Postprint of: Niedźwiecki M., Gańcza A., Kaczmarek P., Identification of Fast Time-varying Communication Channels Using the Preestimation Technique, IFAC-PapersOnLine, Vol. 54, Iss. 7 (2021), pp. 351-356, DOI: 10.1016/j.ifacol.2021.08.384

Identification of Fast Time-varying Communication Channels Using the Preestimation Technique

Maciej Niedźwiecki * **Artur Gańcza** * **Piotr Kaczmarek** *

* *Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control, Gdańsk University of Technology ul. Narutowicza 11/12, Gdańsk , Poland e-mail: maciekn@eti.pg.gda.pl, artgancz@student.pg.edu.pl, piokaczm@pg.edu.pl*

Abstract: Accurate identification of stochastic systems with fast-varying parameters is a challenging task which cannot be accomplished using model-free estimation methods, such as weighted least squares, which assume only that system coefficients can be regarded as locally constant. The current state-of-the-art solutions are based on the assumption that system parameters can be locally approximated by a linear combination of appropriately chosen basis functions. The paper shows that tracking performance of the resulting local basis function estimation algorithms can be further improved by means of regularization. The method is illustrated by an important recent application – identification of fast time-varying acoustic channels used in underwater communication.

Keywords: identification of time-varying systems, underwater acoustic communication channels, local basis function approach, regularization, leave-one-out cross-validation

1. INTRODUCTION

Many nonstationary communication channels (terrestrial, underwater) can be well approximated by a time-varying finite impulse response (FIR) model of the form, Tsatsanis & Giannakis [1996], Stojanovic & Preisig [2009]

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

where $t = \ldots, -1, 0, 1, \ldots$ denotes discrete (normalized) time, $y(t)$ denotes the received complex-valued signal, $\boldsymbol{\varphi}(t) = [u(t), \ldots, u(t-n+1)]^{\mathrm{T}}$ denotes regression vector made up of past values of the complex-valued transmitted signal, $\boldsymbol{\theta}(t) = [\theta_1(t), \ldots, \theta_n(t)]^{\mathrm{T}}$ denotes the vector of time-varying channel impulse response coefficients, and $\{e(t)\}$ denotes white noise independent of $\{u(t)\}$ and $\{\boldsymbol{\theta}(t)\}$. The symbol $*$ denotes complex conjugate and H – complex conjugate transpose (Hermitian transpose).

The application, studied recently, which particularly well fits the technique developed in this paper, is adaptive self-interference cancellation in full-duplex (FD) underwater acoustic (UWA) communication systems, Shen et al. [2020a], Shen et al. [2020b]. FD UWA systems, designed to maximize the limited capacity of acoustic links, simultaneously transmit and receive data in the same frequency band. Due to the close spacing of the transmit and receive antennas, the far-end signal is strongly contaminated by the so-called self-interference (SI) introduced by the near-end transmitter. Self-interference is a multipath propagation effect caused, among others, by multiple reflections of the emitted signal from the water surface and/or the bottom. The model of the received signal is given by (1), where $\{u(t)\}$ denotes the near-end (known) signal and $\{e(t)\}$ is a mixture of the far-end signal and the channel noise (ambient and/or site-specific). Note that in this case our goal is extraction of the signal $\{e(t)\}$ from $\{y(t)\}$, which can be easily done provided that channel parameters are known. Adaptive (on-line) identification of the channel is needed due to its time variability – the effect caused by the transmitter/receiver motion and/or by inherent changes in the propagation medium. An interesting feature of this application is that it allows one to work with a decision delay, which means that estimation of channel parameters can be based not only on past signal samples but also on a certain number of "future" (with respect to the moment of interest) ones. Hence, channel identification can be carried out using noncausal estimation algorithms with improved tracking capabilities, such as the ones described in this paper.

When channel coefficients vary slowly with time, their estimation can be carried out using the localized versions of the least squares approach, such as exponentially weighted least squares or sliding window least squares, Söderström & Stoica [1988]. The corresponding estimation algorithms are not based on any explicit model of parameter variation – it is only assumed that system parameters can be regarded as "locally constant". In the case of rapidly fading channels such a simple estimation

⋆ This work was partially supported by the National Science Center under the agreement UMO-2018/29/B/ST7/00325. Computer simulations were carried out at the Academic Computer Centre in Gdańsk.

strategy fails because the achievable estimation accuracy is not sufficient to guarantee satisfactory operation of the underlying communication system, Shen et al. [2020a]. Fast parameter changes can be tracked successfully using basis expansion methods. In this approach each parameter trajectory is locally approximated by a linear combination of known functions of time, called basis functions. The resulting local basis function (LBF) estimation algorithms, Niedźwiecki & Ciołek [2019], have very good parameter tracking capabilities and have already proven their supremacy over the classical weighted least squares solutions in UWA applications, Shen et al. [2020a]. LBF algorithms are computationally demanding, especially when the number of estimation coefficients is large. In the paper Niedźwiecki, Ciołek & Gańcza [2020a], which is a follow-up to Niedźwiecki & Ciołek [2019], it was shown that the computational structure of LBF estimators can be significantly simplified without compromising their accuracy. 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 current paper shows that the performance of fast LBF algorithms can be further improved using regularization.

2. PREESTIMATION

Preestimation is a technique introduced in Niedźwiecki & Kłaput [2002] and further developed in Niedźwiecki, Ciołek & Gańcza [2020a], Niedźwiecki, Gańcza & Ciołek [2020b]. Preestimates are raw parameter estimates, unbiased but with a very large variability. For this reason, to obtain reliable parameter estimates providing satisfactory bias-variance trade-off, preestimates must be further postfiltered. As shown in Niedźwiecki, Ciołek & Gańcza [2020a], preestimates can be obtained by "inverse filtering" of short-memory exponentially weighted least squares (EWLS) estimates. In the current paper, taking advantage of the fact that the input sequence is white (which is typical of all wireless communication systems), we will apply a computationally simpler solution which incorporates gradient least mean squares (LMS) estimates obtained from

$$\begin{aligned}\widehat{\boldsymbol{\theta}}^{\mathrm{LMS}}(t) &= \widehat{\boldsymbol{\theta}}^{\mathrm{LMS}}(t-1) + \frac{\mu}{\sigma_u^2}\boldsymbol{\varphi}(t)\varepsilon^*(t) \\ \varepsilon(t) &= y(t) - \left[\widehat{\boldsymbol{\theta}}^{\mathrm{LMS}}(t-1)\right]^{\mathrm{H}}\boldsymbol{\varphi}(t)\end{aligned} \tag{2}$$

where $\mu > 0$ denotes a small stepsize parameter. The value of μ should be sufficiently small to guarantee boundedness of the parameter tracking error $\widehat{\boldsymbol{\theta}}^{\mathrm{LMS}}(t) - \boldsymbol{\theta}(t)$. Analysis carried out in the time-invariant case and i.i.d. Gaussian regressors shows that to guarantee stability of the LMS algorithm in the mean square sense, the value of μ should be smaller than $2/(n+1)$.

The preestimates, further denoted by $\widetilde{\boldsymbol{\theta}}(t)$, can be defined as follows

$$\widetilde{\boldsymbol{\theta}}(t) = \frac{\widehat{\boldsymbol{\theta}}^{\mathrm{LMS}}(t) - \lambda\widehat{\boldsymbol{\theta}}^{\mathrm{LMS}}(t-1)}{1-\lambda} \tag{3}$$

where $\lambda = 1 - \mu$.

It is easy to check that

$$\begin{aligned}\widetilde{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t) &= \left[\mathbf{I}_n - \frac{\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{H}}(t)}{\sigma_u^2}\right]\left[\widehat{\boldsymbol{\theta}}^{\mathrm{LMS}}(t-1) - \boldsymbol{\theta}(t)\right] \\ &\quad + \frac{1}{\sigma_u^2}\boldsymbol{\varphi}(t)e^*(t) = \mathbf{z}_1(t) + \mathbf{z}_2(t).\end{aligned} \tag{4}$$

Since $\{e(t)\}$ is a zero-mean white noise, independent of $\{\boldsymbol{\varphi}(t)\}$, it holds that $\mathrm{E}[\mathbf{z}_2(t)] = 0$. Furthermore, since in the case considered $\mathrm{E}[\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{H}}(t)] = \sigma_u^2\mathbf{I}_n$, using the averaging theory, Guo & Ljung [1995], one can show that $\mathrm{E}[\mathbf{z}_1(t)] \cong 0$ (exact equality holds when $\{\boldsymbol{\varphi}(t)\}$ is an i.i.d. sequence). Hence, $\mathrm{E}[\widetilde{\boldsymbol{\theta}}(t)] \cong \boldsymbol{\theta}(t)$, which means that the preestimate $\widetilde{\boldsymbol{\theta}}(t)$ is approximately unbiased and can be written down as

$$\widetilde{\boldsymbol{\theta}}(t) \cong \boldsymbol{\theta}(t) + \mathbf{z}(t) \tag{5}$$

where $\mathbf{z}(t)$ denotes a zero-mean noise with large covariance matrix. Using (5) the problem of identification of the time-varying system (1) can be reformulated as a problem of "denoising" the sequence of parameter preestimates obtained from (3).

3. FAST LOCAL BASIS FUNCTION ESTIMATORS

As a starting point for our further considerations we will use the postfiltering technique based on the local basis function (LBF) approximation. In the LBF approach, which is an extension of Savitzky-Golay filtering, Schafer [2011], the analyzed signal is approximated, in the sliding analysis window $T(t) = [t-k, t+k]$ of width $K = 2k+1$, by a linear combination of known linearly independent functions of time $f_1(i), \dots, f_m(i), i \in I_k = [-k, k]$, called basis functions. Typical choices of basis functions are powers of time (local Taylor expansion) or harmonic functions (local Fourier expansion). In the case considered we will assume that each coefficient of the estimated impulse response can be expressed in the form

$$\begin{aligned}\theta_j(t+i) = \mathbf{f}^{\mathrm{T}}(i)\boldsymbol{\alpha}_j(t), \quad i \in I_k \\ j = 1, \dots, n\end{aligned} \tag{6}$$

where $\mathbf{f}(i) = [f_1(i), \dots, f_m(i)]^{\mathrm{T}}$. Note that the hypermodel (6) can be expressed in a more compact form

$$\boldsymbol{\theta}(t+i) = \mathbf{F}(i)\boldsymbol{\alpha}(t), \quad i \in I_k \tag{7}$$

where

$$\mathbf{F}(i) = \mathbf{I}_n \otimes \mathbf{f}^{\mathrm{T}}(i), \quad \boldsymbol{\alpha}(t) = [\boldsymbol{\alpha}_1^{\mathrm{T}}(t), \dots, \boldsymbol{\alpha}_n^{\mathrm{T}}(t)]^{\mathrm{T}} \tag{8}$$

and $\otimes$ denotes the Kronecker product of the respective vectors and/or matrices.

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. For convenience, but without any loss of generality, we will assume that the adopted basis functions are w-orthonormal, namely $\sum_{i=-k}^{k} w(i)\mathbf{f}(i)\mathbf{f}^{\mathrm{T}}(i) = \mathbf{I}_m$. Orthonormalization of any set of basis functions can be carried out sequentially using the well-known Gram-Schmidt procedure.

Fast local basis function (fLBF) estimates of $\boldsymbol{\theta}(t)$ were defined in Niedźwiecki, Ciołek & Gańcza [2020a] in the form

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t) &= \arg\min_{\boldsymbol{\alpha}} \sum_{i=-k}^{k} w(i)||\widetilde{\boldsymbol{\theta}}(t+i) - \mathbf{F}(i)\boldsymbol{\alpha}||^2 \\ \widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t) &= \mathbf{F}_0\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t)\end{aligned} \tag{9}$$

where $\mathbf{F}_0 = \mathbf{F}(0) = \mathbf{I}_n \otimes \mathbf{f}_0^{\mathrm{T}},\ \mathbf{f}_0 = \mathbf{f}(0)$. The term "fast" refers to the fact that under typical operating conditions fLBF estimators yield results almost indistinguishable from those provided by the, computationally much more involved, LBF estimators (generalized Savitzky-Golay algorithms) proposed in Niedźwiecki & Ciołek [2019].

Using the identity $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = (\mathbf{AC})\otimes(\mathbf{BD})$ which holds true for Kronecker products, one obtains

$$\sum_{i=-k}^{k} w(i)\mathbf{F}^{\mathrm{T}}(i)\mathbf{F}(i) = \mathbf{I}_n \otimes \left[\sum_{i=-k}^{k} w(i)\mathbf{f}(i)\mathbf{f}^{\mathrm{T}}(i)\right] = \mathbf{I}_{mn} \tag{10}$$

leading to

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t) &= \sum_{i=-k}^{k} w(i)\mathbf{F}^{\mathrm{T}}(i)\widetilde{\boldsymbol{\theta}}(t+i)\\ &= \sum_{i=-k}^{k} w(i)[\widetilde{\boldsymbol{\theta}}(t+i)\otimes\mathbf{f}(i)]\end{aligned} \tag{11}$$

and finally

$$\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t) = \sum_{i=-k}^{k} h(i)\widetilde{\boldsymbol{\theta}}(t+i) \tag{12}$$

where

$$h(i) = w(i)\mathbf{f}_0^{\mathrm{T}}\mathbf{f}(i), \quad i \in I_k \tag{13}$$

is the impulse response of the postprocessing FIR filter.

Note that from the computational viewpoint the formula (12) is very simple. Additionally, for some choices of the basis and window functions the fLBF estimates can be expressed in a time-recursive form, Niedźwiecki, Ciołek & Gańcza [2020a].

4. REGULARIZED FAST LBF ESTIMATORS

Regularization is a technique which was originally introduced to solve ill-conditioned inverse problems. As shown later, regularization also allows one to improve the bias-variance trade-off of the applied estimation schemes, and hence – to increase their accuracy, Ljung & Chen [2013]. The idea is to add to the minimized cost function a term, often referred to as a regularizer, which reduces the norm of the solution. In agreement with this principle, we will introduce the L_2 regularizer of the form

$$||\boldsymbol{\theta}(t)||^2_{\mathbf{R}} = \boldsymbol{\theta}^{\mathrm{H}}(t)\mathbf{R}\boldsymbol{\theta}(t) = \boldsymbol{\alpha}^{\mathrm{H}}(t)\mathbf{F}_0^{\mathrm{T}}\mathbf{R}\mathbf{F}_0\boldsymbol{\alpha}(t) \tag{14}$$

where $\mathbf{R} = \mathbf{D}^{\mathrm{H}}\mathbf{D} > 0$ denotes the $n\times n$ positive definite regularization matrix. Note that such a regularization penalizes the norm of $\boldsymbol{\theta}(t)$, the estimation of which is a real purpose of channel identification, and only indirectly penalizes the norm of the vector of hyperparameters $\boldsymbol{\alpha}(t)$, which is not of our primary interest. The fast regularized LBF estimators (fRLBF) will be defined in the form

$$\begin{aligned}&\widehat{\boldsymbol{\alpha}}^{\mathrm{fRLBF}}(t|\mathbf{R}) =\\ &= \arg\min_{\boldsymbol{\alpha}} \left\{ \sum_{i=-k}^{k} w(i)||\widetilde{\boldsymbol{\theta}}(t+i) - \mathbf{F}(i)\boldsymbol{\alpha}||^2 + ||\boldsymbol{\alpha}||^2_{\mathbf{F}_0^{\mathrm{T}}\mathbf{R}\mathbf{F}_0}\right\}\\ &\widehat{\boldsymbol{\theta}}^{\mathrm{fRLBF}}(t|\mathbf{R}) = \mathbf{F}_0\widehat{\boldsymbol{\alpha}}^{\mathrm{fRLBF}}(t|\mathbf{R})\end{aligned} \tag{15}$$

It can be shown that (see Appendix 1)

$$\widehat{\boldsymbol{\theta}}^{\mathrm{fRLBF}}(t|\mathbf{R}) = [\mathbf{I}_n + \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{R}]^{-1}\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t) \tag{16}$$

We will use this formula to optimize the regularization matrix when some prior knowledge about statistical properties of $\{\boldsymbol{\theta}(t)\}$ is available. In the sequel we will assume that $\mathbf{Z} = \mathrm{cov}[\mathbf{z}(t)] = \sigma_z^2\mathbf{I}_n$ (which approximately holds true if the input sequence $\{u(t)\}$ is white) and that $\{\boldsymbol{\theta}(t)\}$ is a wide sense stationary process with known (or preestimated) correlation matrix $\mathrm{E}[\boldsymbol{\theta}(t)\boldsymbol{\theta}^{\mathrm{H}}(t)] = \mathbf{Q} > 0$. Additionally, we will assume that the process $\{\boldsymbol{\theta}(t)\}$ is independent of $\{u(t)\}$.

We will derive the formula for the mean square parameter estimation error matrix in the case where the parameter trajectory obeys the model (7). Note that under (7) it holds that

$$\begin{aligned}\sum_{i=-k}^{k} h(i)\boldsymbol{\theta}(t+i) &= \sum_{i=-k}^{k} w(i)\mathbf{f}_0^{\mathrm{T}}\mathbf{f}(i)[\mathbf{I}_n\otimes\mathbf{f}^{\mathrm{T}}(i)]\boldsymbol{\alpha}(t)\\ &= \left\{\mathbf{I}_n\otimes\left[\mathbf{f}_0^{\mathrm{T}}\sum_{i=-k}^{k} w(i)\mathbf{f}(i)\mathbf{f}^{\mathrm{T}}(i)\right]\right\}\boldsymbol{\alpha}(t)\\ &= [\mathbf{I}_n\otimes\mathbf{f}_0^{\mathrm{T}}]\boldsymbol{\alpha}(t) = \mathbf{F}_0\boldsymbol{\alpha}(t) = \boldsymbol{\theta}(t)\end{aligned} \tag{17}$$

leading to [cf. (5)]

$$\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t) = \boldsymbol{\theta}(t) + \sum_{i=-k}^{k} h(i)\mathbf{z}(t+i) \tag{18}$$

and

$$\begin{aligned}\widehat{\boldsymbol{\theta}}^{\mathrm{fRLBF}}(t|\mathbf{R}) &= [\mathbf{I}_n + \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{R}]^{-1}\boldsymbol{\theta}(t)\\ &+ [\mathbf{I}_n + \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{R}]^{-1}\sum_{i=-k}^{k} h(i)\mathbf{z}(t+i).\end{aligned} \tag{19}$$

Let $\widetilde{\mathbf{R}} = \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{R}$. Since

$$[\mathbf{I}_n + \widetilde{\mathbf{R}}]^{-1}\boldsymbol{\theta}(t) - \boldsymbol{\theta}(t) = -[\mathbf{I}_n + \widetilde{\mathbf{R}}]^{-1}\widetilde{\mathbf{R}}\boldsymbol{\theta}(t)$$

and all matrices involved are Hermitian, one finally obtains

$$\begin{aligned}&\mathrm{MSE}(\widetilde{\mathbf{R}})\\ &= \mathrm{E}\left\{\left[\widehat{\boldsymbol{\theta}}^{\mathrm{fRLBF}}(t|\mathbf{R}) - \boldsymbol{\theta}(t)\right]\left[\widehat{\boldsymbol{\theta}}^{\mathrm{fRLBF}}(t|\mathbf{R}) - \boldsymbol{\theta}(t)\right]^{\mathrm{H}}\right\}\\ &= [\mathbf{I}_n + \widetilde{\mathbf{R}}]^{-1}[\widetilde{\mathbf{R}}\mathbf{Q}\widetilde{\mathbf{R}} + \eta\mathbf{I}_n][\mathbf{I}_n + \widetilde{\mathbf{R}}]^{-1}\\ &= \eta[\mathbf{I}_n + \widetilde{\mathbf{R}}]^{-1}[\widetilde{\mathbf{R}}\widetilde{\mathbf{Q}}\widetilde{\mathbf{R}} + \mathbf{I}_n][\mathbf{I}_n + \widetilde{\mathbf{R}}]^{-1}\end{aligned} \tag{20}$$

where $\eta = \sigma_z^2/N_k$, $\widetilde{\mathbf{Q}} = \mathbf{Q}/\eta$ and $N_k = [\sum_{i=-k}^{k} h^2(i)]^{-1}$ denotes the equivalent width of the analysis window $T(t)$, different from its effective width $L_k = \sum_{i=-k}^{k} w(i)$ – see Niedźwiecki [2000]. The expectation in (20) is carried out over $\{\mathbf{z}(t)\}$ and $\{\boldsymbol{\theta}(t)\}$.

It can be shown that for any nonnegative definite matrix $\widetilde{\mathbf{R}}$ it holds that (see Theorem 1 in Chen, Ohlsson & Ljung [2012])

$$\mathrm{MSE}(\widetilde{\mathbf{R}}) \geq \mathrm{MSE}(\widetilde{\mathbf{Q}}^{-1}) \tag{21}$$

which means that the optimal choice of $\widetilde{\mathbf{R}}$ is given by $\widetilde{\mathbf{R}}_{\mathrm{opt}} = \widetilde{\mathbf{Q}}^{-1}$, i.e.,

$$\mathbf{R}_{\mathrm{opt}} = \frac{\sigma_z^2}{N_k\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}\mathbf{Q}^{-1}\ . \tag{22}$$

So far we have been assuming that the variance σ_z^2 is constant and known. When channel noise intensity varies with time, σ_z^2 can be replaced in (22) with its local estimate

$$\widehat{\sigma}_z^2(t) = \frac{1}{nL_k}\sum_{i=-k}^{k} w(i)||\widehat{\mathbf{z}}(t,i)||^2 \tag{23}$$

where $\widehat{\mathbf{z}}(t,i) = \widetilde{\boldsymbol{\theta}}(t+i) - \mathbf{F}(i)\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t)$. After straightforward calculations incorporating (10) and (11), one gets

$$\widehat{\sigma}_z^2(t) = \frac{1}{nL_k}\left[\sum_{i=-k}^{k} w(i)||\widetilde{\boldsymbol{\theta}}(t+i)||^2 - ||\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t)||^2\right].$$

5. ENHANCED IDENTIFICATION PROCEDURE

Since the LMS estimates, as all causal ones, are biased (the bias increases with decreasing μ and is primarily due to the estimation delay effect, Niedźwiecki [2000]), when the number of estimated parameters is large, the preestimation error (4) is dominated by its first component $\mathbf{z}_1(t)$. Similarly as in Niedźwiecki, Gańcza & Ciołek [2020b], this problem can be circumvented by replacing the preestimates $\widetilde{\boldsymbol{\theta}}(t)$ with their enhanced version

$$\boldsymbol{\theta}^{\dagger}(t) = \widehat{\boldsymbol{\theta}}^{\mathrm{fRLBF}}(t) + \frac{\mu}{\sigma_u^2}\boldsymbol{\varphi}(t)\left[y(t) - [\widehat{\boldsymbol{\theta}}^{\mathrm{fRLBF}}(t)]^{\mathrm{H}}\boldsymbol{\varphi}(t)\right]^*. \tag{24}$$

Note that in this case the preestimation error can be expressed in the form

$$\boldsymbol{\theta}^{\dagger}(t) - \boldsymbol{\theta}(t) = \left[\mathbf{I}_n - \frac{\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{H}}(t)}{\sigma_u^2}\right]\left[\widehat{\boldsymbol{\theta}}^{\mathrm{fRLBF}}(t) - \boldsymbol{\theta}(t)\right] + \frac{1}{\sigma_u^2}\boldsymbol{\varphi}(t)e^*(t) = \mathbf{z}_1'(t) + \mathbf{z}_2(t). \tag{25}$$

Since the estimation error of fRLBF estimates (noncausal) is much smaller than that of LMS estimates (causal), the $\mathbf{z}_1$ component of the preestimation error (25) is much smaller than the analogous component $\mathbf{z}_1'$ in (3), while the $\mathbf{z}_2$ components are identical. Once the enhanced preestimates are evaluated according to (24), they can be smoothed using the same procedure which was described in Section 4. Such a two-stage procedure provides usually a noticeable improvement of the estimation accuracy, especially for lower values of SNR.

6. ADAPTIVE REGULARIZATION

In order to use the optimal regularization formula (22), one needs to know the correlation profile of the process $\{\boldsymbol{\theta}(t)\}$. When the UWA system is fixed in the position, such a statistic can be determined experimentally by averaging identification results obtained in many trials. However, even in this simple case, the correlation matrix $\mathbf{Q} = \mathrm{E}[\boldsymbol{\theta}(t)\boldsymbol{\theta}^{\mathrm{H}}(t)]$ is likely to depend on environmental factors such as the water temperature and weather conditions. Transmitter/receiver motion makes the picture even more complicated, Stojanovic & Preisig [2009]. Therefore, to make the system more robust, at each time instant t the cancellation unit may be allowed to choose the best fitting variant amongst a certain number of the available correlation profiles. As a selection rule, one can use the leave-one-out cross-validation approach. In this framework, the degree of fit of the model is defined as the local sum of squared unbiased interpolation errors (deleted residuals)

$$\varepsilon_0(t|\mathbf{R}) = y(t) - [\widehat{\boldsymbol{\theta}}_0^{\mathrm{fRLBF}}(t|\mathbf{R})]^{\mathrm{H}}\boldsymbol{\varphi}(t) \tag{26}$$

where $\widehat{\boldsymbol{\theta}}_0^{\mathrm{fRLBF}}(t|\mathbf{R})$ denotes the holey estimate of $\boldsymbol{\theta}(t)$, obtained by excluding from the estimation process, governed by (15), the "central" measurement $\widetilde{\boldsymbol{\theta}}(t)$

$$\widehat{\boldsymbol{\alpha}}_0^{\mathrm{fRLBF}}(t|\mathbf{R}) = \arg\min_{\boldsymbol{\alpha}}\left\{\sum_{\substack{i=-k\\ i\neq 0}}^{k} w(i)||\widetilde{\boldsymbol{\theta}}(t+i) - \mathbf{F}(i)\boldsymbol{\alpha}||^2 + ||\boldsymbol{\alpha}||^2_{\mathbf{F}_0^{\mathrm{T}}\mathbf{R}\mathbf{F}_0}\right\}$$

$$\widehat{\boldsymbol{\theta}}_0^{\mathrm{fRLBF}}(t|\mathbf{R}) = \mathbf{F}_0\widehat{\boldsymbol{\alpha}}_0^{\mathrm{fRLBF}}(t|\mathbf{R}) \tag{27}$$

Since $\mathbf{R}$ is a positive definite matrix, it can be expressed in the form $\mathbf{R} = \mathbf{V}\boldsymbol{\Lambda}_n\mathbf{V}^{\mathrm{H}}$, where $\boldsymbol{\Lambda}_n = \mathrm{diag}\{\lambda_1,\ldots,\lambda_n\}$ is a diagonal matrix made up of the eigenvalues of $\mathbf{R}$, and $\mathbf{V}$, $\mathbf{V}^{\mathrm{H}}\mathbf{V} = \mathbf{V}\mathbf{V}^{\mathrm{H}} = \mathbf{I}_n$, is an orthonormal matrix made up of its normalized eigenvectors. Using this decomposition, one can show that (see Appendix 2)

$$\widehat{\boldsymbol{\theta}}_0^{\mathrm{fRLBF}}(t|\mathbf{R}) = \mathbf{V}\boldsymbol{\Gamma}_n\mathbf{V}^{\mathrm{H}}[\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t|\mathbf{R}) - \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\widetilde{\boldsymbol{\theta}}(t)] \tag{28}$$

where

$$\boldsymbol{\Gamma}_n = \mathrm{diag}\left\{\frac{1}{1+(\lambda_1-1)\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0},\ldots,\frac{1}{1+(\lambda_n-1)\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}\right\}. \tag{29}$$

According to (28), the holey fRLBF estimates can be easily obtained by postprocessing the fLBF estimates.

Consider now the case where several fRLBF algorithms, equipped with different regularization matrices $\mathbf{R} \in \mathcal{R} = \{\mathbf{R}_1,\ldots,\mathbf{R}_M\}$, are run simultaneously yielding interpolation errors $\varepsilon_0(t|\mathbf{R}_i), i = 1,\ldots,M$. Selection of the best-fitting value of $\mathbf{R}$ can be made using the following cross-validation decision rule

$$\mathbf{R}_{\mathrm{opt}}(t) = \arg\min_{\mathbf{R}\in\mathcal{R}}\sum_{i=-L}^{L}|\varepsilon_0(t+i|\mathbf{R})|^2 \tag{30}$$

where L determines the size of the local decision window. The same decision rule can be used to select m and k.

7. COMPUTER SIMULATIONS

Simulation was carried out for the model of the self-interference channel of the full-duplex UWA system, decribed in Shen et al. [2020a]. Following Shen et al. [2020a], it was assumed that all complex-valued analog signals are sampled at the rate of 1 kHz, and that the bandwidth of channel coefficient variation is 1 Hz, which can be regarded as fast changes in the UWA case. The channel was modeled as a 50-tap FIR filter with complex-valued coefficients that vary independently of each other. The time-varying impulse response coefficients were generated by lowpass filtering of discrete time circular (with independent real and imaginary components) white Gaussian noise with the variance chosen according to

$$\mathrm{var}[\theta_j(t)] = \zeta^{j-1}, \quad j = 1,\ldots,50$$

which reflects the decaying power delay profile caused by the spreading and absorption loss. The value of ζ was set to 0.69 so that the ratio between the variance of the first arrivals ($j=1$) and that of the latest arrivals ($j=50$) was equal to 80 dB, Shen et al. [2020a]. Typical trajectories of system parameters are shown in Fig. 1.

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

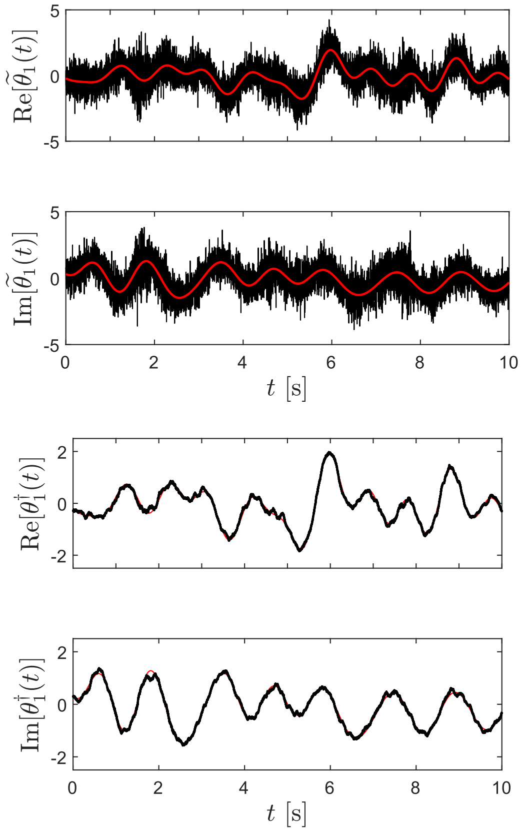

Fig. 1. Typical trajectories of system parameters (red lines) and their preestimates (black lines) obtained for SNR=20 dB. Top figure – LMS-based preestimates, bottom figure – enhanced preestimates.

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

equal to 30 dB, 20 dB and 10 dB, respectively.

The estimation design parameters were set to $k = 100$, $w(i) \equiv 1$ (rectangular window), $m = 3$ (Legendre basis), and $\mu = 0.03$. Based on the available prior knowledge of the estimated impulse response (exponentially decaying and spatially uncorrelated), the regularization matrix was adopted in the form (note that in our case $N_k \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0 = 1$)

$$\mathbf{R}(t) = \widehat{\sigma}_z^2(t)\ \mathrm{diag}\{1, \gamma^{-1}, \gamma^{-2}, \ldots, \gamma^{-49}\}.$$

Three hypothetical values of γ were considered: 0.5, 0.7 and 0.9, none of which was equal to ζ. Optimization was carried out numerically using (30) by searching, at each time instant t, for the best value of $\gamma \in \{0.5, 0.7, 0.9\}$.

Performance was evaluated in terms of the self-interference cancellation factor (SICF) proposed in Shen et al. [2020a]

$$\mathrm{SICF} = \frac{\sum_t |\boldsymbol{\theta}^{\mathrm{H}}(t)\boldsymbol{\varphi}(t)|^2}{\sum_t |[\boldsymbol{\theta}(t) - \widehat{\boldsymbol{\theta}}(t)]^{\mathrm{H}}\boldsymbol{\varphi}(t)|^2} \tag{31}$$

and in terms of the following normalized root mean squared error measure of fit used in Ljung & Chen [2013]

$$\mathrm{FIT}(t) = 100\left(1 - \left[\frac{\sum_{j=1}^{50}|\theta_j(t) - \widehat{\theta}_j(t)|^2}{\sum_{j=1}^{50}|\theta_j(t) - \bar{\theta}(t)|^2}\right]^{1/2}\right) \tag{32}$$

Table 1. FIT[%]/SICF[dB] scores obtained for 3 signal-to-noise ratios for the algorithms described in the text.

Alg. \SNR	30 dB	20 dB	10 dB
LBF	96.0/32.2	87.2/22.2	59.7/12.2
fLBF	80.4/15.8	75.1/13.9	46.0/ 7.6
fRLBF$_1$	80.8/15.2	78.3/14.1	66.5/10.2
fRLBF$_2$	88.5/19.4	85.9/17.8	73.8/12.6
fRLBF$_3$	82.1/16.6	77.7/14.9	58.3/ 9.6
A	87.0/18.4	83.8/16.6	69.3/11.3
fRLBF$_1^\dagger$	89.6/20.4	87.3/18.7	76.3/13.4
fRLBF$_2^\dagger$	89.8/20.7	87.4/18.9	75.8/13.4
fRLBF$_3^\dagger$	89.7/20.8	87.2/18.9	75.3/13.4
A$^+$	89.7/20.7	87.2/18.9	75.6/13.4

where $\bar{\theta}(t) = \frac{1}{50}\sum_{j=1}^{50}\theta_j(t)$. The maximum value of FIT(t), equal to 100, corresponds to the perfect match between the true and estimated impulse response. The final scores, further referred to as FIT (%) and SICF (dB), were obtained by combined time averaging (10000 time steps) and ensemble averaging (20 realizations of scaling coefficients) of the instantaneous/realization-constrained measures. To enable the LMS algorithm reach its steady state behavior, 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 simulation time.

Table 1 compares results obtained for the LBF algorithm, fLBF algorithm, three fRLBF algorithms operating on regular preestimates, with fixed values of γ: fRLBF$_1$ ($\gamma = 0.5$), fRLBF$_2$ ($\gamma = 0.7$) and fRLBF$_3$ ($\gamma = 0.9$), three fRLBF algorithms operating on enhanced preestimates, with fixed values of γ: fRLBF$_1^\dagger$ ($\gamma = 0.5$), fRLBF$_2^\dagger$ ($\gamma = 0.7$) and fRLBF$_3^\dagger$ ($\gamma = 0.9$), and two algorithms with adaptive scheduling of γ (A, A$^+$) with L set to 30.

According to the results summarized in Table 1, regularization improves channel identification results (in spite of the discrepancy between the true value of γ and the assumed one). Furthermore, adaptive scheduling of γ yields performance comparable with that given by the best algorithms incorporated in the parallel estimation scheme.

8. CONCLUSION

A new method of identification of time-varying linear systems, based on the concepts of preestimation and regularization, was proposed and applied to identification of underwater acoustic channels. The new approach allows one to achieve considerable performance gains at very moderate computational costs.

REFERENCES

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

Guo, L. & Ljung, L. (1995). Performance analysis of general tracking algorithms. *IEEE Trans. Automat. Contr.*, (40), 1388–1402.

Ljung, L. & Chen, T. (2013). 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, 1-8.

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

Niedźwiecki, M. & Kłaput, T. (2002). Fast recursive basis function estimators for identification of nonstationary systems. *IEEE Transactions on Signal Processing*, (50), 1925–1934.

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

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

Niedźwiecki, M., Gańcza A. & Ciołek, M. (2020b). On the preestimation technique and its application to identification of nonstationary systems. *59th IEEE Conf. on Decision and Control*, Jeju Island, South Korea, 286-293.

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

Shen, L., Henson, B., Zakharov, Y., Morozs, N. & Mitchell, P. (2020a) Adaptive filtering for full-duplex UWA systems with time-varying self-interference channel. *IEEE Access*, (8), 187590 - 187604.

Shen, L., Henson, B., Zakharov, Y., & Mitchell, P. (2020b). Digital self-interference cancellation for underwater acoustic systems. *IEEE Transactions on Circuits and Systems II: Express Briefs*, (67), 192-196.

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.

Tsatsanis, M. K. & Giannakis, G. B. (1996). Modelling and equalization of rapidly fading channels. *International Journal of Adaptive Control and Signal Processing*, (310), 515–522.

APPENDIX 1 [derivation of (16)]

Denote by $\mathbf{B} = \mathbf{D}\mathbf{F}_0 = \mathbf{D} \otimes \mathbf{f}_0^{\mathrm{T}}$ the $n \times mn$ matrix. It is easy to show that

$$\widehat{\boldsymbol{\alpha}}^{\mathrm{fRLBF}}(t|\mathbf{R}) = [\mathbf{I}_{mn} + \mathbf{B}^{\mathrm{H}}\mathbf{B}]^{-1}\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t)$$

where $\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t)$ is given by (11). 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

$$[\mathbf{I}_{mn} + \mathbf{B}^{\mathrm{H}}\mathbf{B}]^{-1} = \mathbf{I}_{mn} - \mathbf{B}^{\mathrm{H}}[\mathbf{I}_n + \mathbf{B}\mathbf{B}^{\mathrm{H}}]^{-1}\mathbf{B}$$

Note that

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

Hence

$$\begin{aligned}
&[\mathbf{I}_{mn} + \mathbf{B}^{\mathrm{H}}\mathbf{B}]^{-1} \\
&\quad = \mathbf{I}_{mn} - [\mathbf{D}^{\mathrm{H}} \otimes \mathbf{f}_0][\mathbf{I}_n + \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{D}\mathbf{D}^{\mathrm{H}}]^{-1}[\mathbf{D} \otimes \mathbf{f}_0^{\mathrm{T}}] \\
&\quad = \mathbf{I}_{mn} - \big\{\mathbf{D}^{\mathrm{H}}[\mathbf{I}_n + \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{D}\mathbf{D}^{\mathrm{H}}]^{-1}\mathbf{D}\big\} \otimes [\mathbf{f}_0\mathbf{f}_0^{\mathrm{T}}].
\end{aligned}$$

Observe that

$$\begin{aligned}
&[\mathbf{I}_n + \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{D}^{\mathrm{H}}\mathbf{D}]^{-1} \\
&\quad = \mathbf{I}_n - \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{D}^{\mathrm{H}}[\mathbf{I}_n + \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{D}\mathbf{D}^{\mathrm{H}}]^{-1}\mathbf{D}
\end{aligned}$$

Combining the last two results, and noting that $\mathbf{D}^{\mathrm{H}}\mathbf{D} = \mathbf{R}$, one arrives at

$$\begin{aligned}
[\mathbf{I}_{mn} + \mathbf{B}^{\mathrm{H}}\mathbf{B}]^{-1} &= \mathbf{I}_{mn} - \mathbf{I}_n \otimes \left[\frac{\mathbf{f}_0\mathbf{f}_0^{\mathrm{T}}}{\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}\right] \\
&\quad + [\mathbf{I}_n + \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{R}]^{-1} \otimes \left[\frac{\mathbf{f}_0\mathbf{f}_0^{\mathrm{T}}}{\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}\right]
\end{aligned} \tag{33}$$

and

$$\begin{aligned}
\widehat{\boldsymbol{\alpha}}^{\mathrm{fRLBF}}(t|\mathbf{R}) &= \widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t) - \widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t) \otimes \left[\frac{\mathbf{f}_0}{\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}\right] \\
&\quad + \big\{[\mathbf{I}_n + \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{R}]^{-1}\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t)\big\} \otimes \left[\frac{\mathbf{f}_0}{\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}\right].
\end{aligned}$$

leading to

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}^{\mathrm{fRLBF}}(t|\mathbf{R}) &= (\mathbf{I}_n \otimes \mathbf{f}_0^{\mathrm{T}})\widehat{\boldsymbol{\alpha}}^{\mathrm{fRLBF}}(t|\mathbf{R}) \\
&= [\mathbf{I}_n + \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{R}]^{-1}\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t)
\end{aligned}$$

which is nothing but (16).

APPENDIX 2 [derivation of (28)]

Note that the estimate $\widehat{\boldsymbol{\theta}}_0^{\mathrm{fRLBF}}(t|\mathbf{R})$ can be expressed in the form

$$\widehat{\boldsymbol{\theta}}_0^{\mathrm{fRLBF}}(t|\mathbf{R}) = \mathbf{F}_0[\mathbf{S} - \mathbf{F}_0^{\mathrm{T}}\mathbf{F}_0]^{-1}[\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t) - \mathbf{F}_0\widetilde{\boldsymbol{\theta}}(t)] \tag{34}$$

where $\mathbf{S} = \mathbf{I}_{mn} + \mathbf{B}^{\mathrm{H}}\mathbf{B}$.

Using the Woodbury identity, one arrives at

$$[\mathbf{S} - \mathbf{F}_0^{\mathrm{T}}\mathbf{F}_0]^{-1} = \mathbf{S}^{-1} + \mathbf{S}^{-1}\mathbf{F}_0^{\mathrm{T}}[\mathbf{I}_n - \mathbf{F}_0\mathbf{S}^{-1}\mathbf{F}_0^{\mathrm{T}}]^{-1}\mathbf{F}_0\mathbf{S}^{-1}$$

According to (16), it holds that

$$\mathbf{F}_0\mathbf{S}^{-1}\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t) = \widehat{\boldsymbol{\alpha}}^{\mathrm{fRLBF}}(t|\mathbf{R}) = \mathbf{G}^{-1}\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t)$$

where $\mathbf{G} = [\mathbf{I}_n + \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{R}]$. Furthermore, using (34), one obtains

$$\begin{aligned}
&\mathbf{F}_0\mathbf{S}^{-1}\mathbf{F}_0^{\mathrm{T}} = (\mathbf{I}_n \otimes \mathbf{f}_0^{\mathrm{T}})(\mathbf{I}_n \otimes \mathbf{f}_0) \\
&- (\mathbf{I}_n \otimes \mathbf{f}_0^{\mathrm{T}})\left(\mathbf{I}_n \otimes \left[\frac{\mathbf{f}_0\mathbf{f}_0^{\mathrm{T}}}{\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}\right]\right)(\mathbf{I}_n \otimes \mathbf{f}_0) \\
&+ (\mathbf{I}_n \otimes \mathbf{f}_0^{\mathrm{T}})\left(\mathbf{G}^{-1} \otimes \left[\frac{\mathbf{f}_0\mathbf{f}_0^{\mathrm{T}}}{\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}\right]\right)(\mathbf{I}_n \otimes \mathbf{f}_0) = \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{G}^{-1}
\end{aligned}$$

where the second transition follows from the fact that the first two terms cancel out. Combining the last three relationships, one arrives at

$$\begin{aligned}
&\mathbf{F}_0[\mathbf{S} - \mathbf{F}_0^{\mathrm{T}}\mathbf{F}_0]^{-1}\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t) = \\
&\mathbf{G}^{-1}\Big\{\mathbf{I}_n + \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0[\mathbf{I}_n - \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{G}^{-1}]^{-1}\mathbf{G}^{-1}\Big\}\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t).
\end{aligned}$$

Note that

$$\mathbf{G}^{-1} = \mathbf{V}\ \mathrm{diag}\left\{\frac{1}{1+\lambda_1\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}, \ldots, \frac{1}{1+\lambda_n\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}\right\}\mathbf{V}^{\mathrm{H}}$$

In a similar way, one can show that

$$\begin{aligned}
&[\mathbf{I}_n - \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{G}^{-1}]^{-1} \\
&= \mathbf{V}\ \mathrm{diag}\left\{\frac{1+\lambda_1\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}{1+(\lambda_1-1)\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}, \ldots, \frac{1+\lambda_n\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}{1+(\lambda_n-1)\mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0}\right\}\mathbf{V}^{\mathrm{H}}
\end{aligned}$$

Exploiting the fact that $\mathbf{V}^{\mathrm{H}}\mathbf{V} = \mathbf{I}_n$, one arrives at

$$\mathbf{F}_0[\mathbf{S} - \mathbf{F}_0^{\mathrm{T}}\mathbf{F}_0]^{-1}\widehat{\boldsymbol{\alpha}}^{\mathrm{fLBF}}(t) = \mathbf{V}\boldsymbol{\Gamma}_n\mathbf{V}^{\mathrm{H}}\widehat{\boldsymbol{\theta}}^{\mathrm{fLBF}}(t)$$

where $\boldsymbol{\Gamma}_n$ is given by (29). Similarly, one can show that

$$\mathbf{F}_0[\mathbf{S} - \mathbf{F}_0^{\mathrm{T}}\mathbf{F}_0]^{-1}\mathbf{F}_0\widetilde{\boldsymbol{\theta}}(t) = \mathbf{f}_0^{\mathrm{T}}\mathbf{f}_0\mathbf{V}\boldsymbol{\Gamma}_n\mathbf{V}^{\mathrm{H}}\widetilde{\boldsymbol{\theta}}(t)$$

Finally, combining (34) with the last two relationships, one obtains (28).