mathbf{f}^{\mathrm{T}}(i)\boldsymbol{\alpha}_j|^2 \\ &= \sum_{i=-k}^{k} w(i)\widetilde{\theta}_j(t)\mathbf{f}(i) \\ \widehat{\theta}_j^{\mathrm{fLBF}}(t) &= \mathbf{f}_0^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}_j^{\mathrm{fLBF}}(t) = \sum_{i=-k}^{k} h(i)\widetilde{\theta}_j(t+i) \\ j &= 1, \ldots, n\end{aligned} \tag{8}$$

where

$$h(i) = w(i)\mathbf{f}_0^{\mathrm{T}}\mathbf{f}(i), \quad i \in I_k \tag{9}$$

denotes the impulse response of the FIR filter associated with the LBF estimator and $\{\widetilde{\theta}_j(t)\}$ denotes the preestimated trajectory of the j-th system parameter, obtained by means of "inverse filtering" of the estimates yielded by the short-memory exponentially weighted least squares (EWLS) algorithm. The EWLS estimates $\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t) = \left[\widehat{\theta}_1^{\mathrm{EWLS}}(t), \ldots, \widehat{\theta}_n^{\mathrm{EWLS}}(t)\right]^{\mathrm{T}}$, defined as

$$\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t) = \arg\min_{\boldsymbol{\theta}} \sum_{i=0}^{t-1} \lambda_0^i |y(t-i) - \boldsymbol{\theta}^{\mathbf{H}} \boldsymbol{\varphi}(t-i)|^2 \quad (10)$$

where λ_0, $0 < \lambda_0 < 1$, denotes the so-called forgetting constant, can be computed using the well-known recursive algorithm [1]

$$\begin{aligned} \varepsilon(t) &= y(t) - \left[\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1)\right]^{\mathbf{H}} \boldsymbol{\varphi}(t) \\ \mathbf{k}(t) &= \frac{\mathbf{R}(t-1)\boldsymbol{\varphi}(t)}{\lambda_0 + \boldsymbol{\varphi}^{\mathbf{H}}(t)\mathbf{R}(t-1)\boldsymbol{\varphi}(t)} \\ \widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t) &= \widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(t-1) + \mathbf{k}(t)\varepsilon^*(t) \\ \mathbf{R}(t) &= \frac{1}{\lambda_0}\left[\mathbf{I}_n - \mathbf{k}(t)\boldsymbol{\varphi}^{\mathbf{H}}(t)\right]\mathbf{R}(t-1) \end{aligned} \quad (11)$$

with initial conditions $\widehat{\boldsymbol{\theta}}^{\mathrm{EWLS}}(0) = 0$ and $\mathbf{R}(0) = c\mathbf{I}_n$, where c denotes a large positive constant. The inverse filtering formula has the form

$$\widetilde{\theta}_j(t) = L_t \widehat{\theta}_j^{\mathrm{EWLS}}(t) - \lambda_0 L_{t-1} \widehat{\theta}_j^{\mathrm{EWLS}}(t-1) \quad (12)$$

where $L_t = \sum_{i=0}^{t-1} \lambda_0^i = \lambda_0 L_{t-1} + 1$, $L_0 = 1$, denotes the effective width of the exponential window. For large values of t, when the effective window width reaches its steady state value $L_\infty = 1/(1-\lambda_0)$, the formula (12) can be replaced with

$$\widetilde{\theta}_j(t) = \frac{1}{1-\lambda_0}\left[\widehat{\theta}_j^{\mathrm{EWLS}}(t) - \lambda_0 \widehat{\theta}_j^{\mathrm{EWLS}}(t-1)\right] \quad (13)$$

Since under assumptions (A1) - (A3) each preestimate $\widetilde{\theta}_j(t)$ can be written down in the form

$$\widetilde{\theta}_j(t) \cong \theta_j(t) + z_j(t) \quad (14)$$

where $z_j(t)$ denotes zero-mean white noise with large variance, the fLBF estimate $\widehat{\theta}_j^{\mathrm{fLBF}}(t)$ can be regarded as a result of "denoising" $\widetilde{\theta}_j(t)$ using the basis function approach [21].

The recommended choice of the forgetting factor is $\lambda_0 = \max\{0.9, 1-2/n\}$ so that the number of estimated parameters is not larger than the so-called equivalent width of the applied exponential window equal to $(1+\lambda)/(1-\lambda) \cong 2/(1-\lambda)$ [3]. Since the resulting effective memory L_∞ of the EWLS algorithm is usually short, or very short, the assumption about stationarity of the input signal postulated in (A3) is by no means critical - local ("L_∞-wide") stationarity is sufficient to guarantee good preestimation results. The same applies to the assumption about Gaussianity of $u(t)$.

In spite of its computational simplicity, the fLBF scheme has similar parameter tracking capabilities as the LBF scheme. As a matter of fact in many cases the fLBF estimates are almost indistinguishable from the LBF ones [21].

III. Regularized LBF and fLBF estimators

The main goal of identification is to minimize the mean squared parameter estimation (tracking) error (MSE). It is well-known that MSE can be written down as a sum of its bias and variance components. Hence, minimization of the MSE requires finding a good trade-off between estimation bias and estimation variance. Very often, an improvement of this trade-off can be achieved by means of regularization. Probably the simplest approach is to add to the minimized cost function, such as (6) or (8), an additional term that penalizes large values of the estimated parameters.

A. Ridge LBF and fLBF estimators

Denote by $||\mathbf{x}||_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$ the L^2 norm of a complex-valued vector $\mathbf{x} = [x_1, \ldots, x_n]^{\mathrm{T}}$. When the penalty mentioned above is proportional to $||\boldsymbol{\alpha}||_2^2$, the technique is called ridge regression. The resulting ridge LBF (RLBF) estimates can be obtained from

$$\begin{aligned} \widehat{\boldsymbol{\alpha}}^{\mathrm{RLBF}}(t) &= \arg\min_{\boldsymbol{\alpha}} \left\{ \sum_{i=-k}^{k} w(i) |y(t+i) - \boldsymbol{\alpha}^{\mathbf{H}} \boldsymbol{\psi}(t,i)|_2^2 \right. \\ &\qquad \left. + \mu ||\boldsymbol{\alpha}||_2^2 \right\} = \mathbf{S}^{-1}(t)\mathbf{p}(t), \\ \widehat{\boldsymbol{\theta}}^{\mathrm{RLBF}}(t) &= \mathbf{F}_0 \widehat{\boldsymbol{\alpha}}^{\mathrm{RLBF}}(t), \end{aligned} \quad (15)$$

where $\mu > 0$ denotes regularization gain (which will be optimized later on) and $\mathbf{S}(t) = \mathbf{P}(t) + \mu \mathbf{I}_{mn}$. Note that, unlike $\mathbf{P}(t)$, the matrix $\mathbf{S}(t)$ is guaranteed to be nonsingular and hence invertible, which is a clear advantage of regularization. For $m = 1$ the necessary and sufficient conditions of stochastic inveribility of $\mathbf{P}(t)$ are given in [22] but it seems that an extension of this result to the $m > 1$ case has not yet been worked out. However, from the practical viewpoint, invertibility of $\mathbf{P}(t)$ is not a problem unless the ratio $K/(mn)$ becomes too close to 1.

In a similar way, one may design the ridge version of the fLBF estimator, further denoted by fRLBF

$$\begin{aligned} \widehat{\boldsymbol{\alpha}}_j^{\mathrm{fRLBF}}(t) &= \arg\min_{\boldsymbol{\alpha}_j} \left\{ \sum_{i=-k}^{k} w(i) |\widetilde{\theta}_j(t) - \mathbf{f}^{\mathrm{T}}(i)\boldsymbol{\alpha}_j|_2^2 \right. \\ &\qquad \left. + \mu ||\boldsymbol{\alpha}_j||_2^2 \right\} = \frac{\widehat{\boldsymbol{\alpha}}_j^{\mathrm{fLBF}}(t)}{1+\mu} \\ \widehat{\theta}_j^{\mathrm{fRLBF}}(t) &= \mathbf{f}_0^{\mathrm{T}} \widehat{\boldsymbol{\alpha}}_j^{\mathrm{fRLBF}}(t) = \frac{\widehat{\theta}_j^{\mathrm{fLBF}}(t)}{1+\mu} \\ &\qquad j = 1, \ldots, n \end{aligned} \quad (16)$$

Since $\mu > 0$, the fRLBF estimates $\widehat{\theta}_j^{\mathrm{fRLBF}}(t)$ can be obtained directly by shrinking the corresponding fLBF estimates, i.e., there is no need to evaluate $\widehat{\boldsymbol{\alpha}}_j^{\mathrm{fRLBF}}(t)$.

Note that, unlike RLBF, in the fRLBF case system parameters are estimated independently of each other. This means that, in principle, one could incorporate in (16) different regularization gains $\mu_1, \ldots, \mu_n$ instead of a single gain μ. Since for large values of n the price that has to be paid for this increased estimation flexibility is a dramatic increase of the gain optimization cost (which is typical of all combinatorial optimization problems), here and later we will not take advantage of this opportunity.

B. Reweighted LBF and fLBF estimators

While ridge regression reduces (shrinks) the values of parameter estimates, in many cases a better bias-variance trade-off can be reached by setting some of them (the least "informative" ones) to zero. This can be achieved if the L^2 penalty terms in (15) and (16) are replaced with the L^1 ones, leading to the well-known LASSO approach [27]. The LASSO variants of the minimization problems carried out in (15) and (16) take the form

$$\arg\min_{\boldsymbol{\alpha}} \left\{ \sum_{i=-k}^{k} w(i)|y(t+i)-\boldsymbol{\alpha}^{\mathbf{H}}\boldsymbol{\psi}(t,i)|^2+\mu||\boldsymbol{\alpha}||_1 \right\} \quad (17)$$

and

$$\arg\min_{\boldsymbol{\alpha}_j} \left\{ \sum_{i=-k}^{k} w(i)|\widetilde{\theta}_j(t)-\mathbf{f}^{\mathrm{T}}(i)\boldsymbol{\alpha}_j|^2+\mu||\boldsymbol{\alpha}_j||_1 \right\} \quad (18)$$
$$j=1,\ldots,n,$$

respectively, where $||\mathbf{x}||_1=\sum_{i=1}^n |x_i|$ denotes the L^1 norm of a complex-valued vector $\mathbf{x}=[x_1,\ldots,x_n]^{\mathrm{T}}$.

In the time-varying estimation context the problem with LASSO emerges when it comes to optimization of the regularization gain μ. The presence of the weighting sequence $\{w(i)\}$ in (17) and (18) prevents the application of the so-called empirical Bayesian approach [29], based on stochastic embedding/reinterpretation of the minimized cost function (unless some heuristic techniques are used). Although applicable, the second, frequently used optimization approach, based on cross-validation, is computationally very expensive (especially in the LBF case). This is a serious drawback when estimation is performed in the sliding window mode, i.e., when it must be repeated for consecutive values of t.

The gain optimization problem mentioned above is a motivation to replace the L^1 regularization terms in (17) and (18) with the appropriately reweighted L^2 regularizers. Reweighting is a well-known optimization technique [30], [31]. Note that the L^1 norm of $\mathbf{x}$ can be written down in the form

$$||\mathbf{x}||_1=\sum_{i=1}^{n}\frac{x_i^* x_i}{|x_i|}=\mathbf{x}^{\mathrm{H}}\mathbf{W}\mathbf{x}=||\mathbf{W}^{1/2}\mathbf{x}||_2^2 \quad (19)$$

where

$$\mathbf{W}=\mathrm{diag}\left\{\frac{1}{|x_1|},\ldots,\frac{1}{|x_n|}\right\}.$$

Since in the case considered the values of the hyperparameters making up the vector $\boldsymbol{\alpha}$ are not known, one can proceed in two steps as follows:

Step 1

Compute the LBF/fLBF estimates of $\boldsymbol{\alpha}$ using (6)/(8).

Step 2

Solve (17)/(18) after replacing the L^1 penalty terms with their first order approximations $||\mathbf{W}^{1/2}(t)\boldsymbol{\alpha}||_2^2\cong||\boldsymbol{\alpha}||_1$ and $||\mathbf{W}_j^{1/2}(t)\boldsymbol{\alpha}_j||_2^2\cong||\boldsymbol{\alpha}_j||_1$, where

$$\begin{aligned}\mathbf{W}(t)&=\mathrm{diag}\Big\{|\widehat{a}_{11}^{\mathrm{LBF}}(t)|^{-1},\ldots,|\widehat{a}_{1m}^{\mathrm{LBF}}(t)|^{-1},\ldots,\\&\qquad|\widehat{a}_{n1}^{\mathrm{LBF}}(t)|^{-1},\ldots,|\widehat{a}_{nm}^{\mathrm{LBF}}(t)|^{-1}\Big\},\\ \mathbf{W}_j(t)&=\mathrm{diag}\Big\{|\widehat{a}_{j1}^{\mathrm{fLBF}}(t)|^{-1},\ldots,|\widehat{a}_{jm}^{\mathrm{fLBF}}(t)|^{-1}\Big\}.\end{aligned} \quad (20)$$

Using this technique, the reweighted LBF (rLBF) estimators can be obtained in the form

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}^{\mathrm{rLBF}}(t)&=\arg\min_{\boldsymbol{\alpha}}\Big\{\sum_{i=-k}^{k}w(i)|y(t+i)-\boldsymbol{\alpha}^{\mathbf{H}}\boldsymbol{\psi}(t,i)|_2^2\\&\qquad+\mu||\mathbf{W}^{1/2}(t)\boldsymbol{\alpha}||_2^2\Big\}=\mathbf{Q}^{-1}(t)\mathbf{p}(t)\\ \widehat{\boldsymbol{\theta}}^{\mathrm{rLBF}}(t)&=\mathbf{F}_0\widehat{\boldsymbol{\alpha}}^{\mathrm{rLBF}}(t),\end{aligned} \quad (21)$$

where $\mathbf{Q}(t)=\mathbf{P}(t)+\mu\mathbf{W}(t)$.

Similarly, the fast reweighted LBF (frLBF) estimators can be evaluated using

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_j^{\mathrm{frLBF}}(t)&=\arg\min_{\boldsymbol{\alpha}_j}\Big\{\sum_{i=-k}^{k}w(i)|\widetilde{\theta}_j(t)-\mathbf{f}^{\mathrm{T}}(i)\boldsymbol{\alpha}_j|_2^2\\&\qquad+\mu||\mathbf{W}^{1/2}(t)\boldsymbol{\alpha}_j||_2^2\Big\}=[\mathbf{I}_m+\mu\mathbf{W}_j(t)]^{-1}\widehat{\boldsymbol{\alpha}}_j^{\mathrm{fLBF}}(t)\\ \widehat{\theta}_j^{\mathrm{frLBF}}(t)&=\mathbf{f}_0^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}_j^{\mathrm{frLBF}}(t)\\&\quad j=1,\ldots,n.\end{aligned} \quad (22)$$

Since it holds that

$$\begin{aligned}\mathbf{V}_j^{-1}(t)&=[\mathbf{I}_m+\mu\mathbf{W}_j(t)]^{-1}\\&=\mathrm{diag}\left\{\frac{|\widehat{a}_{j1}^{\mathrm{fLBF}}(t)|}{\mu+|\widehat{a}_{j1}^{\mathrm{fLBF}}(t)|},\ldots,\frac{|\widehat{a}_{jm}^{\mathrm{fLBF}}(t)|}{\mu+|\widehat{a}_{jm}^{\mathrm{fLBF}}(t)|}\right\}\end{aligned} \quad (23)$$

evaluation of (22) does not require matrix inversion.

In the next section we will show that the problem of optimization of the regularization gain μ for the proposed reweighted LBF/fLBF estimators can be solved in a pretty straightforward and computationally efficient way using the localized version of the cross-validation approach.

Remark 1

Reweighted L^2 regularization does not retain the thresholding property of LASSO, owing to which some of the estimated coefficients may be completely discarded. Note, however, that when the estimated values of the hyperparameters a_{jl} are close to zero, the corresponding weights in (20) take very large values, which strengthens the shrinking effect compared to LBF/fLBF.

Remark 2

Similar to the approach taken in [30], [31], the two-step procedure described above can be easily extended to the multi-step (iterative) one. The iterative frLBF estimation algorithm can be obtained by replacing (22) with

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_j^{\text{frLBF}}(t,i+1) &= [\mathbf{I}_m + \mu \mathbf{W}_j(t,i)]^{-1}\widehat{\boldsymbol{\alpha}}_j^{\text{frLBF}}(t,i)\\ \widehat{\theta}_j^{\text{frLBF}}(t,i+1) &= \mathbf{f}_0^{\text{T}}\widehat{\boldsymbol{\alpha}}_j^{\text{frLBF}}(t,i+1)\end{aligned}$$

where i denotes the iteration number,

$$\mathbf{W}_j(t,i) = \text{diag}\left\{|\widehat{a}_{j1}^{\text{frLBF}}(t,i)|^{-1},\ldots,|\widehat{a}_{jm}^{\text{frLBF}}(t,i)|^{-1}\right\}$$

and initial conditions are set to $\mathbf{W}_j(t,1) = \mathbf{W}_j(t)$ and $\widehat{\boldsymbol{\alpha}}_j^{\text{frLBF}}(t,1) = \widehat{\boldsymbol{\alpha}}_j^{\text{fLBF}}(t)$.

In this way one can get closer to the results that would have been provided by the LASSO approach. As shown below, the iterative procedure can improve the MSE score by setting the values of less important coefficients closer to zero, while, unlike LASSO, allowing one to evaluate the leave-one-out cross-validation statistic in a computationally efficient way. The same technique can be used to extend the rLBF estimates, but in this case the associated computational burden is much higher.

IV. Optimization

Optimization of the regularization gain μ can be performed by a grid search. In this case several estimation algorithms, equipped with different regularization gains, are run simultaneously and compared. As a selection rule one can use the leave-one-out cross-validation approach, which is a time-localized version of the cross-validation test used for time-invariant systems. In this framework, the degree of fit of the model is defined as a local sum of squared unbiased interpolation errors (deleted residuals)

$$\varepsilon_0(t|\mu) = y(t) - [\widehat{\boldsymbol{\theta}}_0(t|\mu)]^{\mathbf{H}}\boldsymbol{\varphi}(t) \tag{24}$$

where $\widehat{\boldsymbol{\theta}}_0(t|\mu)$ is the holey estimate of $\boldsymbol{\theta}(t)$, obtained by excluding from the estimation process the "central" measurement $y(t)$. At each time instant the best fitting gain is chosen according to

$$\mu(t) = \arg\min_{\mu\in M}\sum_{i=-L}^{L}|\varepsilon_0(t+i|\mu)|^2 \tag{25}$$

where L determines the size of the local decision window and $M = \{\mu_1,\ldots,\mu_N\}$ denotes the set of grid points.

A. rLBF and RLBF estimators

The holey rLBF estimator has the form

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_0^{\text{rLBF}}(t|\mu) = \arg\min_{\boldsymbol{\alpha}}\Bigg\{&\sum_{\substack{i=-k\\ i\neq 0}}^{k} w(i)|y(t+i) - \boldsymbol{\alpha}^{\mathbf{H}}\boldsymbol{\psi}(t,i)|^2\\ &+\mu||\mathbf{W}^{-1/2}(t)\boldsymbol{\alpha}||_2^2\Bigg\} = \mathbf{Q}_0^{-1}(t)\mathbf{p}_0(t)\end{aligned}$$

where, due to the fact that $w(0)=1$, it holds that

$$\begin{aligned}\mathbf{Q}_0(t) &= \mathbf{Q}(t) - \boldsymbol{\psi}(t,0)\boldsymbol{\psi}^{\mathbf{H}}(t,0)\\ \mathbf{p}_0(t) &= \mathbf{p}(t) - y^*(t)\boldsymbol{\psi}(t,0)\end{aligned}$$

Using the matrix inversion lemma [1], one arrives at

$$\mathbf{Q}_0^{-1}(t) = \mathbf{Q}^{-1}(t) + \frac{\mathbf{Q}^{-1}(t)\boldsymbol{\psi}(t,0)\boldsymbol{\psi}^{\mathbf{H}}(t,0)\mathbf{Q}^{-1}(t)}{1-\boldsymbol{\psi}^{\mathbf{H}}(t,0)\mathbf{Q}^{-1}(t)\boldsymbol{\psi}(t,0)}$$

Straightforward calculations yield

$$\widehat{\boldsymbol{\alpha}}_0^{\text{rLBF}}(t|\mu) = \widehat{\boldsymbol{\alpha}}^{\text{rLBF}}(t|\mu) - \frac{\mathbf{Q}^{-1}(t)\boldsymbol{\psi}(t,0)[\varepsilon^{\text{rLBF}}(t|\mu)]^*}{1-\beta(t)}$$

where

$$\varepsilon^{\text{rLBF}}(t|\mu) = y(t) - [\widehat{\boldsymbol{\alpha}}^{\text{rLBF}}(t|\mu)]^{\mathbf{H}}\boldsymbol{\psi}(t,0)$$

and $\beta(t) = \boldsymbol{\psi}^{\mathbf{H}}(t,0)\mathbf{Q}^{-1}(t)\boldsymbol{\psi}(t,0)$.

Since

$$[\varepsilon_0^{\text{rLBF}}(t|\mu)]^* = y^*(t) - \boldsymbol{\psi}^{\mathbf{H}}(t,0)\widehat{\boldsymbol{\alpha}}_0^{\text{rLBF}}(t|\mu),$$

one finally obtains

$$\begin{aligned}[\varepsilon_0^{\text{rLBF}}(t|\mu)]^* &= [\varepsilon^{\text{rLBF}}(t|\mu)]^* + \frac{\beta(t)}{1-\beta(t)}[\varepsilon^{\text{rLBF}}(t|\mu)]^*\\ &= \frac{[\varepsilon^{\text{rLBF}}(t|\mu)]^*}{1-\beta(t)},\end{aligned}$$

which is equivalent to

$$\varepsilon_0^{\text{rLBF}}(t|\mu) = \frac{\varepsilon^{\text{rLBF}}(t|\mu)}{1-\beta(t)}. \tag{26}$$

The analogous expression for the RLBF scheme can be obtained by setting $\mathbf{Q}(t) = \mathbf{S}(t)$.

B. frLBF estimator

The holey frLBF estimator has the form

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_{j0}^{\text{frLBF}}(t|\mu) = \arg\min_{\boldsymbol{\alpha}_j}\Bigg\{&\sum_{\substack{i=-k\\ i\neq 0}}^{k} w(i)|\widetilde{\theta}_j(t) - \mathbf{f}^{\text{T}}(i)\boldsymbol{\alpha}_j|^2\\ &+\mu||\mathbf{W}_j^{-1/2}(t)\boldsymbol{\alpha}||_2^2\Bigg\}\\ &= [\mathbf{V}_j(t) - \mathbf{f}_0\mathbf{f}_0^{\text{T}}]^{-1}[\widehat{\boldsymbol{\alpha}}_j^{\text{fLBF}}(t) - \mathbf{f}_0\widetilde{\theta}_j(t)]\end{aligned}$$

Using the matrix inversion lemma, one gets

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_{j0}^{\text{frLBF}}(t|\mu) = &\left[\mathbf{I}_m + \mathbf{V}_j^{-1}(t)\frac{\mathbf{f}_0\mathbf{f}_0^{\text{T}}}{1-\mathbf{f}_0^{\text{T}}\mathbf{V}_j^{-1}(t)\mathbf{f}_0}\right]\widehat{\boldsymbol{\alpha}}_j^{\text{frLBF}}(t)\\ &- \frac{\mathbf{V}_j^{-1}(t)\mathbf{f}_0}{1-\mathbf{f}_0^{\text{T}}\mathbf{V}_j^{-1}(t)\mathbf{f}_0}\widetilde{\theta}_j(t)\end{aligned}$$

Since $\widehat{\boldsymbol{\theta}}_{j0}^{\text{frLBF}}(t|\mu) = \mathbf{f}_0^{\text{T}}\widehat{\boldsymbol{\alpha}}_{j0}^{\text{frLBF}}(t|\mu)$, one obtains

$$\widehat{\boldsymbol{\theta}}_{j0}^{\text{frLBF}}(t|\mu) = \frac{1}{1-c_j(t)}[\widehat{\boldsymbol{\theta}}_j^{\text{frLBF}}(t|\mu) - c_j(t)\widetilde{\boldsymbol{\theta}}_j(t)] \tag{27}$$

where $c_j(t) = \mathbf{f}_0^{\text{T}}\mathbf{V}_j^{-1}(t)\mathbf{f}_0$.

C. fRLBF estimator

In this case

$$\widehat{\boldsymbol{\alpha}}_0^{\mathrm{fRLBF}}(t|\mu) = \arg\min_{\boldsymbol{\alpha}} \Bigg\{ \sum_{\substack{i=-k\\ i\neq 0}}^{k} w(i)|\widetilde{\theta}(t) - \mathbf{F}^{\mathrm{T}}(i)\boldsymbol{\alpha}|^2 + \mu||\boldsymbol{\alpha}||_2^2 \Bigg\}$$
$$= [(1+\mu)\mathbf{I}_{mn} - \mathbf{F}_0\mathbf{F}_0^{\mathrm{T}}]^{-1}[\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t) - \mathbf{F}_0\widetilde{\theta}(t)],$$

Using the matrix inversion lemma [1] and exploiting the well-known property of Kronecker products $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = \mathbf{AC}\otimes\mathbf{BD}$, one obtains

$$[(1+\mu)\mathbf{I}_{mn} - \mathbf{F}_0\mathbf{F}_0^{\mathrm{T}}]^{-1} = \frac{1}{1+\mu}\mathbf{I}_{mn} + $$
$$+ \frac{1}{(1+\mu)^2}[\mathbf{I}_n - \frac{1}{1+\mu}\mathbf{I}_n \otimes \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0]^{-1} \otimes \mathbf{f}_0\mathbf{f}_0^{\mathrm{T}}$$
$$= \frac{1}{1+\mu}\mathbf{I}_n \otimes \left[\mathbf{I}_m + \frac{\mathbf{f}_0\mathbf{f}_0^{\mathrm{T}}}{1+\mu-\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}\right].$$

Note also that

$$\frac{1}{1+\mu}\mathbf{I}_n \otimes \left[\mathbf{I}_m + \frac{\mathbf{f}_0\mathbf{f}_0^{\mathrm{T}}}{1+\mu-\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}\right]\mathbf{F}_0\widetilde{\theta}(t) =$$
$$= \left[\mathbf{I}_n \otimes \frac{\mathbf{f}_0}{1+\mu-\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}\right]\widetilde{\boldsymbol{\theta}}(t)$$

Since $\widehat{\boldsymbol{\theta}}_0^{\mathrm{fRLBF}}(t|\mu) = \mathbf{F}_0^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}_0^{\mathrm{fRLBF}}(t|\mu)$, after straightforward calculations one obtains

$$\widehat{\boldsymbol{\theta}}_0^{\mathrm{fRLBF}}(t|\mu) = \frac{1}{1+\mu-\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}[\widehat{\boldsymbol{\theta}}^{\mathrm{fRLBF}}(t|\mu) - \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\widetilde{\boldsymbol{\theta}}(t)],$$

which finally leads to

$$\varepsilon_0^{\mathrm{fRLBF}}(t|\mu) = \frac{\varepsilon^{\mathrm{fRLBF}}(t|\mu) - \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\widetilde{\varepsilon}(t) + \mu y(t)}{1+\mu-\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0} \tag{28}$$

where $\varepsilon^{\mathrm{fRLBF}}(t|\mu) = y(t) - [\widehat{\boldsymbol{\theta}}^{\mathrm{fRLBF}}(t|\mu)]^{\mathbf{H}}\boldsymbol{\varphi}(t)$ and $\widetilde{\varepsilon}(t) = y(t) - \widetilde{\boldsymbol{\theta}}^{\mathbf{H}}(t)\boldsymbol{\varphi}(t)$.

Remark 3

Note that in all cases discussed above, the leave-one-out interpolation errors can be determined without the need to implement the corresponding holey estimation schemes.

V. Simulation results

Computer simulations were arranged to compare the plain LBF and fLBF algorithms with their regularized (ridge and reweighted) versions. A simulated underwater acoustic (UWA) communication channel, described in [28], was used as a testbed for comparison. Such a system can be modeled as a 50-tap FIR system with time-varying impulse response. Following [28], the complex-valued impulse response coefficients $\theta_j(t), j = 1, \ldots, 50$, varying independently of each other, were generated by lowpass filtering of a circular white Gaussian noise. The cutoff frequency of the forming filter was set to 1 Hz under 1 kHz sampling. The resulting parameter changes, shown in Fig. 1, can be regarded as fast in the UWA case. Each parameter trajectory was randomly scaled. To reflect the exponentially decaying power delay profile, caused by the spreading and absorbtion loss [7], the scaling coefficients were chosen so that

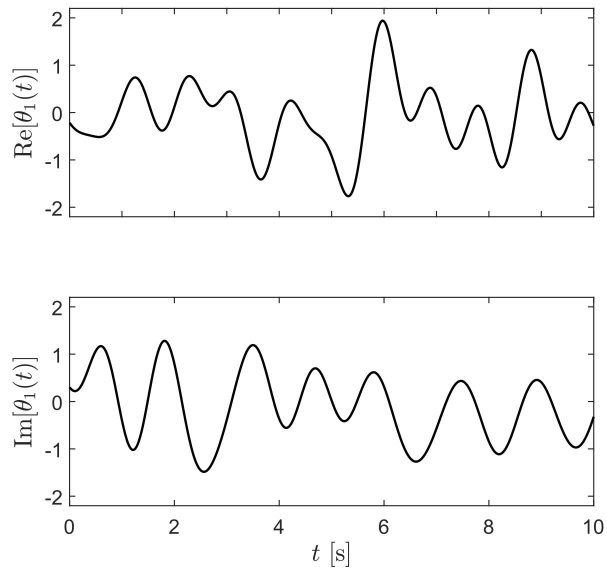

Fig. 1: Real and imaginary parts of a typical parameter trajectory prior to scaling.

$$\mathrm{var}[\theta_j(t)] = \gamma^{j-1}, \tag{29}$$

where $\gamma = 0.69$. Under such settings the ratio between the variance of the first arrival ($j = 1$) and the last arrival ($j = 50$) is equal to 80 dB. A snapshot of the time-varying impulse response of the simulated underwater communication channel and its rLBF/frLBF estimates obtained for SNR=20 dB are shown in Fig. 2.

The input signal was circular white binary $u(t) = \pm 1 \pm i$ and the measurement noise was circular white Gaussian with variance σ_e^2 equal to 0.65, 0.065 and 0.0065 which corresponds to the signal-to-noise ratio (SNR)

$$\mathrm{SNR} = \frac{\mathrm{E}[|\boldsymbol{\theta}^{\mathbf{H}}\boldsymbol{\varphi}(t)|^2]}{\sigma_e^2} = \frac{\sigma_u^2}{\sigma_e^2}\sum_{j=1}^{50}\mathrm{var}[\theta_j(t)] \tag{30}$$

equal to 10, 20 and 30 dB respectively.

To avoid boundary problems, data generation was started 1000 time instants prior to $t = 1$ and was continued for 1000 time instants after $t = T_s$, where $T_s = 10000$ denotes the simulation time. The preestimation forgetting constant was set to $\lambda_0 = 0.96$ and the adopted weighting sequence had the form $w(i) = \cos\frac{\pi i}{2k}, i \in I_k$ (recursively computable cosinusoidal window). Prior to orthonormalization the basis set was made up of powers of time: $g_l(i) = i^{l-1}, l = 1, \ldots, m, i \in I_k$. The LBF and fLBF algorithms were run with $k = 100$, $m = 3$, and the choice of μ was restricted to the set $M = \{0.1, 0.2, 0.4, 0.8, 1.6\}$. The half-width of the decision window was set to $L = 30$.

Performance of the compared algorithms was evaluated

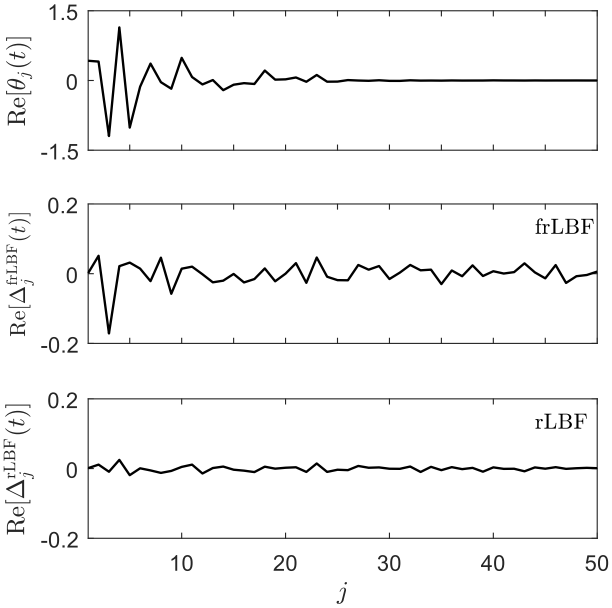

Fig. 2: A snapshot of the time-varying impulse response of the simulated underwater communication channel (top figure) and the corresponding parameter estimation errors obtained using the frLBF approach and rLBF approach, respectively (two lower figures).

using the FIT measure, defined as follows [24]

$$\mathrm{FIT}(t) = 100 \left(1 - \left[\frac{\sum_{j=1}^{50} |\theta_j(t) - \widehat{\theta}_j(t)|^2}{\sum_{j=1}^{50} |\theta_j(t) - \overline{\theta}_j(t)|^2} \right]^{1/2} \right), \quad (31)$$

where $\overline{\theta}_j(t) = \frac{1}{50} \sum_{j=1}^{50} \theta_j(t)$. The maximum value of this measure, equal to 100, corresponds to the perfect match between the true and estimated impulse responses. The final scores, further referred to as FIT (%), were obtained by combined time averaging (over $[1, T_s]$) and ensemble averaging (over 20 independent realizations of $\boldsymbol{\theta}(t)$).

According to the simulation results, summarized in Tab. I, ridge regularization noticeably improves the MSE score only in the low SNR case. For SNR>10 dB, a very small improvement of approximately 0.02% can be observed in the range of very small values of μ (around 0.002). Since improvement of this order is negligible from the practical viewpoint, it was not included in our comparison. On the other hand, the well-tuned reweighted L^2 regularization yields improvement for all considered SNR's. As expected, the greatest performance gain can be observed in the low SNR range.

Fig. 3 shows the time-averaged FIT scores obtained for the adaptive rLBF/frLBF algorithms for all 20 realizations of parameter trajectories (corresponding to different sets of randomly selected scaling coefficients). Note that the adaptive algorithms with regularization yield consistently better results (better in all cases) than the not regularized ones.

TABLE I: Average FIT [%] scores obtained for LBF/fLBF estimators and their regularized versions: ridge (RLBF/fRLBF) and reweighted (rLBF/frLBF). The algorithm with adaptive choice of the regularization gain is labelled by A.

	Method \ μ	0.1	0.2	0.4	0.8	1.6	A
10 dB	LBF	57.51					
	RLBF	63.52	63.48	60.13	52.60	41.84	64.49
	rLBF	65.65	69.40	73.48	76.79	**77.45**	74.55
	fLBF	65.83					
	fRLBF	67.33	66.36	61.45	50.65	36.12	67.34
	frLBF	69.55	72.09	75.16	**77.10**	74.83	74.81
20 dB	LBF	86.56					
	RLBF	82.96	77.24	68.30	56.69	43.61	82.96
	rLBF	90.76	**91.55**	91.30	89.09	84.41	91.39
	fLBF	85.71					
	fRLBF	83.71	78.87	68.92	54.07	37.38	83.71
	frLBF	87.93	**88.52**	87.85	84.55	77.83	88.36
30 dB	LBF	95.74					
	RLBF	86.63	79.27	69.30	57.14	43.79	86.63
	rLBF	**96.77**	95.84	93.86	90.51	85.27	96.77
	fLBF	89.76					
	fRLBF	86.51	80.65	69.81	54.44	37.51	86.51
	frLBF	91.01	90.73	89.03	85.01	77.97	**91.04**

In the second experiment the same data set was used to check whether the iterative multi-step procedure, described in the previous section (see Remark 2), can improve the MSE scores. The obtained results, shown in Tab. II, confirm that this is the case provided that the regularization gain is well tuned (if not, the results deteriorate as the number of iterations grows). Since the adaptive algorithm always picks the best variant, successive iterations improve its performance – steadily albeit not significantly.

Finally, we note that the situation does not change if the design parameters k and m are chosen in an adaptive manner described in [20] – the regularized algorithms continue to perform better than the not regularized ones.

TABLE II: Average FIT [%] scores obtained for the iterative multi-step frLBF algorithm (the number of iterations is given next to the method label).

	Method \ μ	0.1	0.2	0.4	0.8	1.6	A
10 dB	fLBF	65.83					
	frLBF 1	69.55	72.09	75.16	77.10	74.83	74.81
	frLBF 2	70.06	73.35	77.35	78.59	72.55	75.61
	frLBF 3	70.15	73.73	78.19	78.99	71.27	75.56
	frLBF 4	70.18	73.87	78.57	79.14	70.59	75.48
	frLBF 5	70.19	73.93	78.78	**79.21**	70.19	75.42
20 dB	fLBF	85.71					
	frLBF 1	87.93	88.52	87.85	84.55	77.83	88.36
	frLBF 2	88.41	89.05	87.61	82.63	73.42	88.73
	frLBF 3	88.56	89.25	87.47	81.88	71.75	88.84
	frLBF 4	88.63	89.34	87.39	81.53	70.96	88.88
	frLBF 5	88.66	**89.39**	87.33	81.33	70.52	88.90
30dB	fLBF	89.76					
	frLBF 1	91.01	90.73	89.03	85.01	77.97	91.04
	frLBF 2	91.24	90.71	88.22	82.79	73.46	91.27
	frLBF 3	91.33	90.68	87.90	82.00	71.79	91.36
	frLBF 4	91.37	90.66	87.75	81.63	71.00	91.40
	frLBF 5	91.39	90.65	87.66	81.44	70.56	**91.42**

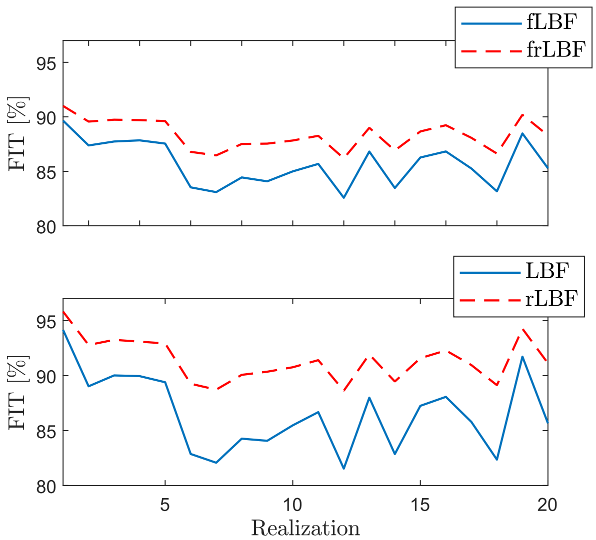

Fig. 3: Time-averaged FIT [%] scores obtained for all 20 realizations of parameter trajectories. Upper figure shows the results obtained for the fLBF algorithm and its adaptively reweighted version. Lower figure shows the analogous results obtained for the LBF algorithm and its adaptively reweighted version (SNR=20 dB).

VI. Conclusion

The problem of identification of a time-varying FIR system was considered. It was shown that accuracy of the recently proposed local basis function (LBF) estimators and their computationally fast versions (fLBF) can be improved by means of regularization. Two variants of regularization were examined: the classical L^2 (ridge) regularization and a new, reweighted L^2 regularization. In both cases optimization of the regularization gain was carried out using the time-localized version of the cross-validation approach. As shown in a realistic underwater acoustic channel identification experiment, the reweighted LBF/fLBF estimators perform better, in terms of the FIT measure, than their original and ridge versions. One of the interesting directions of the future research would be to include in the identification process some prior knowledge (whenever available) about the estimated impulse response, such as degree of its smoothness or the rate of decay.

References

[1] T. Söderström and P. Stoica, *System Identification*, Prentice-Hall, 1988.

[2] S. Haykin, *Adaptive Filter Theory*. Prentice-Hall, 1996.

[3] M. Niedźwiecki, *Identification of Time-varying Processes*, Wiley, 2000.

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

[5] R. Dahlhaus, "Local inference for locally sationary time series based on the empirical spectral measure," *Journal of Econometrics*, vol. 151, pp. 101–112, 2009.

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

[7] M. Stojanovic and J. Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," *IEEE Communications Magazine*, vol. 47, pp. 84-89, 2009.

[8] 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.

[9] L. A. Liporace, "Linear estimation of nonstationary signals," *Journal of the Acoustical Society of America*, vol. 58, pp. 1288–1295, 1975.

[10] Y. Grenier, "Time-dependent ARMA modeling of nonstationary signals," *IEEE Trans. Acoust. speech Signal Process.*, vol. 31, pp. 899-911, 1981.

[11] M. Niedźwiecki, "Functional series modeling approach to identification of nonstationary stochastic systems," *IEEE Transactions on Automatic Control*, vol. 33, pp. 955-961, 1988.

[12] M. K. Tsatsanis and G. B. Giannakis "Time-varying system identification and model reduction using wavelets," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3512-3523, 1994.

[13] G. B. Giannakis and C. Tepedelenlioĝlu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proceedings of IEEE*, vol. 86, pp. 1969–1986, 1998.

[14] H. L. Wei, J. J. Liu and S. A. Billings, "Identification of time-varying systems using multi-resolution wavelet models," *International Journal of Systems Science*, vol. 33, pp. 1217-1228, 2002.

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

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

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

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

[19] M. Niedźwiecki, "Locally adaptive cooperative Kalman smoothing and its application to identification of nonstationary stochastic systems," *IEEE Transactions on Signal Processing*, vol. 60, pp. 48-59, 2012.

[20] M. Niedźwiecki and M. Ciołek, "Generalized Savitzky-Golay filters for identification of nonstationary systems," *Automatica*, vol. 108, no. 108522, 2019.

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

[22] M. Niedźwiecki and L. Guo, "Nonasymptotic results for finite-memory WLS filters," *IEEE Transactions on Automatic Control*, vol. AC-36, pp. 515-522, 1991.

[23] M. Niedźwiecki, A. Gańcza and M. Ciołek, "On the preestimation technique and its application to identification of nonstationary systems," 59th Conference on Decision and Control CDC 2020, Jeju Island, Republic of Korea, pp. 286-293, 2020.

[24] 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.

[25] L. Ljung and T. Chen, "What can regularization offer for estimation of dynamical systems?," *Proc. of the 11th IFAC Workshop on Adaptation and Learning in Control and Signal Processing*, Caen, France, pp. 1-8, 2013.

[26] A.E. Hoerl, "Application of Ridge Analysis to Regression Problems," Chemical Engineering Progress. vol. 58, pp. 54–59, 1962.

[27] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. R. Statist. Soc. B*, vol. 58, pp. 267-288, 1996.

[28] L. Shen, Y. Zakharov, B. Henson, 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.

[29] B.P. Carlin and T.A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman & Hall, 1996.

[30] M.S. Bani and B.L. Chalmers, "Best approximation in L^∞ via iterative Hilbert space procedures," *Journal of Approximation Theory*,vol. 42, pp. 173-180, 1984.

[31] C.S. Burrus, J.A. Barreto and I.W. Selesnick, "Iterative reweighted least squares design of FIR filters," *IEEE Transactions on Signal Processing*, vol. 42, pp.2926-2936, 1994.