IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 25, 2021, accepted May 11, 2021, date of publication May 19, 2021, date of current version May 26, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3081561

Decoupled Kalman Filter Based Identification of Time-Varying FIR Systems

MARCIN CIOŁEK, (Member, IEEE), MACIEJ NIEDŹWIECKI, (Senior Member, IEEE), AND ARTUR GAŃCZA, (Member, IEEE)

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

Corresponding author: Marcin Ciołek (marciole@pg.edu.pl)

This work was supported in part by the National Science Center under Grant UMO-2018/29/B/ST7/00325.

ABSTRACT When system parameters vary at a fast rate, identification schemes based on model-free local estimation approaches do not yield satisfactory results. In cases like this, more sophisticated parameter tracking procedures must be used, based on explicit models of parameter variation (often referred to as hypermodels), either deterministic or stochastic. Kalman filter trackers, which belong to the second category, are seldom used in practice due to difficulties in adjusting their internal parameters such as the smoothness coefficient and the order of the hypermodel. The paper presents a new solution to this problem, based on the concept of preestimation of system parameters. The resulting identification algorithms, which can be characterized as decoupled Kalman trackers, are computationally attractive, easy to tune and can be optimized in an adaptive fashion using the parallel estimation approach. The decoupled KF algorithms can be regarded as an attractive alternative to the state-of-the-art algorithms which are much more computationally demanding.

INDEX TERMS Kalman filter, parallel estimation, preestimation of system parameters, system identification.

I. INTRODUCTION

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

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

where $t = \ldots, -1, 0, 1, \ldots$ denotes discrete (normalized, i.e., dimensionless) time, $y(t)$ denotes system output, $\boldsymbol{\varphi}(t) = [u(t-1), \ldots, u(t-n)]^T$ is the regression vector made up of past measurements of the observable (locally) wide sense stationary input signal $\{u(t)\}$, $\{e(t)\}$ denotes white measurement noise, and $\boldsymbol{\theta}(t) = [\theta_1(t), \ldots, \theta_n(t)]^T$ is the parameter vector made up of unknown time-varying system coefficients, independent of $\{u(t)\}$ and $\{e(t)\}$.

Linear time-varying FIR models are used, among others, to describe rapidly fading mobile communication channels. Their identification allows one to efficiently solve the channel equalization (inverse filtering) problem [1], [2] or to mitigate self-interference in full-duplex communication systems [3], [4]. The FIR structure describes well the so-called multi-path effect: due to scattering the transmitted signal reaches the receiver along different paths, i.e., with different time delays; the values of FIR coefficients depend on the strength of ''natural reflectors'' and their time variation is caused by the receiver motion [1], [5].

When system parameters vary slowly with time, their estimation can be successfully carried out using the local estimation approach, such as the method of weighted least squares (WLS) [6], [7]. Since local methods fail in the presence of fast parameter changes, in cases like this more sophisticated frameworks must be used, such as those incorporating explicit models (hypermodels) of parameter variation. Hypermodels can be deterministic or stochastic. In the first case, parameter trajectory is modeled as a linear combination of known functions of time, called basis functions (BF), such as powers of time (polynomial basis) or sine and cosine functions (harmonic basis). The BF approach can be traced back to the paper of Subba Rao [8], which was followed by a large number of contributions exploring different aspects of the BF-based estimation [1]–[4], [7], [9]–[23]. This research covered different types of system/signal models, such as FIR [1]–[4], [22], [23], AR/ARX (autoregressive/autoregressive with exogenous inputs) [8], [18]–[21] and

The associate editor coordinating the review of this manuscript and approving it for publication was Claudia Raibulet.

ARMA (autoregressive moving average) [12], [16], [19], and their various applications in telecommunications [1]–[4], biomedical signal analysis [18], [20], speech analysis [11] and diagnostics of mechanical systems [19]. It involved different functional bases, such as polynomial [3], [4], [8]–[11], [15], [16], [21], [22], harmonic [1], [2], prolate spheroidal [12] and wavelets [17], [20]. For FIR systems and free choice of basis functions some pretty general analytical results, describing both static and dynamic parameter tracking characteristics of BF algorithms, were presented in [7], [14], [15], [22] and [23].

In the case of stochastic hypermodels it is assumed that system parameters change in a random way, e.g. that their evolution can be described by a generalized random walk equation. Then the problem of estimation of $\boldsymbol{\theta}(t)$ can be expressed and solved as the problem of estimation of a state vector of an associated dynamical system. In such a setup parameter estimation can be carried out using appropriately designed Kalman filtering (KF) algorithms [24]–[31]. The available research results are focused on identification of ARX [24], [25] and FIR [31] systems, and AR signals [28]–[30], and on different approaches to tuning of KF trackers, such as the smoothness priors technique [27]–[29] or cross-validation [31]. Some interesting applications of KF-based identification algorithms can be found in [26], [29] and [30].

In spite of qualitative similarities, pointed out in Section II, the KF approach is less frequently used than the BF approach, mainly because of problems with tuning its design parameters. In the current contribution we propose decoupled versions of the KF-based identification algorithms, free of the drawback mentioned above. Decoupling means that identification is carried out independently for each coefficient of the analyzed system. Such a solution is possible owing to a new estimation paradigm for identification of nonstationary stochastic systems, based on the concept of preestimation. Preestimates are very noisy but unbiased estimates of parameter trajectories. They can be used to "X-ray" the structure of system parameter variation without making any assumptions about its functional form or speed. Due to large variability, preestimates must be postprocessed. We will use for this purpose Kalman filters/smoothers.

It will be shown that, unlike in the case of the classical KF-based parameter trackers, the estimation memory span of the proposed algorithm is data-independent and hence it can be easily controlled by the user. It will be also shown that optimization of the tracking performance of the decoupled KF algorithm, i.e., tuning of its design parameters to the unknown and possibly time-varying form and rate of parameter changes, can be achieved by means of parallel estimation and cross-validation. Finally, we will derive the simplified, steady state version of the proposed identification algorithm with reduced computational load (depending linearly on the number of estimated parameters), which is a computationally attractive alternative to the current state-of-the-art.

II. CLASSICAL KALMAN FILTER BASED IDENTIFICATION ALGORITHMS

The integrated random walk (IRW) model of parameter variation is the most frequently used stochastic hypermodel [7]. The IRW model [25], [28] of order m is governed by

$$\nabla^m \boldsymbol{\theta}(t) = \boldsymbol{w}(t) \tag{2}$$

where $\nabla^m \boldsymbol{\theta}(t)$ denotes the m-th order difference of $\boldsymbol{\theta}(t)$: $\nabla^m \boldsymbol{\theta}(t) = (1-q^{-1})^m \boldsymbol{\theta}(t) = \sum_{i=0}^{m} f_i \boldsymbol{\theta}(t-i), f_i = (-1)^i \binom{m}{i}$, $i = 0, \ldots, m$ (q^{-1} is the backward shift operator), and $\{\boldsymbol{w}(t)\}$ denotes a zero-mean i.i.d. sequence, independent of $\{e(t)\}$ and $\{\boldsymbol{\varphi}(t)\}$. To reduce the number of degrees of freedom it is usually assumed that $\mathrm{cov}[\boldsymbol{w}(t)] = \sigma_w^2 \boldsymbol{I}$, i.e., that the rate of variation is the same for all system coefficients. Generally, the larger the order m of the IRW model, the smoother the corresponding parameter trajectory.

Equations (1) and (2) can be put in the state space form [25], [28]

$$\begin{aligned} \boldsymbol{x}(t) &= \boldsymbol{F}_m \boldsymbol{x}(t-1) + \boldsymbol{C}_m \boldsymbol{w}(t) \\ y(t) &= \boldsymbol{\varphi}^T(t) \boldsymbol{C}_m^T \boldsymbol{x}(t) + e(t) \end{aligned} \tag{3}$$

where $\boldsymbol{x}(t) = [\boldsymbol{\theta}^T(t), \boldsymbol{\theta}^T(t-1), \ldots, \boldsymbol{\theta}^T(t-m+1)]^T$ is the $mn \times 1$ regression vector,

$$\boldsymbol{F}_m = \begin{bmatrix} -f_1 \boldsymbol{I} & -f_2 \boldsymbol{I} & \ldots & -f_{m-1} \boldsymbol{I} & -f_m \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{O} & \ldots & \boldsymbol{O} & \boldsymbol{O} \\ & & & & \\ \boldsymbol{O} & \boldsymbol{O} & \ldots & \boldsymbol{I} & \boldsymbol{O} \end{bmatrix}$$

is the $mn \times mn$ state transition matrix, and $\boldsymbol{C}_m = [\boldsymbol{I}, \boldsymbol{O}, \ldots, \boldsymbol{O}]^T$ is the $mn \times n$ output matrix ($\boldsymbol{I}$ and $\boldsymbol{O}$ denote $n \times n$ identity and null matrices, respectively).

Since $\boldsymbol{\theta}(t) = \boldsymbol{C}_m^T \boldsymbol{x}(t)$, the problem of estimation of $\boldsymbol{\theta}(t)$ can be formulated and solved as the problem of estimation of the state vector $\boldsymbol{x}(t)$ of the dynamical system (3). The optimal, in the mean squared sense, estimate of $\boldsymbol{x}(t)$ based on the available observation history $\Omega(t) = \{y(i), \boldsymbol{\varphi}(i), i \leq t\}$ has the form $\widehat{\boldsymbol{x}}(t|t) = E[\boldsymbol{x}(t)|\Omega(t)]$. Under Gaussian assumptions the conditional mean can be computed recursively using the Kalman filtering algorithm. Whenever noncausal estimation – based on the prerecorded data set $\Omega(N)$, containing both "past" and "future" measurements (relative to t) – is feasible, the minimum-variance estimate has the form $\widehat{\boldsymbol{x}}(t|N) = E[\boldsymbol{x}(t)|\Omega(N)]$ and can be evaluated using the algorithm known as Kalman smoother. The corresponding causal and noncausal estimates of $\boldsymbol{\theta}(t)$ have the form

$$\widehat{\boldsymbol{\theta}}(t|t) = \boldsymbol{C}_m^T \widehat{\boldsymbol{x}}(t|t), \quad \widehat{\boldsymbol{\theta}}(t|N) = \boldsymbol{C}_m^T \widehat{\boldsymbol{x}}(t|N).$$

It is well known [7] that parameter tracking properties of the Kalman filtering/smoothing algorithms based on the hypermodel (2) depend on the variance quotient $\xi = \sigma_w^2/\sigma_e^2$, further referred to as the smoothness coefficient (both Kalman filtering and Kalman smoothing algorithms can be written down in a normalized form which depends on ξ, rather than separately on σ_w^2 and σ_e^2). Different values of ξ

and m correspond to different estimation memory settings of KF algorithms. In practice ξ and m serve as user-dependent "knobs", allowing one to tune KF tracker/smoother to the rate of system nonstationarity, or in statistical terms – to trade off the bias and variance components of the mean squared parameter estimation error. Tuning can be achieved by running several KF algorithms, equipped with different hypermodel order and smoothness coefficients, and choosing at each time instant the estimates yielded by the algorithm which proves to be "locally the best" [32].

It is straightforward to show that when $\boldsymbol{w}(t)$ in (2) is set to zero, i.e., when $\nabla^m \boldsymbol{\theta}(t) = 0$, system parameters must obey

$$\theta_j(t) = \sum_{l=1}^{m} a_{jl} t^{l-1}, \quad j = 1, \ldots, n. \tag{4}$$

Hence, the IRW model can be regarded as a local, or perturbed, power series model of parameter variation, which corresponds to the deterministic BF hypermodel incorporating polynomial basis $\{1, t, t^2, \ldots, t^{m-1}\}$. This is an obvious point of tangency of the deterministic BF approach and the stochastic KF one. The difference lies in the way the polynomial description of parameter variation is time-localized. In the deterministic case the memory of the estimation algorithm is controlled by restricting the model to a local interval only. In the stochastic case the same goal is achieved by adding to the polynomial generating function a random perturbation. However, while the estimation memory of the BF algorithm can be easily evaluated and controlled by changing the local approximation range (see Section 7), the dependence of the estimation memory of the KF algorithm on ξ is not known and obscured by the fact that it depends also on the characteristics of the regression vector $\boldsymbol{\varphi}(t)$ [7]. This significantly complicates design of KF algorithms and KF-based parallel estimation schemes mentioned above.

III. PREESTIMATION TECHNIQUE

Preestimation is a technique introduced in [33] and further developed in [34]–[36]. 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 [34], preestimates, further denoted by $\boldsymbol{\theta}^*(t)$, can be obtained by "inverse filtering" short-memory exponentially weighted least squares (EWLS) estimates, namely

$$\boldsymbol{\theta}^*(t) = L_t \widehat{\boldsymbol{\theta}}^{EWLS}(t) - \lambda_0 L_{t-1} \widehat{\boldsymbol{\theta}}^{EWLS}(t-1) \tag{5}$$

where λ_0, $0 < \lambda_0 < 1$, denotes the so-called forgetting constant and $L_t = \sum_{i=0}^{t-1} \lambda_0^i = \lambda_0 L_{t-1} + 1$ denotes the effective width of the exponential window. EWLS estimates can be computed recursively [6]

$$\begin{aligned}
\varepsilon(t) &= y(t) - \widehat{\boldsymbol{\theta}}^T(t-1)\boldsymbol{\varphi}(t) \\
\boldsymbol{g}(t) &= \frac{\boldsymbol{R}(t-1)\boldsymbol{\varphi}(t)}{\lambda_0 + \boldsymbol{\varphi}^T(t)\boldsymbol{R}(t-1)\boldsymbol{\varphi}(t)} \\
\widehat{\boldsymbol{\theta}}(t) &= \widehat{\boldsymbol{\theta}}(t-1) + \boldsymbol{g}(t)\varepsilon(t) \\
\boldsymbol{R}(t) &= \frac{1}{\lambda_0}\left[\boldsymbol{R}(t-1) - \frac{\boldsymbol{R}(t-1)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^T(t)\boldsymbol{R}(t-1)}{\lambda_0 + \boldsymbol{\varphi}^T(t)\boldsymbol{R}(t-1)\boldsymbol{\varphi}(t)}\right].
\end{aligned}$$

For large values of t the effective window width reaches its steady state value equal to $L_\infty = 1/(1-\lambda_0)$. In this case the preestimate (5) can be evaluated using the following simplified formula

$$\boldsymbol{\theta}^*(t) = \frac{1}{1-\lambda_0}[\widehat{\boldsymbol{\theta}}^{EWLS}(t) - \lambda_0 \widehat{\boldsymbol{\theta}}^{EWLS}(t-1)]\,. \tag{6}$$

It can be shown that if the input signal $\{u(t)\}$ is (locally) stationary, and the measurement noise $\{e(t)\}$ is white, the preestimates defined in this way are approximately unbiased [34], namely

$$\boldsymbol{\theta}^*(t) = \boldsymbol{\theta}(t) + z(t) \tag{7}$$

where $z(t)$ denotes (approximately) zero-mean white noise with large covariance matrix $\boldsymbol{\Sigma}_z$.

Fig. 1 shows the preestimated parameter trajectories obtained for a nonstationary two-tap FIR system governed by

$$y(t) = \theta_1(t)u(t-1) + \theta_2(t)u(t-2) + e(t) \tag{8}$$

excited by a zero-mean stationary autoregressive Gaussian process with autocorrelation function $E[u(t)u(t-i)] = (0.8)^i$, and corrupted by white Gaussian noise with variance $\sigma_e^2 = 0.0025$ (SNR = 25 dB). Parameter $\theta_1(t)$ was changing in a chirp-like way, and parameter $\theta_2(t)$ was piecewise constant – see Fig. 1. The forgetting constant λ_0 was set to 0.9.

Note that preestimates provide interesting insights into the structure of parameter variation, and they do so without making any assumptions about the speed and mode of parameter variation. Additionally, such a "prescreening" is provided separately for each system coefficient, which allows one to

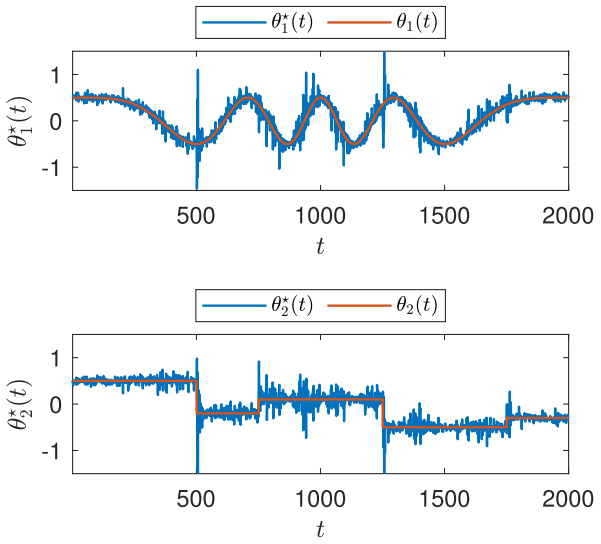

FIGURE 1. Preestimated parameter trajectories (blue lines) of a nonstationary two-tap FIR system, superimposed on the true trajectories (red lines).

individually adjust the postprocessing scheme (as different components of the parameter vector may require using different algorithm settings).

IV. DECOUPLED KALMAN FILTER BASED IDENTIFICATION ALGORITHMS

Consider the j-th component of the parameter vector $\boldsymbol{\theta}(t)$. As shown above the preestimate of $\theta_j(t)$ can be written down in the form

$$\theta_j^*(t) = \theta_j(t) + z_j(t) \tag{9}$$

where $z_j(t)$ denotes zero-mean white "noise" with a large variance σ_z^2 (for a stationary input signal the variance of $z_j(t)$ does not depend on j).

The proposed postprocessing will be based on Kalman filtering/smoothing. Similar to the classical case, we will assume that $\theta_j(t)$ obeys the IRW model of order m

$$\nabla^m \theta_j(t) = w_j(t) \tag{10}$$

where $w_j(t)$ denotes white perturbation noise with variance σ_w^2.

Rewriting (9) and (10) in the state space form, one obtains

$$\begin{aligned} \boldsymbol{x}_j(t) &= \boldsymbol{F}_m \boldsymbol{x}_j(t-1) + \boldsymbol{c}_m w_j(t) \\ \theta_j^*(t) &= \boldsymbol{c}_m^T \boldsymbol{x}_j(t) + z_j(t) \end{aligned} \tag{11}$$

where $\boldsymbol{x}_j(t) = [\theta_j(t), \theta_j(t-1), \ldots, \theta_j(t-m+1)]^T$ and $\boldsymbol{c}_m = [1, 0, \ldots, 0]^T$ are $m \times 1$ vectors and

$$\boldsymbol{F}_m = \begin{bmatrix} -f_1 & -f_2 & \ldots & -f_{m-1} & -f_m \\ 1 & 0 & \ldots & 0 & 0 \\ & & & & \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}$$

is a $m \times m$ matrix.

Let $\pi = \{m, \xi\}$, where $\xi = \sigma_w^2/\sigma_z^2$ is a smoothness coefficient, which can be used for tuning purposes. Equations of the normalized version of Kalman filter, which constitute the causal identification algorithm, have the form

$$\begin{aligned} \widehat{\boldsymbol{x}}_{j|\pi}(t|t-1) &= \boldsymbol{F}_m \widehat{\boldsymbol{x}}_{j|\pi}(t-1|t-1) \\ \boldsymbol{P}_\pi(t|t-1) &= \boldsymbol{F}_m \boldsymbol{P}_\pi(t-1|t-1) \boldsymbol{F}_m^T + \boldsymbol{c}_m \boldsymbol{c}_m^T \xi \\ \varepsilon_{j|\pi}(t) &= \theta_j^*(t) - \boldsymbol{c}_m^T \widehat{\boldsymbol{x}}_{j|\pi}(t|t-1) \\ q_\pi(t) &= \boldsymbol{c}_m^T \boldsymbol{P}_\pi(t|t-1) \boldsymbol{c}_m + 1 \\ \boldsymbol{k}_\pi(t) &= \frac{\boldsymbol{P}_\pi(t|t-1)\boldsymbol{c}_m}{q_\pi(t)} \\ \widehat{\boldsymbol{x}}_{j|\pi}(t|t) &= \widehat{\boldsymbol{x}}_{j|\pi}(t|t-1) + \boldsymbol{k}_\pi(t)\varepsilon_{j|\pi}(t) \\ \boldsymbol{P}_\pi(t|t) &= \boldsymbol{P}_\pi(t|t-1) - \boldsymbol{k}_\pi(t)\boldsymbol{k}_\pi^T(t) q_\pi(t) \\ \widehat{\theta}_{j|\pi}(t|t) &= \boldsymbol{c}_m^T \widehat{\boldsymbol{x}}_{j|\pi}(t|t) \\ t &= 1, 2, \ldots \end{aligned} \tag{12}$$

where $\boldsymbol{P}_\pi(t|t-1) = \mathrm{cov}[\widehat{\boldsymbol{x}}_{j|\pi}(t|t-1)]/\sigma_z^2$ and $\boldsymbol{P}_\pi(t|t) = \mathrm{cov}[\widehat{\boldsymbol{x}}_{j|\pi}(t|t)]/\sigma_z^2$ denote normalized *a priori* and *a posteriori* covariance matrices, respectively. Note that for fixed values of m and ξ, the quantities $\boldsymbol{P}_\pi(t|t-1)$, $\boldsymbol{P}_\pi(t|t)$, $\boldsymbol{k}_\pi(t)$ and $q_\pi(t)$ can be computed once for all values of $j = 1, \ldots, n$.

The estimation memory of the KF tracker can be quantified in terms of the noise reduction rate observed when $\theta_j(t) = \theta_j^0 = \mathrm{const}$, i.e., $\theta_j^*(t) = \theta_j^0 + z_j(t)$. Its steady state value is given by

$$N_\infty = \lim_{t\to\infty} \frac{\sigma_{z_j}^2}{\mathrm{var}[\widehat{\theta}_j(t|t)]} \tag{13}$$

and can be easily evaluated numerically for different values of ξ and m. The quantity N_∞ can be interpreted as the width (number of taps) of the averaging filter which in the constant excitation case provides the same noise reduction rate as the KF algorithm. Unlike the classical estimation case, the noise-equivalent width N_∞ *does not* depend on the characteristics of the regression vector $\boldsymbol{\varphi}(t)$.

For $m = 1$ the following expressions

$$N_\infty \cong \frac{2}{\sqrt{\xi}}, \quad \xi \cong \left(\frac{2}{N_\infty}\right)^2 \tag{14}$$

can be easily derived by means of analyzing equations of the steady state Kalman filter – see Appendix A.

For $m > 1$ such calculations become cumbersome, but there are some analytical clues suggesting a possible inverse dependence of N_∞ on $\sqrt[2m]{\xi}$. For $m = 1, 2, 3$ the empirical formulas which provide a pretty good approximation of $N_\infty(\xi)$ and $\xi(N_\infty)$ have the form

$$N_\infty \cong \frac{2 - 0.1(m-1)}{m\sqrt[2m]{\xi}} \tag{15}$$

$$\xi \cong \left[\frac{2 - 0.1(m-1)}{mN_\infty}\right]^{2m}. \tag{16}$$

When both "past" and "future" measurements are available, as in the channel identification case [1], more accurate parameter estimates can be obtained using the fixed interval Kalman smoother. Smoothing is achieved by means of backward-time processing of the KF estimates, which in the so-called Bryson-Frazier realization [7], [31] takes the form

$$\begin{aligned} \boldsymbol{B}_\pi(t) &= \boldsymbol{F}_m[\boldsymbol{I} - \boldsymbol{k}_\pi(t)\boldsymbol{c}_m^T] \\ \boldsymbol{r}_\pi(t-1) &= \boldsymbol{B}_\pi^T(t)\boldsymbol{r}_\pi(t) + \boldsymbol{c}_m \frac{\varepsilon_{j|\pi}(t)}{q_\pi(t)} \\ \boldsymbol{R}_\pi(t-1) &= \boldsymbol{B}_\pi^T(t)\boldsymbol{R}_\pi(t)\boldsymbol{B}_\pi(t) + \frac{\boldsymbol{c}_m \boldsymbol{c}_m^T}{q_\pi(t)} \\ \widehat{\boldsymbol{x}}_{j|\pi}(t|N) &= \widehat{\boldsymbol{x}}_{j|\pi}(t|t-1) + \boldsymbol{P}_\pi(t|t-1)\boldsymbol{r}_\pi(t-1) \\ \boldsymbol{P}_\pi(t|N) &= \boldsymbol{P}_\pi(t|t-1) - \boldsymbol{P}_\pi(t|t-1)\boldsymbol{R}_\pi(t-1)\boldsymbol{P}_\pi(t|t-1) \\ \widehat{\theta}_{j|\pi}(t|N) &= \boldsymbol{c}_m^T \widehat{\boldsymbol{x}}_{j|\pi}(t|N) \\ t &= N-1, \ldots, 1 \end{aligned} \tag{17}$$

with initial conditions $\boldsymbol{r}_\pi(N) = 0$ and $\boldsymbol{R}_\pi(N) = \mathbf{O}$. The steady state value of the estimation memory span is in this case equal to $M_\infty = \lim_{t\to\infty} \sigma_{z_j}^2/\mathrm{var}[\widehat{\theta}_j(t|2t)] = 2\,mN_\infty$. The values of the smoothness coefficient ξ corresponding to the selected values of the estimation memory spans N_∞ and M_∞, and to the order of the IRW model m, are listed in Table 1.

TABLE 1. The values of the smoothness coefficient ξ corresponding to the selected values of the estimation memory spans N_∞ (upper table) and M_∞ (lower table), and to the order of the IRW model m.

$N_\infty \setminus m$	1	2	3
20	1.00E-02	5.09E-06	7.29E-10
40	2.50E-03	3.18E-07	1.14E-11
80	6.25E-04	1.99E-08	1.78E-13
160	1.56E-04	1.24E-09	2.78E-15

$M_\infty \setminus m$	1	2	3
20	4.00E-02	1.30E-03	6.40E-05
40	1.00E-02	8.15E-05	3.97E-07
80	2.50E-03	5.09E-06	9.67E-09
160	6.25E-04	3.18E-07	1.20E-10

V. SIMPLIFIED ALGORITHMS

To further reduce computational load of the KF tracker, the algorithm (12) can be replaced by its steady state version. Denote by $\boldsymbol{P}_\pi^\infty$ and $\boldsymbol{k}_\pi^\infty$ the steady state values of the *a priori* covariance matrix and Kalman gain, respectively

$$\boldsymbol{P}_\pi^\infty = \lim_{t\to\infty} \boldsymbol{P}_\pi(t|t-1), \quad \boldsymbol{k}_\pi^\infty = \frac{\boldsymbol{P}_\pi^\infty \boldsymbol{c}_m}{1+\boldsymbol{c}_m^T \boldsymbol{P}_\pi^\infty \boldsymbol{c}_m}.$$

The steady state version of (12) takes the form

$$\begin{aligned}
\widehat{\boldsymbol{x}}_{j|\pi}(t|t-1) &= \boldsymbol{F}_m \widehat{\boldsymbol{x}}_{j|\pi}(t-1|t-1) \\
\varepsilon_{j|\pi}(t) &= \theta_j^*(t) - \boldsymbol{c}_m^T \widehat{\boldsymbol{x}}_{j|\pi}(t|t-1) \\
\widehat{\boldsymbol{x}}_{j|\pi}(t|t) &= \widehat{\boldsymbol{x}}_{j|\pi}(t|t-1) + \boldsymbol{k}_\pi^\infty \varepsilon_{j|\pi}(t) \\
\widehat{\theta}_{j|\pi}(t|t) &= \boldsymbol{c}_m^T \widehat{\boldsymbol{x}}_{j|\pi}(t|t) \\
t &= 1, 2, \ldots
\end{aligned} \tag{18}$$

The steady state values $\boldsymbol{P}_\pi^\infty$ and $\boldsymbol{k}_\pi^\infty$ can be obtained by solving the associated Riccati equation, or by iterating KF equations for a sufficiently long time (in the second case the so-called doubling algorithms can be used to reduce the number of iterations [37]).

Similarly, the smoothing algorithm (17) can be reduced down to

$$\begin{aligned}
\boldsymbol{r}_\pi(t-1) &= [\boldsymbol{B}_\pi^\infty]^T \boldsymbol{r}_\pi(t) + \boldsymbol{c}_m \frac{\varepsilon_{j|\pi}(t)}{q_\pi^\infty} \\
\widehat{\boldsymbol{x}}_{j|\pi}(t|N) &= \widehat{\boldsymbol{x}}_{j|\pi}(t|t-1) + \boldsymbol{P}_\pi^\infty \boldsymbol{r}_\pi(t-1) \\
\widehat{\theta}_{j|\pi}(t|N) &= \boldsymbol{c}_m^T \widehat{\boldsymbol{x}}_{j|\pi}(t|N) \\
t &= N-1, \ldots, 1
\end{aligned} \tag{19}$$

where

$$\boldsymbol{B}_\pi^\infty = \boldsymbol{F}_m[\boldsymbol{I} - \boldsymbol{k}_\pi^\infty \boldsymbol{c}_m^T], \quad q_\pi^\infty = 1 + \boldsymbol{c}_m^T \boldsymbol{P}_\pi^\infty \boldsymbol{c}_m.$$

Exploiting symmetry of all covariance matrices and a special structure of the matrix $\boldsymbol{F}_m$ and the vector $\boldsymbol{c}_m$, the computational cost of running n filtering algorithms (12) can be reduced to $\frac{3}{2}m^2 + \frac{5}{2}m + 2mn$ multiply-add operations per time update. The same count for the smoothing algorithm (17) yields $3m^3 + m^2 + m + (2m^2 + 1)n$ operations per time update. When the matrices/vectors are precomputed, or when their steady state values are used, the computational load is further reduced to $2mn$ (filtering) and $(2m^2+1)n$ (smoothing) operations, respectively. Note that in both cases the number of operations depends linearly on the number of estimated coefficients.

VI. PARALLEL ESTIMATION

As already mentioned in Section II, tuning of the identification algorithm to the unknown, and possibly time-varying form and rate of parameter changes can be achieved by means of parallel estimation. Consider L KF-based identification algorithms, equipped with different settings $\pi = \{m, \xi\} \in \Pi = \{\pi_1, \ldots, \pi_L\}$, $\pi_i = \{m_i, \xi_i\}$, $i = 1, \ldots, L$, yielding the estimates

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}_\pi(t|t) &= [\widehat{\theta}_{1|\pi}(t|t), \ldots, \widehat{\theta}_{n|\pi}(t|t)]^T \\
\widehat{\boldsymbol{\theta}}_\pi(t|N) &= [\widehat{\theta}_{1|\pi}(t|N), \ldots, \widehat{\theta}_{n|\pi}(t|N)]^T \\
l &= 1, \ldots, L.
\end{aligned}$$

The algorithms are run simultaneously and at each time instant only one of the competing estimates is selected, i.e., the estimated parameter trajectory has the form

$$\widehat{\boldsymbol{\theta}}(t|t) = \widehat{\boldsymbol{\theta}}_{\widehat{\pi}(t)}(t|t), \quad \widehat{\boldsymbol{\theta}}(t|N) = \widehat{\boldsymbol{\theta}}_{\widehat{\pi}(t)}(t|N)$$

where

$$\widehat{\pi}(t) = \{\widehat{m}(t), \widehat{\xi}(t)\} = \arg\min_{\pi\in\Pi} J_\pi(t) \tag{20}$$

and $J_\pi(t)$ denotes the local performance index.

In the causal (filtering) case, selection can be based on minimization of the sum of squared one-step-ahead output prediction errors observed in the recent past [32]

$$J_\pi(t) = \sum_{i=0}^{I} \varepsilon_\pi^2(t-i) \tag{21}$$

where

$$\varepsilon_\pi(t) = y(t) - \boldsymbol{\varphi}^T(t)\widehat{\boldsymbol{\theta}}_\pi(t|t-1) \tag{22}$$

and

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}_\pi(t|t-1) &= [\widehat{\theta}_{1|\pi}(t|t-1), \ldots, \widehat{\theta}_{n|\pi}(t|t-1)]^T \\
\widehat{\theta}_{j|\pi}(t|t-1) &= \boldsymbol{c}_m^T \widehat{\boldsymbol{x}}_{j|\pi}(t|t-1), \quad j = 1, \ldots, n.
\end{aligned}$$

In the noncausal (smoothing) case, one can use the approach known as cross-validation. In this approach prediction errors are replaced with leave-one-out output interpolation errors (sometimes referred to as deleted residuals)

$$\eta_\pi^\circ(t) = y(t) - \boldsymbol{\varphi}^T(t)\widehat{\boldsymbol{\theta}}_\pi^\circ(t|N) \tag{23}$$

where

$$\widehat{\boldsymbol{\theta}}_\pi^\circ(t|N) = [\widehat{\theta}_{1|\pi}^\circ(t|N), \ldots, \widehat{\theta}_{n|\pi}^\circ(t|N)]^T$$

denotes the holey estimate of $\boldsymbol{\theta}(t)$, i.e., the one that excludes from the estimation process the interpolated sample $y(t)$

$$\begin{aligned}
\widehat{\theta}_{j|\pi}^\circ(t|N) &= \boldsymbol{c}_m^T E[\boldsymbol{x}_j(t)|\Omega^\circ(N)] \\
\Omega^\circ(N) &= \Omega(N) - \{y(t)\}.
\end{aligned}$$

TABLE 2. The values of the width $K = 2k + 1$ of the analysis interval corresponding to the selected values of the estimation memory span M_∞, and to the number of basis functions m.

$M_\infty \setminus m$	1	2	3
20	21	21	45
40	41	41	91
80	81	81	181
160	161	161	361

It can be shown that (see Appendix B)

$$\widehat{\theta}^{\circ}_{j|\pi}(t|N) = \frac{\widehat{\theta}_{j|\pi}(t|N) - \beta_\pi(t|N)\theta^*_j(t)}{1 - \beta_\pi(t|N)} \tag{24}$$

where

$$\beta_\pi(t|N) = \boldsymbol{c}_m^T \boldsymbol{P}_\pi(t|N)\boldsymbol{c}_m.$$

More generally

$$\widehat{\boldsymbol{\theta}}^{\circ}_{\pi}(t|N) = \frac{\widehat{\boldsymbol{\theta}}_{\pi}(t|N) - \beta_\pi(t|N)\boldsymbol{\theta}^*(t)}{1 - \beta_\pi(t|N)} \tag{25}$$

which means that there is no need to implement the holey estimation scheme in order to evaluate interpolation errors (23). When the steady state version of the KF algorithm is used, the formula (24) can be further simplified by replacing the time-dependent coefficient $\beta_\pi(t|N)$ with its asymptotic (constant) version $\beta^\infty_\pi = \boldsymbol{c}_m^T \widetilde{\boldsymbol{P}}^\infty_\pi \boldsymbol{c}_m$, where $\widetilde{\boldsymbol{P}}^\infty_\pi = \lim_{t\to\infty} \boldsymbol{P}_\pi(t|2t)$.

The local cross-validation decision statistic can be defined in the following way

$$J_\pi(t) = \sum_{i=-I/2}^{I/2} [\eta^\circ_\pi(t+i)]^2. \tag{26}$$

The recommended values of I are those from the interval [20, 50] for the filtering algorithm, and from the interval [50, 100] for the smoothing algorithm. Another practical advice concerns selection of smoothness coefficients: for a fixed value of m the competing values of ξ should be chosen in such a way that the corresponding memory spans N_∞ and M_∞ form a geometric progression [32], [38], e.g. N_∞, $M_\infty = 20, 40, 80, 160$ etc.

VII. COMPUTER SIMULATIONS

In our simulation study we will focus on comparison of the noncausal KF algorithm (18)-(19) and simplified cross-validation decision rule incorporating β^∞_π, with the state-of-the-art BF smoother proposed recently in [23]. The noncausal BF estimates can be obtained by minimizing the local sum of squared output modeling errors

$$\widehat{\boldsymbol{\theta}}_{m|k}(\cdot) = arg\min_{\boldsymbol{\theta}(\cdot)} \sum_{i=-k}^{k} [y(t+i) - \boldsymbol{\varphi}^T(t+i)\boldsymbol{\theta}(t+i)]^2 \tag{27}$$

under the constraints

$$\theta_j(t+i) = \sum_{l=1}^{m} a_{jl} i^{l-1}, \quad i \in T_k(t),\ j = 1, \ldots, n \tag{28}$$

TABLE 3. Mean squared parameter estimation errors obtained for 12 noncausal basis function and Kalman filter estimators with different memory settings (M_∞) and different orders of the hypermodel (m), and for the corresponding adaptive parallel estimation schemes (A). Simulations were carried out for 2 signal-to-noise ratios (SNR) and 3 speeds of parameter variation (SoV). All averages were computed for 100 process realizations.

LBF

SNR	SoV	slow			medium			fast		
	$M_\infty \setminus m$	1	2	3	1	2	3	1	2	3
15 dB	20	3.40E-03	3.65E-03	3.46E-03	4.31E-03	4.24E-03	4.15E-03	6.24E-03	5.27E-03	5.35E-03
	40	2.44E-03	2.25E-03	2.34E-03	3.90E-03	3.20E-03	3.40E-03	7.68E-03	5.62E-03	5.69E-03
	80	2.76E-03	2.36E-03	2.54E-03	5.91E-03	4.67E-03	4.68E-03	1.95E-02	1.60E-02	1.84E-02
	160	4.93E-03	4.29E-03	4.13E-03	1.74E-02	1.53E-02	1.72E-02	7.80E-02	7.39E-02	1.01E-01
	A		**1.81E-03**			**2.75E-03**			**4.45E-03**	
25 dB	20	9.94E-04	8.24E-04	8.95E-04	1.68E-03	1.26E-03	1.39E-03	3.81E-03	2.36E-03	2.72E-03
	40	1.28E-03	1.03E-03	1.16E-03	2.55E-03	1.91E-03	2.12E-03	6.56E-03	4.34E-03	4.48E-03
	80	2.20E-03	1.78E-03	1.99E-03	5.20E-03	4.05E-03	4.01E-03	1.88E-02	1.53E-02	1.77E-02
	160	4.66E-03	3.96E-03	3.86E-03	1.70E-02	1.51E-02	1.69E-02	7.78E-02	7.37E-02	1.01E-01
	A		**6.60E-04**			**1.14E-03**			**2.33E-03**	

KF

SNR	SoV	slow			medium			fast		
	$M_\infty \setminus m$	1	2	3	1	2	3	1	2	3
15 dB	20	3.23E-03	3.26E-03	3.55E-03	4.06E-03	4.10E-03	4.45E-03	5.94E-03	5.84E-03	6.28E-03
	40	2.09E-03	2.15E-03	2.17E-03	3.04E-03	3.07E-03	3.18E-03	5.97E-03	5.03E-03	5.24E-03
	80	2.06E-03	2.10E-03	2.21E-03	4.49E-03	3.69E-03	3.83E-03	1.54E-02	1.33E-02	1.32E-02
	160	3.98E-03	3.19E-03	3.43E-03	1.44E-02	1.25E-02	1.42E-02	4.51E-02	5.89E-02	7.06E-02
	A		**1.70E-03**			**2.63E-03**			**4.53E-03**	
25 dB	20	9.02E-04	9.19E-04	9.72E-04	1.54E-03	1.55E-03	1.63E-03	3.69E-03	3.58E-03	3.79E-03
	40	9.18E-04	9.69E-04	1.06E-03	1.76E-03	1.77E-03	1.95E-03	4.83E-03	3.89E-03	4.17E-03
	80	1.47E-03	1.51E-03	1.62E-03	3.84E-03	3.03E-03	3.16E-03	1.48E-02	1.27E-02	1.25E-02
	160	3.67E-03	2.89E-03	3.14E-03	1.41E-02	1.22E-02	1.39E-02	4.47E-02	5.86E-02	7.02E-02
	A		**7.02E-04**			**1.24E-03**			**3.08E-03**	

where $T_k(t) = [t-k, t+k]$ denotes the local analysis interval of width $K = 2k+1$, centered at t. Even though such identification scheme allows one to estimate the entire segment of the parameter trajectory $\{\boldsymbol{\theta}(s), s \in T_k(t)\}$, it is used to generate a sequence of point estimates instead of interval ones, i.e., only the "central" estimate $\widehat{\boldsymbol{\theta}}_{m|k}(t)$ is utilized at the instant t, and the estimation process is repeated for a new position of the sliding analysis window $T_k(t)$. The resulting local basis function (LBF) algorithm can be regarded as a generalization – to the system identification case – of the classical signal smoothing technique known as Savitzky-Golay filtering [39], [40].

The estimation memory of the LBF algorithm can be obtained from

$$M_\infty = \left[\sum_{i=-k}^{k} h_{m|k}^2(i)\right]^{-1} \tag{29}$$

where $h_{m|k}(i)$ denotes the so-called impulse response associated with the LBF estimator [23]

$$h_{m|k}(i) = \boldsymbol{f}_m^T(0)\left[\sum_{i=-k}^{k} \boldsymbol{f}_m(i)\boldsymbol{f}_m^T(i)\right]^{-1}\boldsymbol{f}_m(i) \tag{30}$$

and $\boldsymbol{f}_m(i) = [1, i, \ldots, i^{m-1}]^T$. The values of the width $K = 2k+1$ of the analysis interval corresponding to the selected values of M_∞ and to the order m of the polynomial basis are shown in Table 2.

The identified nonstationary system was that presented earlier in Section III. In addition to the medium speed parameter variation scenario, depicted in Fig. 1, the two times faster and two times slower variations were considered, obtained by "resampling" the parameter trajectory without changing its shape. As a result, the length of the simulation interval T_s was equal to 1000, 2000 and 4000 for fast, medium speed and slow changes, respectively. To check behavior of the compared algorithms under different noise conditions, two average signal-to-noise ratios were considered: 25 dB ($\sigma_e^2 = 0.0025$) and 15 dB ($\sigma_e^2 = 0.025$).

The KF-based parallel estimation scheme was made up of 12 smoothers designed for 3 IRW models of orders $m = 1, 2, 3$ and 4 estimation memory spans $M_\infty = 20, 40, 80, 160$. Analogously, the adaptive BF algorithm combined 12 LBF smoothers designed for 3 polynomial bases of order $m = 1, 2, 3$ and 4 estimation memory spans $M_\infty = 20, 40, 80, 160$. In both cases the width of the decision window was set to $I + 1 = 61$. Under such conditions the KF and BF approaches can be expected to behave comparably, as the corresponding hypermodels and their memory settings are statistically compatible. Data generation was started 1000 instants prior to $t = 1$ and was continued for 1000 instants after $t = T_s$, so that, irrespective of settings, the estimation process and evaluation of its results could be, for all algorithms, started at the instant 1 and ended at the instant T_s. For $t < 1$ and $t > T_s$ system parameters were constant and equal to $\boldsymbol{\theta}(1)$ and $\boldsymbol{\theta}(T_s)$, respectively.

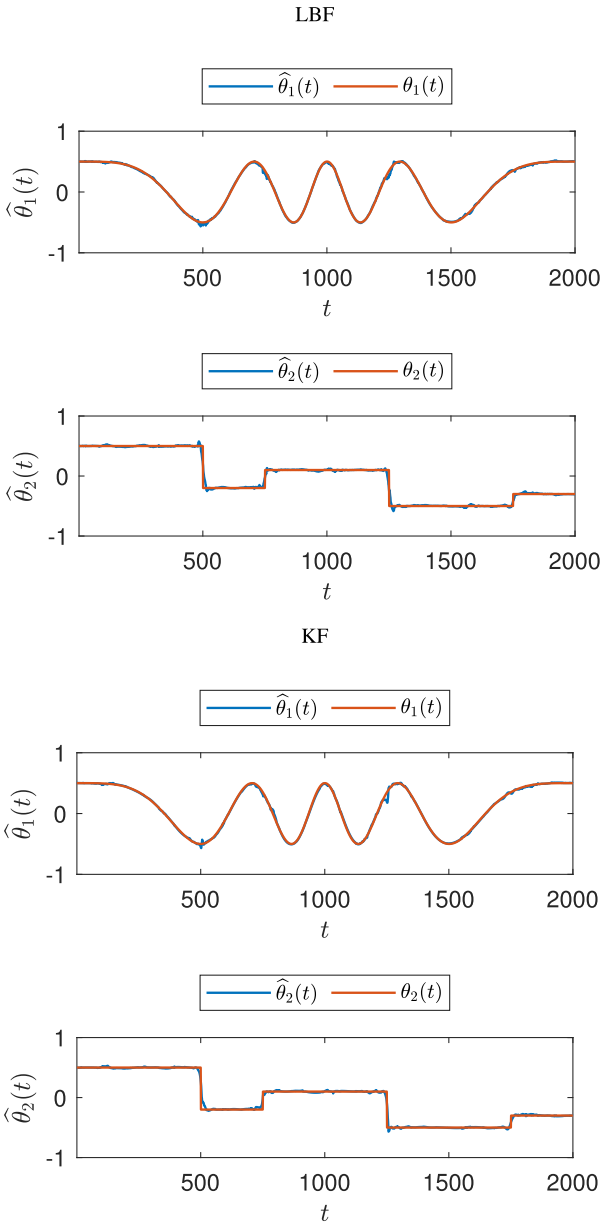

FIGURE 2. Identification results obtained using the adaptive LBF algorithm (two upper plots) and adaptive decoupled KF algorithm (two lower plots). The estimated trajectories (blue lines), obtained by postprocessing the preestimates shown in Fig. 1, are superimposed on the true trajectories (red lines).

Table 3 presents estimation results - the squared parameter estimation errors $\| \widehat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t) \|^2$ averaged over time and over 100 process realizations. The best results in each group are shown in boldface. Noticeably, in all cases considered, the adaptive parallel estimation algorithms yielded results that were better than those provided by their component algorithms with fixed settings. In the case of low SNR the adaptive KF algorithm yielded slightly better results than the adaptive LBF algorithm, while for the higher SNR the adaptive LBF performed slightly better than the KF one. Overall, however, for the same values of M_∞ and m both types

of algorithms yielded a similar tracking performance. Typical identification results yielded by the adaptive scheme are shown in Fig. 2. As expected, the results yielded by the deterministic BF approach are comparable with those obtained for the stochastic KF approach. This makes the decoupled KF algorithm a computationally attractive alternative to the LBF algorithm. The high computational requirements of the LBF algorithm, of order $O(m^3n^3)$, are due to the fact that it requires evaluation and inversion, every time instant t, of the $mn \times mn$-dimensional generalized regression matrix $\boldsymbol{G}_{m|k}(t) = \sum_{i=-k}^{k} \boldsymbol{\psi}_m(t,i)\boldsymbol{\psi}_m^T(t,i)$, where $\boldsymbol{\psi}_m(t,i) = \boldsymbol{\varphi}(t+i) \otimes \boldsymbol{f}_m(i)$ and $\otimes$ denotes the Kronecker product.

Remark: We note that a meaningful comparison of the proposed decoupled KF approach with the classical one is not possible since without equalization of the estimation memory spans of the compared algorithms – which cannot be performed due to limitations of the classical scheme, mentioned in Section II – such comparison would not make much sense.

VIII. CONCLUSION

The problem of identification of a nonstationary FIR system was considered and solved using the decoupled Kalman filtering/smoothing approach. Unlike the classical KF-based solutions, the proposed estimation algorithms are easy to tune. It was shown that optimization of their tracking performance can be carried out using parallel estimation and cross-validation. Due to low computational requirements (achieved without compromising good tracking capabilities) the decoupled KF algorithms can be regarded as an attractive alternative to the state-of-the-art algorithms based on functional series approximation, known as basis function algorithms.

APPENDIX A
DERIVATION OF (14)

Let $m = 1$. The steady state value of the *a priori* variance $p_\pi^\infty = \lim_{t\to\infty} p_\pi(t|t-1)$ is a solution of the Riccati equation

$$p_\pi^\infty = \frac{p_\pi^\infty}{1+p_\pi^\infty} + \xi$$

leading to

$$p_\pi^\infty = \frac{\xi + \sqrt{\xi^2+4\xi}}{2} \cong \sqrt{\xi}$$

where the approximation holds for $\xi_j \ll 1$ (cf. Table 1).

Without any loss of generality, one can assume that $\theta_j^0 = 0$. In this case the steady state relationship between $\widehat{\theta}_{j|\pi}(t|t) = \widehat{x}_{j|\pi}(t|t)$ and $z_j(t)$ has the form

$$\widehat{\theta}_{j|\pi}(t|t) = (1-k_\pi^\infty)\widehat{\theta}_{j|\pi}(t-1|t-1) + k_\pi^\infty z_j(t)$$

where $k_\pi^\infty = p_\pi^\infty/(1+p_\pi^\infty) \cong \sqrt{\xi}$ denotes the steady state Kalman gain. Based on this relationship, one obtains

$$\sigma_{\theta_j}^2 = \frac{k_\pi^\infty}{2-k_\pi^\infty}\sigma_{z_j}^2 \cong \frac{k_\pi^\infty}{2}\sigma_{z_j}^2$$

which leads to (14).

APPENDIX B
DERIVATION OF (24)

Let $\Omega(t) = \{y(i), \boldsymbol{\varphi}(i), 1 \le i \le t\}$ and $\Omega^+(t) = \{y(i), \boldsymbol{\varphi}(i), t \le i \le N\}$. Using the Mayne-Fraser two-filter formula, the smoothed estimate $\widehat{\boldsymbol{x}}_{j|\pi}(t|N)$ can be expressed in the form [41]

$$\begin{aligned}\widehat{\boldsymbol{x}}_{j|\pi}(t|N) &= \boldsymbol{P}_\pi(t|N)\big\{[\boldsymbol{P}_\pi(t|t)]^{-1}\widehat{\boldsymbol{x}}_{j|\pi}(t|t)\\ &\quad +[\boldsymbol{P}_\pi^+(t|t+1)]^{-1}\widehat{\boldsymbol{x}}_{j|\pi}^+(t|t+1)\big\}\\ \boldsymbol{P}_\pi(t|N) &= \big\{[\boldsymbol{P}_\pi(t|t)]^{-1} + [\boldsymbol{P}_\pi^+(t|t+1)]^{-1}\big\}^{-1} \quad (31)\end{aligned}$$

where $\widehat{\boldsymbol{x}}_{j|\pi}^+(t|t+1) = E[\boldsymbol{x}_j(t)|\Omega^+(t+1)]$ denotes the backward-time (anticausal) predictor, designed for the backwards Markovian model of $\boldsymbol{x}_j(t)$ and operated in reverse time. Similarly, $\boldsymbol{P}_\pi^+(t|t+1)$ denotes the associated covariance matrix.

The analogous expression for the holey estimator of $\boldsymbol{x}_j(t)$ is

$$\begin{aligned}\widehat{\boldsymbol{x}}_{j|\pi}^\circ(t|N) &= \boldsymbol{P}_\pi^\circ(t|N)\big\{[\boldsymbol{P}_\pi(t|t-1)]^{-1}\widehat{\boldsymbol{x}}_{j|\pi}(t|t-1)\\ &\quad +[\boldsymbol{P}_\pi^+(t|t+1)]^{-1}\widehat{\boldsymbol{x}}_{j|\pi}^+(t|t+1)\big\}\\ \boldsymbol{P}_\pi^\circ(t|N) &= \big\{[\boldsymbol{P}_\pi(t|t-1)]^{-1} + [\boldsymbol{P}_\pi^+(t|t+1)]^{-1}\big\}^{-1}. \quad (32)\end{aligned}$$

In order to prove (24), we first notice that

$$\begin{aligned}\boldsymbol{P}_\pi(t|t) &= \boldsymbol{P}_\pi(t|t-1) - \frac{\boldsymbol{P}_\pi(t|t-1)\boldsymbol{c}_m\boldsymbol{c}_m^T\boldsymbol{P}_\pi(t|t-1)}{1+\boldsymbol{c}_m^T\boldsymbol{P}_\pi(t|t-1)\boldsymbol{c}_m}\\ &= \big[\boldsymbol{P}_\pi^{-1}(t|t-1) + \boldsymbol{c}_m\boldsymbol{c}_m^T\big]^{-1} \quad (33)\end{aligned}$$

where the transition follows from the well-known matrix inversion lemma [6]. Combining (31) - (33) and using matrix inversion lemma again, one arrives at

$$\begin{aligned}\boldsymbol{P}_\pi^\circ(t|N) &= \big[\boldsymbol{P}_\pi^{-1}(t|N) - \boldsymbol{c}_m\boldsymbol{c}_m^T\big]^{-1}\\ &= \boldsymbol{P}_\pi(t|N) + \frac{\boldsymbol{P}_\pi(t|N)\boldsymbol{c}_m\boldsymbol{c}_m^T\boldsymbol{P}_\pi(t|N)}{1-\beta_\pi(t|N)} \quad (34)\end{aligned}$$

where $\beta_\pi(t|N) = \boldsymbol{c}_m^T\boldsymbol{P}_\pi(t|N)\boldsymbol{c}_m$. Furthermore, since [cf. (33)]

$$\boldsymbol{P}_\pi^{-1}(t|t) = \boldsymbol{P}_\pi^{-1}(t|t-1) + \boldsymbol{c}_m\boldsymbol{c}_m^T \quad (35)$$

and [cf. (12)]

$$\begin{aligned}\widehat{\boldsymbol{x}}_{j|\pi}(t|t) &= \widehat{\boldsymbol{x}}_{j|\pi}(t|t-1)\\ &\quad + \frac{\boldsymbol{P}_\pi(t|t-1)\boldsymbol{c}_m}{1+\boldsymbol{c}_m^T\boldsymbol{P}_\pi(t|t-1)\boldsymbol{c}_m}[\theta_j^*(t) - \boldsymbol{c}_m^T\widehat{\boldsymbol{x}}_{j|\pi}(t|t-1)] \quad (36)\end{aligned}$$

after straightforward calculations, one arrives at

$$\boldsymbol{P}_\pi^{-1}(t|t)\widehat{\boldsymbol{x}}_{j|\pi}(t|t) = \boldsymbol{P}_\pi^{-1}(t|t-1)\widehat{\boldsymbol{x}}_{j|\pi}(t|t-1) + \boldsymbol{c}_m\theta_j^*(t). \quad (37)$$

Substituting (34) and (37) into (32), one obtains

$$\begin{aligned}\widehat{\boldsymbol{x}}_{j|\pi}^\circ(t|N) &= \left[\boldsymbol{P}_\pi(t|N) + \frac{\boldsymbol{P}_\pi(t|N)\boldsymbol{c}_m\boldsymbol{c}_m^T\boldsymbol{P}_\pi(t|N)}{1-\beta_\pi(t|N)}\right]\\ &\quad \times[\boldsymbol{P}_\pi^{-1}(t|N)\widehat{\boldsymbol{x}}_{j|\pi}(t|N) - \boldsymbol{c}_m\theta_j^*(t)]\end{aligned}$$

$$= \widehat{\boldsymbol{x}}_{j|\pi}(t|N) + \frac{\boldsymbol{P}_{\pi}(t|N)\boldsymbol{c}_m\boldsymbol{c}_m^T}{1-\beta_{\pi}(t|N)}\widehat{\boldsymbol{x}}_{j|\pi}(t|N)$$
$$-\boldsymbol{P}_{\pi}(t|N)\boldsymbol{c}_m\theta_j^*(t) - \frac{\beta_{\pi}(t|N)\boldsymbol{P}_{\pi}(t|N)\boldsymbol{c}_m}{1-\beta_{\pi}(t|N)}\theta_j^*(t). \quad (38)$$

Finally, after multiplying both sides of (38) by $\boldsymbol{c}_m^T$ and noting that $\boldsymbol{c}_m^T\widehat{\boldsymbol{x}}^{\circ}_{j|\pi}(t|N) = \widehat{\theta}^{\circ}_{j|\pi}(t|N)$ and $\boldsymbol{c}_m^T\widehat{\boldsymbol{x}}_{j|\pi}(t|N) = \widehat{\theta}_{j|\pi}(t|N)$, one obtains (24).

ACKNOWLEDGMENT

Computer simulations were carried out at the Academic Computer Centre in Gdańsk.

REFERENCES

[1] M. K. Tsatsanis and G. B. Giannakis, "Modelling and equalization of rapidly fading channels," *Int. J. Adapt. Control Signal Process.*, vol. 10, nos. 2–3, pp. 159–176, 1996.
[2] J. Bakkoury, D. Roviras, M. Ghogho, and F. Castanie, "Adaptive MLSE receiver over rapidly fading channels," *Signal Process.*, vol. 80, no. 7, pp. 1347–1360, Jul. 2000.
[3] L. Shen, Y. Zakharov, B. Henson, N. Morozs, and P. D. Mitchell, "Adaptive filtering for full-duplex UWA systems with time-varying self-interference channel," *IEEE Access*, vol. 8, pp. 187590–187604, Oct. 2020.
[4] L. Shen, B. Henson, Y. Zakharov, and P. Mitchell, "Digital self-interference cancellation for full-duplex underwater acoustic systems," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 1, pp. 192–196, Jan. 2020.
[5] M. Stojanovic and J. Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 84–89, Jan. 2009.
[6] T. Söderström and P. Stoica, *System Identification*. Englewoods Cliffs, NJ, USA: Prentice-Hall, 1988.
[7] M. Niedźwiecki, *Identification of Time-varying Processes*. New York, NY, USA: Wiley, 2000.
[8] T. S. Rao, "The fitting of nonstationary time-series models with time-dependent parameters," *J. Roy. Stat. Soc., Ser. B (Methodol.)*, vol. 32, no. 2, pp. 312–322, 1970.
[9] J. M. Mendel, *Discrete Techniques of Parameter Estimation: The Equation Error Formulation*. New York, NY, USA: Marcel Dekker, 1973.
[10] L. A. Liporace, "Linear estimation of nonstationary signals," *J. Acoust. Soc. Amer.*, vol. 58, no. 6, pp. 1288–1295, Dec. 1975.
[11] M. G. Hall, A. V. Oppenheim, and A. S. Willsky, "Time-varying parametric modeling of speech," *Signal Process.*, vol. 5, no. 3, pp. 267–285, May 1983.
[12] Y. Grenier, "Time-dependent ARMA modeling of nonstationary signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, no. 4, pp. 899–911, Aug. 1983.
[13] R. Charbonnier, M. Barlaud, G. Alengrin, and J. Menez, "Results on AR-modelling of nonstationary signals," *Signal Process.*, vol. 12, no. 2, pp. 143–151, Mar. 1987.
[14] M. Niedzwiecki, "Functional series modeling approach to identification of nonstationary stochastic systems," *IEEE Trans. Autom. Control*, vol. 33, no. 10, pp. 955–961, Oct. 1988.
[15] M. Niedzwiecki, "Recursive functional series modeling estimators for identification of time-varying plants-more bad news than good?" *IEEE Trans. Autom. Control*, vol. 35, no. 5, pp. 610–616, May 1990.
[16] R. B. Mrad, S. D. Fassois, and J. A. Levitt, "A polynomial-algebraic method for non-stationary TARMA signal analysis–Part I: The method," *Signal Process.*, vol. 65, pp. 1–19, Feb. 1998, doi: 10.1016/S0165-1684(97)00145-X.
[17] H.-L. Wei and S. A. Billings, "Identification of time-varying systems using multiresolution wavelet models," *Int. J. Syst. Sci.*, vol. 33, no. 15, pp. 1217–1228, Jan. 2003.
[18] R. Zou, H. Wang, and K. H. Chon, "A robust time-varying identification algorithm using basis functions," *Ann. Biomed. Eng.*, vol. 31, no. 7, pp. 840–853, Jul. 2003.
[19] A. G. Poulimenos and S. D. Fassois, "Parametric time-domain methods for non-stationary random vibration modelling and analysis—A critical survey and comparison," *Mech. Syst. Signal Process.*, vol. 20, no. 4, pp. 763–816, May 2006.
[20] 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," *Int. J. Model., Identificat. Control*, vol. 9, no. 3, pp. 215–224, 2010.
[21] S. C. Chan and Z. G. Zhang, "Local polynomial modeling and variable bandwidth selection for time-varying linear systems," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 3, pp. 1102–1117, Mar. 2011.
[22] M. Niedzwiecki and S. Gackowski, "New approach to noncausal identification of nonstationary stochastic FIR systems subject to both smooth and abrupt parameter changes," *IEEE Trans. Autom. Control*, vol. 58, no. 7, pp. 1847–1853, Jul. 2013.
[23] M. Niedzwiecki and M. Ciołek, "Generalized Savitzky–Golay filters for identification of nonstationary systems," *Automatica*, vol. 108, Oct. 2019, Art. no. 108477.
[24] D. Q. Mayne, "Optimal non-stationary estimation of the parameters of a linear system with Gaussian inputs," *J. Electron. Control*, vol. 14, no. 1, pp. 101–112, Jan. 1963.
[25] J. P. Norton, "Optimal smoothing in the identification of linear time-varying systems," *Proc. Inst. Elect. Eng.*, vol. 122, pp. 663–668, Jun. 1975.
[26] T. Bohlin, "Four cases of identification of changing systems," in *Mathematics in Science and Engineering*, vol. 126, R. K. Mehra and D. G. Lainiotis, Eds. Amsterdam, The Netherlands: Elsevier, 1976, pp. 441–518.
[27] H. Akaike, "Likelihood and the Bayes procedure," in *Bayesian Statistics*, J. M. Bernado, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, Eds. Valencia, Spain: Univ. Press, 1980, pp. 143–166.
[28] G. Kitagawa and W. Gersch, "A smoothness priors time-varying AR coefficient modeling of nonstationary covariance time series," *IEEE Trans. Autom. Control*, vol. AC-30, no. 1, pp. 48–56, Jan. 1985.
[29] X.-Q. Jiang and G. Kitagawa, "A time varying coefficient vector AR modeling of nonstationary covariance time series," *Signal Process.*, vol. 33, no. 3, pp. 315–331, Sep. 1993.
[30] M. Arnold, X. H. R. Milner, H. Witte, R. Bauer, and C. Braun, "Adaptive AR modeling of nonstationary time series by means of Kalman filtering," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 5, pp. 553–562, May 1998.
[31] M. Niedzwiecki, "Locally adaptive cooperative Kalman smoothing and its application to identification of nonstationary stochastic systems," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 48–59, Jan. 2012.
[32] M. Niedzwiecki, "Identification of nonstationary stochastic systems using parallel estimation schemes," *IEEE Trans. Autom. Control*, vol. 35, no. 3, pp. 329–334, Mar. 1990.
[33] M. Niedźwiecki and T. Kłaput, "Fast recursive basis function estimators for identification of nonstationary systems," *IEEE Trans. Signal Process.*, vol. 50, pp. 1925–1934, 2002.
[34] M. J. Niedzwiecki, M. Ciołek, and A. Gancza, "A new look at the statistical identification of nonstationary systems," *Automatica*, vol. 118, Aug. 2020, Art. no. 109037.
[35] M. Niedzwiecki, M. Ciolek, A. Gancza, and P. Kaczmarek, "A new method of noncausal identification of time-varying systems," in *Proc. Signal Process., Algorithms, Archit., Arrangements, Appl. (SPA)*, Poznań, Poland, Sep. 2020, pp. 48–52.
[36] M. Niedzwiecki, A. Gancza, and M. Ciolek, "On the preestimation technique and its application to identification of nonstationary systems," in *Proc. 59th IEEE Conf. Decis. Control (CDC)*, Jeju Island, Republic of Korea, Dec. 2020, pp. 286–293.
[37] N. Assimakis and M. Adam, "Iterative and algebraic algorithms for the computation of the steady state Kalman filter gain," *Int. Scholarly Res. Notices*, vol. 2014, May 2014, Art. no. 417623.
[38] M. Niedzwiecki, M. Ciołek, and Y. Kajikawa, "On adaptive covariance and spectrum estimation of locally stationary multivariate processes," *Automatica*, vol. 82, pp. 1–12, Aug. 2017.
[39] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, Jul. 1964.
[40] R. W. Schafer, "What is a Savitzky-Golay filter?" *IEEE Signal Process. Mag.*, vol. 28, no. 4, pp. 111–117, Jul. 2011.
[41] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewoods Cliffs, NJ, USA: Prentice-Hall, 1979.

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

MACIEJ NIEDŹWIECKI (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees from the Technical University of Gdańsk, Gdańsk, Poland, in 1977 and 1981, respectively, and the Dr.Hab. (D.Sc.) degree from the Technical University of Warsaw, Warsaw, Poland, in 1991.

From 1986 to 1989, he was a Research Fellow with the Department of Systems Engineering, Australian National University. From 1990 to 1993, he worked as the Vice Chairman of the Technical Committee on Theory of the International Federation of Automatic Control (IFAC). He currently works as a Professor and the Head of the Department of Automatic Control, Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology. He is the author of the book *Identification of Time-varying Processes* (Wiley, 2000). His main research interests include system identification, statistical signal processing, and adaptive systems. He is also 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).

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