

IEEE Copyright Notice

- ©20xx IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.
- This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder.

Identification of Quasi-Periodically Varying Systems Using the Combined Nonparametric/Parametric Approach

Maciej Niedźwiecki and Piotr Kaczmarek

Abstract—The problem of identification of quasi-periodically varying finite impulse response systems is considered. Neither the number of system frequency modes nor the initial frequency values are assumed to be known a priori. The proposed solution is a blend of the parametric (model based) and nonparametric (discrete Fourier transform based) approach to system identification. It is shown that the results of nonparametric analysis can be used to identify the number of frequency modes and to determine initial conditions needed to smoothly start (or restart) the model-based tracking algorithm. Such a combined nonparametric/parametric approach allows one to preserve advantages of both frameworks, leading to an estimation procedure which guarantees global frequency search, high-frequency resolution, fast initial convergence, and good steady-state tracking capabilities.

Index Terms—Basis function approach, frequency estimation, system identification, time-varying processes.

I. INTRODUCTION

A. Problem Statement

CONSIDER the problem of identification/tracking of coefficients of a quasi-periodically varying complex system (system with input/output signals and parameters described by complex numbers) governed by

$$y(t) = \sum_{l=1}^n \theta_l(t)u(t-l+1) + v(t) = \boldsymbol{\varphi}^T(t)\boldsymbol{\theta}(t) + v(t) \quad (1)$$

where $t = 1, 2, \dots$, denotes the normalized discrete time, $y(t)$ denotes the system output, $\boldsymbol{\varphi}(t) = [u(t), \dots, u(t-n+1)]^T$ is the regression vector made up of the past input samples, $v(t)$ is an additive (white) noise, uncorrelated with $u(t)$, and $\boldsymbol{\theta}(t) = [\theta_1(t), \dots, \theta_n(t)]^T$ denotes the vector of time varying impulse response coefficients, modeled as weighted sums of complex exponentials

$$\theta_l(t) = \sum_{i=1}^k a_{li}(t)e^{j \sum_{s=1}^t \omega_i(s)}, \quad l = 1, \dots, n. \quad (2)$$

Manuscript received July 28, 2004. This work was supported by KBN under Grant 4 T11A 01225. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Abdelhak M. Zoubir.

The authors are with the Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control, Gdańsk University of Technology, Gdańsk, Poland (e-mail: maciekn@eti.pg.gda.pl; piokacz@proterians.net.pl).

Digital Object Identifier 10.1109/TSP.2005.859221

Since the amplitudes $a_{li}(t)$ are complex valued, they incorporate both magnitude and phase information. Therefore an explicit phase component is not needed in (2).

We will assume that both the amplitudes $a_{li}(t)$, $l = 1, \dots, n$, and frequencies $\omega_i(t)$ in (2) are slowly time-varying, and that $v(t)$ is a complex white measurement noise of variance σ_v^2 , independent of the input signal $u(t)$. Neither the number of frequency modes nor the initial frequency values are supposed to be known a priori.

One of the interesting modern applications, which admits such problem formulation, is identification of rapidly fading mobile radio channels. In a typical mobile radio scenario, one station is fixed in position and the other station is moving. In urban environments the direct line between transmitter and receiver is usually obstructed by buildings, so propagation of electromagnetic energy to and from the mobile unit is largely by scattering, e.g., by reflection from the flat sides of buildings. When scattering is caused by a few strong reflectors the impulse response of the channel $\theta_l(t)$, $l = 1, \dots, n$ can be approximately written down in the form (2)—see, e.g., [1]–[3]. In this particular application k is the number of dominant reflectors (note that such reflectors may suddenly appear and/or disappear as the terrain changes) and the angular frequencies $\omega_1, \dots, \omega_k$ correspond to Doppler shifts along different paths of signal arrival (when the speed of the vehicle changes over time, Doppler shifts are also time-varying).

When $n = 1$ and $u(t) = 1, \forall t$ (1) and (2) become a description of a noisy nonstationary multifrequency signal

$$y(t) = \theta(t) + v(t) = \sum_{i=1}^k a_i(t)e^{j \sum_{s=1}^t \omega_i(s)} + v(t). \quad (3)$$

The problem of either elimination or extraction of complex sinusoidal signals (called cisoids) buried in noise can be solved using adaptive notch filtering—see, e.g., Tichavsky and Nehorai [4] and the references therein. For this reason the system identification/tracking algorithm described below can be considered a generalized adaptive notch filter.

B. Contribution and Novelty

There are two broad approaches to system identification. The first one, known as parametric approach, attempts to build a mathematical model of the analyzed process. The name stems from the fact that the functional form of the model (e.g., system equation) is assumed to be known and the identification

task is reduced to estimation, from experimental data, of a certain number of model parameters. In contrast with this, the nonparametric identification methods attempt to characterize system behavior without using a given parametrized model set. The frequency-domain approach to system identification, which attempts to estimate the system's frequency response, is an example of such a model-free technique.

The results of comparison of the parametric and nonparametric techniques is far from conclusive [5]. Parametric methods usually outperform nonparametric methods if the identified process (system or signal) fits into the assumed model class, i.e., if it can be closely approximated within a given model structure. On the other hand, such methods are usually quite sensitive to model misspecification. As a result they yield substantial bias errors if the model structure is chosen inadequately and/or if the order of the adopted model is underestimated.

The identification method presented in this paper combines elements of both frameworks in a way that preserves their advantages. We start from reviewing known properties of the exponentially weighted basis function (EWBF) and gradient basis function (GBF) estimators obtained for the model [(1) and (2)]. The frequency-adaptive versions of the EWBF/GBF algorithms, based on a simple gradient search strategy, are next derived. The EWBF/GBF trackers are capable of following slow changes in the amplitudes and frequencies of $\theta(t)$, but they require prior knowledge of the number of frequency modes, and may fail to work correctly in the presence of frequency jumps. Additionally, they may have problems with initial convergence if the initial frequency estimates are not sufficiently close to the true frequencies. To overcome difficulties mentioned above, a special nonparametric identification technique, based on the generalized (system) periodogram, is proposed. We show that both the number of frequency components and the corresponding frequency values can be estimated by means of localizing "significant" periodogram peaks. We prove that the frequency-selection problem can be solved using the suitably modified version of the Akaike's final prediction error criterion. Finally, we show that all initial conditions, needed to start (or restart) the EWBF/GBF algorithms, can be established using the results of nonparametric, periodogram-based analysis.

II. TRACKING ALGORITHMS

For the sake of clarity, the development of the parameter tracking algorithm will be carried out in two steps. First, we will review known results on the recursive EWBF and GBF algorithms (generalized notch filters). In the second step, we will show how both algorithms can be made frequency-adaptive.

A. Known Frequencies

Suppose that only the complex amplitudes in (2) are (slowly) time-varying, whereas the frequencies are constant and known, i.e.,

$$\theta_l(t) = \sum_{i=1}^k a_{li}(t)e^{j\omega_i t}.$$

Denote by $\alpha_i(t) = [a_{1i}(t), \dots, a_{ni}(t)]^T$ the vector of system coefficients associated with a particular frequency ω_i . Similarly, let $\psi_i(t) = \varphi(t)e^{j\omega_i t}$ be the generalized regression vector associated with the i th frequency component. Using the short-hand notation introduced above, (1) can be rewritten in the form

$$y(t) = \sum_{i=1}^k \psi_i^T(t)\alpha_i(t) + v(t) = \psi^T(t)\alpha(t) + v(t)$$

where $\alpha(t) = [\alpha_1^T(t), \dots, \alpha_k^T(t)]^T$, $\psi(t) = [\psi_1^T(t), \dots, \psi_k^T(t)]^T = \mathbf{f}(t) \otimes \varphi(t)$, $\mathbf{f}(t) = [e^{j\omega_1 t}, \dots, e^{j\omega_k t}]^T$, and \otimes denotes the Kronecker product of two vectors/matrices.

The estimator of $\theta(t)$, capable of tracking slow changes in $\alpha(t)$, can be obtained from

$$\begin{aligned} \hat{\alpha}(t) &= \arg \min_{\alpha} \sum_{s=1}^t \lambda^{t-s} \left| y(s) - \psi^T(s)\alpha \right|^2 \\ \hat{\theta}(t) &= \sum_{i=1}^k e^{j\omega_i t} \hat{\alpha}_i(t) \end{aligned} \quad (4)$$

where λ ($0 < \lambda < 1$, $1 - \lambda \ll 1$) denotes the so-called forgetting constant—the design parameter which controls the memory of the estimator, and hence allows one to trade off between its tracking speed and tracking accuracy.

Since the estimator (4) combines the basis function parameterization (it is assumed that system parameters can be expressed as linear combinations of known functions of time, called basis functions) with exponentially weighted least squares estimation, it will be further referred to as the exponentially weighted basis function estimator [6].

Straightforward calculations lead to

$$\begin{aligned} \hat{\alpha}(t) &= \mathbf{R}^{-1}(t)\mathbf{p}(t) \\ \mathbf{R}(t) &= \sum_{s=1}^t \lambda^{t-s} \psi^*(s)\psi^T(s) \\ \mathbf{p}(t) &= \sum_{s=1}^t \lambda^{t-s} y(s)\psi^*(s). \end{aligned} \quad (5)$$

The EWBF estimate can be evaluated recursively using

$$\begin{aligned} \varepsilon(t) &= y(t) - \psi^T(t)\hat{\alpha}(t-1) \\ \mathbf{k}(t) &= \frac{\mathbf{Q}(t-1)\psi(t)}{\lambda + \psi^H(t)\mathbf{Q}(t-1)\psi(t)} \\ \hat{\alpha}(t) &= \hat{\alpha}(t-1) + \mathbf{k}^*(t)\varepsilon(t) \\ \mathbf{Q}(t) &= \frac{1}{\lambda} \left[\mathbf{Q}(t-1) - \mathbf{k}(t)\psi^H(t)\mathbf{Q}(t-1) \right] \\ \hat{\theta}(t) &= \sum_{i=1}^k e^{j\omega_i t} \hat{\alpha}_i(t) \end{aligned} \quad (6)$$

where $\mathbf{Q}(t) = (\mathbf{R}^*(t))^{-1}$.

To avoid inversion of the matrix $\mathbf{R}^*(t)$ in the startup phase of the estimation, the initial conditions for (6) should be set to $\hat{\alpha}(0) = \mathbf{0}$ and $\mathbf{Q}(0) = \eta \mathbf{I}_{kn}$, where η denotes a large positive

constant—this is a standard initialization procedure for all recursive least square type recursive estimation algorithms [7].

The tracking characteristics of the EWBF algorithm, such as its equivalent estimation memory (deciding upon the estimation bandwidth, i.e., the frequency range in which estimation can be carried out successfully) and the associated impulse response, were established and analyzed in Niedźwiecki and Kłaput [8]. As argued in [8], when the input sequence $u(t)$ is wide-sense stationary and persistently exciting, the tracking characteristics mentioned above remain practically unchanged if the regression matrix $\mathbf{R}(t)$ in (5) is replaced with its expectation, i.e., if $\hat{\boldsymbol{\alpha}}(t)$ is replaced with

$$\tilde{\boldsymbol{\alpha}}(t) = \bar{\mathbf{R}}^{-1}(t)\mathbf{p}(t) \quad (7)$$

where

$$\bar{\mathbf{R}}(t) = \mathbb{E}[\mathbf{R}(t)] = \mathbf{F}(t) \otimes \boldsymbol{\Phi}, \quad \mathbf{F}(t) = \sum_{s=1}^t \lambda^{t-s} \mathbf{f}^*(s) \mathbf{f}^T(s),$$

$$\boldsymbol{\Phi} = \mathbb{E}[\boldsymbol{\varphi}^*(s) \boldsymbol{\varphi}^T(s)].$$

We will rely on this observation in Section III.

Even for moderate values of n (the number of system coefficients) and k (the number of frequency modes) the EWBF algorithm is computationally very demanding, as it requires updating the $nk \times nk$ matrix $\mathbf{Q}(t)$. The low complexity stochastic gradient counterpart of (6) can be obtained by replacing the matrix $\mathbf{Q}(t)$ with a scalar gain. The resulting gradient basis function algorithm can be written in the form

$$\begin{aligned} \varepsilon(t) &= y(t) - \boldsymbol{\psi}^T(t) \hat{\boldsymbol{\alpha}}(t-1) \\ \hat{\boldsymbol{\alpha}}(t) &= \hat{\boldsymbol{\alpha}}(t-1) + \mu \boldsymbol{\psi}^*(t) \varepsilon(t) \\ \hat{\boldsymbol{\theta}}(t) &= \sum_{i=1}^k e^{j\omega_i t} \hat{\boldsymbol{\alpha}}_i(t) \end{aligned} \quad (8)$$

where $\mu > 0$ denotes a small stepsize coefficient.

Despite its simplicity the GBF filter can be shown to have similar tracking capabilities as the EWBF filter. The main difference lies in initial convergence, which for the GBF algorithm may be very slow [8].

B. Unknown Frequencies

Even though the EWBF filter is robust to small local changes in frequencies, i.e., changes around the known nominal values $\omega_1^o, \dots, \omega_k^o$, it will fail to identify the system correctly in the presence of a frequency drift [8]. We will adopt a very simple gradient search strategy to derive the frequency-adaptive version of the EWBF algorithm.

Denote by

$$J(t, \boldsymbol{\omega}) = \frac{1}{2} |\varepsilon(t, \boldsymbol{\omega})|^2 \quad (9)$$

where $\boldsymbol{\omega} = [\omega_1, \dots, \omega_k]^T$, the instantaneous measure of fit. The gradient algorithm for minimization of (9) can be expressed in the form

$$\hat{\boldsymbol{\omega}}(t+1) = \hat{\boldsymbol{\omega}}(t) - \eta J'(\hat{\boldsymbol{\omega}}(t)) \quad (10)$$

where $J'(\hat{\boldsymbol{\omega}}(t))$ denotes derivative of $J(t, \boldsymbol{\omega})$ with respect to $\boldsymbol{\omega}$, evaluated at the point $\hat{\boldsymbol{\omega}}(t)$, and $\eta > 0$ is a small adaptation gain.

Observe that in the general time-varying case

$$\varepsilon(t) = y(t) - \boldsymbol{\psi}^T(t) \hat{\boldsymbol{\alpha}}(t-1) = y(t) - \sum_{i=1}^k \boldsymbol{\psi}_i^T(t) \hat{\boldsymbol{\alpha}}_i(t-1)$$

where

$$\boldsymbol{\psi}_i(t) = e^{j \sum_{s=1}^t \omega_i(s)} \boldsymbol{\varphi}(t).$$

Therefore

$$\begin{aligned} J'_i(\hat{\boldsymbol{\omega}}(t)) &= \left. \frac{\partial J(t, \boldsymbol{\omega})}{\partial \omega_i(t)} \right|_{\boldsymbol{\omega}(s) = \hat{\boldsymbol{\omega}}(s), s=1, \dots, t} \\ &= \text{Re} \left\{ j \varepsilon(t) e^{-j \sum_{s=1}^t \hat{\omega}_i(s)} \boldsymbol{\varphi}^H(t) \hat{\boldsymbol{\alpha}}_i^*(t-1) \right\} \\ &= \text{Im} \left\{ \varepsilon^*(t) \hat{\boldsymbol{\psi}}_i^T(t) \hat{\boldsymbol{\alpha}}_i(t-1) \right\} \end{aligned}$$

$$\text{where } \hat{\boldsymbol{\psi}}_i(t) = \hat{f}_i(t) \boldsymbol{\varphi}(t), \quad \hat{f}_i(t) = e^{j \hat{\omega}_i(t)} \hat{f}_i(t-1).$$

Using the notation introduced above, the frequency-adaptive version of the EWBF algorithm (6) can be summarized as follows:

$$\begin{aligned} \hat{f}_i(t) &= e^{j \hat{\omega}_i(t)} \hat{f}_i(t-1) \\ \hat{\boldsymbol{\psi}}_i(t) &= \hat{f}_i(t) \boldsymbol{\varphi}(t), \\ & \quad i=1, \dots, k \\ \hat{\boldsymbol{\psi}}(t) &= \left[\hat{\boldsymbol{\psi}}_1^T(t), \dots, \hat{\boldsymbol{\psi}}_k^T(t) \right]^T \\ \varepsilon(t) &= y(t) - \hat{\boldsymbol{\psi}}^T(t) \hat{\boldsymbol{\alpha}}(t-1) \\ g_i(t) &= \text{Im} \left\{ \varepsilon^*(t) \hat{\boldsymbol{\psi}}_i^T(t) \hat{\boldsymbol{\alpha}}_i(t-1) \right\} \\ \hat{\omega}_i(t+1) &= \hat{\omega}_i(t) - \eta g_i(t), \\ & \quad i=1, \dots, k \\ \mathbf{k}(t) &= \frac{\mathbf{Q}(t-1) \hat{\boldsymbol{\psi}}(t)}{\lambda + \hat{\boldsymbol{\psi}}^H(t) \mathbf{Q}(t-1) \hat{\boldsymbol{\psi}}(t)} \\ \hat{\boldsymbol{\alpha}}(t) &= \hat{\boldsymbol{\alpha}}(t-1) + \mathbf{k}^*(t) \varepsilon(t) \\ \mathbf{Q}(t) &= \frac{1}{\lambda} \left[\mathbf{Q}(t-1) - \mathbf{k}(t) \hat{\boldsymbol{\psi}}^H(t) \mathbf{Q}(t-1) \right] \\ \hat{\boldsymbol{\theta}}(t) &= \sum_{i=1}^k \hat{f}_i(t) \hat{\boldsymbol{\alpha}}_i(t). \end{aligned} \quad (11)$$

The frequency-adaptive version of the GBF algorithm (8) can be written down in the form

$$\begin{aligned}
 \hat{f}_i(t) &= e^{j\hat{\omega}_i(t)} \hat{f}_i(t-1) \\
 \hat{\psi}_i(t) &= \hat{f}_i(t) \boldsymbol{\varphi}(t), \\
 & i = 1, \dots, k \\
 \varepsilon(t) &= y(t) - \sum_{i=1}^k \hat{\psi}_i^T(t) \hat{\boldsymbol{\alpha}}_i(t-1) \\
 g_i(t) &= \text{Im} \left\{ \varepsilon^*(t) \hat{\psi}_i^T(t) \hat{\boldsymbol{\alpha}}_i(t-1) \right\} \\
 \hat{\omega}_i(t+1) &= \hat{\omega}_i(t) - \eta g_i(t) \\
 \hat{\boldsymbol{\alpha}}_i(t) &= \hat{\boldsymbol{\alpha}}_i(t-1) + \mu \hat{\psi}_i^*(t) \varepsilon(t), \\
 & i = 1, \dots, k \\
 \hat{\boldsymbol{\theta}}(t) &= \sum_{i=1}^k \hat{f}_i(t) \hat{\boldsymbol{\alpha}}_i(t). \tag{12}
 \end{aligned}$$

To smooth out frequency estimates, the instantaneous measure of fit (9) can be replaced with the following averaged measure of fit

$$\bar{J}(t, \boldsymbol{\omega}) = \frac{1}{M} \sum_{n=t-M+1}^t J(n, \boldsymbol{\omega}) \tag{13}$$

where $M > 1$. All that one needs to do to incorporate this change is replace the "instantaneous" gradient term $g_i(t)$ in (11) or (12) with the averaged gradient term $\bar{g}_i(t) = (1/M) \sum_{n=t-M+1}^t g_i(n)$.

To our best knowledge the algorithms (11) and (12) are the first working gradient-based frequency-adaptive EWBF/GBF procedures for system identification. Derivations of the two gradient algorithms proposed earlier by Tsatsanis and Giannakis [1] and Bakkoury *et al.* [3] were based on the assumption that $\boldsymbol{\psi}_i(t) = \boldsymbol{\varphi}(t) e^{j\omega_i t}$, i.e., that the estimated frequencies are unknown but *constant*. For this reason both algorithms are not capable of tracking time-varying frequencies unless some heuristic modifications are introduced.

A different method of frequency tracking, based on the recursive prediction error (RPE) approach, was proposed in [9]. Even though the computational complexity of gradient algorithms is much lower than the complexity of the RPE algorithms, the corresponding filters have similar tracking capabilities. The price paid for reduced computational complexity of gradient algorithms is their poor initial convergence. First, when the starting values $\hat{\omega}_1(0), \dots, \hat{\omega}_k(0)$ are too far from the true values, the gradient-based algorithm may not "lock" on the true frequencies. Second, even if the convergence takes place, the convergence rate is usually much slower than that of the comparable RPE algorithm. In the next section we will show how both drawbacks can be eliminated using the nonparametric identification technique.

III. INITIALIZATION

In this section we will show that all initial conditions needed to start (or restart) the EWBF or GBF algorithm can be inferred from the nonparametric [discrete Fourier transform

(DFT)-based] system identification results, obtained for a short startup fragment of the input–output data.

A. Selection of Frequencies

Consider the case where the frequencies are constant in the time interval $T = [1, \dots, N]$ covering the first N samples. For constant frequencies it is reasonable to set $\lambda = 1$, so that the estimation is carried out without forgetting. In this case the EWBF estimator turns into the BF (basis function) estimator for which it holds

$$\begin{aligned}
 \hat{\boldsymbol{\alpha}}(N) &= \mathbf{R}^{-1}(N) \mathbf{p}(N) \\
 \mathbf{R}(N) &= \sum_{t=1}^N \boldsymbol{\psi}^*(t) \boldsymbol{\psi}^T(t), \quad \mathbf{p}(N) = \sum_{t=1}^N y(t) \boldsymbol{\psi}^*(t) \\
 \boldsymbol{\psi}(t) &= [e^{j\omega_1 t} \boldsymbol{\varphi}(t), \dots, e^{j\omega_k t} \boldsymbol{\varphi}(t)]^T. \tag{14}
 \end{aligned}$$

Based on (14), the estimate of the time-varying parameter trajectory of the identified system can be obtained from

$$\hat{\boldsymbol{\theta}}(t|N) = \sum_{i=1}^k e^{j\omega_i t} \hat{\boldsymbol{\alpha}}_i(N), \quad t \in T. \tag{15}$$

The residual sum of squares associated with (15) can be expressed in the form

$$\begin{aligned}
 \rho(k) &= \sum_{t=1}^N |y(t) - \boldsymbol{\varphi}^T(t) \hat{\boldsymbol{\theta}}(t|N)|^2 \\
 &= \sum_{t=1}^N |y(t)|^2 - \mathbf{p}^H(N) \mathbf{R}^{-1}(N) \mathbf{p}(N). \tag{16}
 \end{aligned}$$

In order to arrive at DFT approximations we will assume that the frequencies, forming the set $\Omega = \{\omega_1, \dots, \omega_k\}$, are constrained to a grid of N equidistant frequencies $\Omega_N = \{2\pi i/N, i = 0, \dots, N-1\}$, but neither the number of frequency components k nor their exact location is known a priori. The grid constraint will be relaxed later on. Additionally we will assume that the input sequence is wide-sense stationary with positive-definite covariance matrix $\boldsymbol{\Phi} = \text{E}[\boldsymbol{\varphi}^*(t) \boldsymbol{\varphi}^T(t)] > 0$.

Note that in the case considered, the basis functions are mutually orthogonal in the interval T , that is, for any selection of $\omega_1, \dots, \omega_k$ from Ω_N , it holds that

$$\mathbf{F}(N) = \sum_{t=1}^N \mathbf{f}^*(t) \mathbf{f}^T(t) = N \mathbf{I}_k. \tag{17}$$

Therefore

$$\begin{aligned}
 \bar{\mathbf{R}}(N) &= \mathbf{F}(N) \otimes \boldsymbol{\Phi} = N \text{diag}_k \{\boldsymbol{\Phi}, \dots, \boldsymbol{\Phi}\} \\
 &\cong N \text{diag}_k \{\hat{\boldsymbol{\Phi}}(N), \dots, \hat{\boldsymbol{\Phi}}(N)\}
 \end{aligned}$$

and [see (7)]

$$\tilde{\boldsymbol{\alpha}}_i(N) \cong \frac{1}{N} \hat{\boldsymbol{\Phi}}^{-1}(N) \mathbf{p}_i(N), \quad i = 1, \dots, k \tag{18}$$

where

$$\mathbf{p}_i(N) = \sum_{t=1}^N e^{-j\omega_i t} y(t) \boldsymbol{\varphi}^*(t)$$

and $\widehat{\boldsymbol{\Phi}}(N)$ denotes the sample estimate of $\boldsymbol{\Phi}$

$$\widehat{\boldsymbol{\Phi}}(N) = \frac{1}{N} \sum_{t=1}^N \boldsymbol{\varphi}^*(t) \boldsymbol{\varphi}^T(t). \quad (19)$$

When the input signal is known to be white, as in the channel equalization application, one can set $\widehat{\boldsymbol{\Phi}}(N) = \widehat{\sigma}_u^2(N) \mathbf{I}_n$, where $\widehat{\sigma}_u^2(N) = (1/N) \sum_{t=1}^N |u(t)|^2$.

The quantities $\mathbf{p}_i(N)$, $\omega_i \in \Omega_N$ can be evaluated using the discrete Fourier transform. Let

$$q_l(\omega_i) = \sum_{t=1}^N e^{-j\omega_i t} y(t) u^*(t-l+1).$$

Observe that $\mathbf{p}_i(N) = [q_1(\omega_i), \dots, q_n(\omega_i)]^T$ and

$$\left\{ q_l \left(\frac{2\pi i}{N} \right), i = 0, \dots, N-1 \right\} \\ = \text{DFT} \{ y(1)u^*(2-l), \dots, y(N)u^*(N-l+1) \} \quad l=1, \dots, n.$$

If N is the power of two, then the DFT can be efficiently computed using a fast Fourier transform (FFT).

Note that the quantity $q_l(\omega_i)$ can be regarded an estimate of the cyclic cross-correlation coefficient of $y(t)$ and $u(t)$

$$R_{yu}(\omega_i, l) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N e^{-j\omega_i t} r_{yu}(t, l)$$

where

$$r_{yu}(t, l) = \mathbb{E}[y(t)u^*(t-l+1)].$$

Cyclic cross-correlation is a standard tool for analysis of cyclostationary systems—see, e.g., Giannakis [10].

The quantity

$$\widehat{P}(\omega_i) = N \widetilde{\boldsymbol{\alpha}}_i^H(N) \widehat{\boldsymbol{\Phi}}(N) \widetilde{\boldsymbol{\alpha}}_i(N) \\ = \frac{1}{N} \mathbf{p}_i^H(N) \widehat{\boldsymbol{\Phi}}^{-1}(N) \mathbf{p}_i(N), \quad \omega_i \in \Omega_N \quad (20)$$

will be further referred to as generalized periodogram, or system periodogram. The name stems from the fact that (20) can be considered generalization, to the system case, of the classical concept of signal periodogram. Actually, observe that in the signal case, where $n = 1$ and $u(t) = 1, \forall t$, that (20) reduces to

$$\widehat{P}(\omega_i) = \frac{1}{N} |Y(j\omega_i)|^2$$

where

$$\left\{ Y \left(j \left(\frac{2\pi i}{N} \right) \right), i=0, \dots, N-1 \right\} = \text{DFT} \{ y(1), \dots, y(N) \}$$

i.e., the generalized periodogram becomes identical with the classical (signal) periodogram.

According to (7) and (18), the sum of generalized periodogram values

$$\sum_{i=1}^k \widehat{P}(\omega_i) = \frac{1}{N} \sum_{i=1}^k \mathbf{p}_i^H(N) \widehat{\boldsymbol{\Phi}}^{-1}(N) \mathbf{p}_i(N)$$

can be regarded an estimate of the quantity

$$\mathbf{p}^H(N) \mathbf{R}^{-1}(N) \mathbf{p}(N).$$

Hence, to minimize the residual sum of squares (16), i.e., to maximize the least squares fit, one should pick frequencies $\omega_1, \dots, \omega_k$ which correspond to the k largest values of system periodogram.

The system periodogram is a statistic which combines in a meaningful way information contained in the entire set of cyclic cross-correlation functions $q_1(\omega), \dots, q_n(\omega)$. Combined analysis is needed, otherwise some of the frequency components may be not detected. Note, for example, that for uncorrelated input the condition $a_{li} = 0$ entails $q_l(\omega_i) \cong 0$. In a case like this the frequency ω_i cannot be detected by examining the cyclic cross-correlation spectrum $|q_l(\omega)|^2$. However, when at least one of the coefficients $a_{li}, l = 1, \dots, n$ differs from zero, joint analysis of the spectra $|q_1(\omega)|^2, \dots, |q_n(\omega)|^2$ should reveal the presence of ω_i . System periodogram allows one to perform such analysis.

B. Selection of the Number of Frequency Modes

We will show that the problem of selection of the number of system frequencies can be solved using the Akaike's model order selection criterion. Although the Akaike information criterion (AIC) [11] is better known than its forerunner, the final prediction error (FPE) criterion [12], it is the latter that seems to better fit our current purposes. By final prediction error Akaike meant the steady-state variance of the one-step-ahead prediction error evaluated for the identified stationary autoregressive process. Quite obviously, this is not an adequate measure of fit for a nonstationary finite impulse response system, analyzed in this paper.

Denote by $\widetilde{Y}(N) = \{\widetilde{y}(1), \dots, \widetilde{y}(N)\}$ another set of measurements, *independent* of the set $Y(N) = \{y(1), \dots, y(N)\}$ that was used for parameter estimation. We will assume that the sequences $\widetilde{y}(t)$ and $y(t)$ arise from the identified system under the same experimental conditions, that is, they are obtained for the same input sequence $u(t)$ but different (mutually independent) noise sequences $\widetilde{v}(t)$ and $v(t)$, respectively.

To evaluate predictive capability of the model (15) in the analysis interval $T = [1, \dots, N]$, we will use the following measure of fit:

$$\sigma(k) = \mathbb{E}_{\widetilde{Y}} \left[\sum_{t=1}^N | \widetilde{y}(t) - \boldsymbol{\varphi}^T(t) \widehat{\boldsymbol{\theta}}(t|N) |^2 \right] \quad (21)$$

where the expectation is taken over $\{\tilde{y}(t), t \leq N\}$, i.e., over $\{\tilde{v}(t), t \leq N\}$. Note that this measure fulfills the basic principle of model evaluation: the model should be tested on a different data set than that used for determination of its parameters.

Denote by $\Omega_o \subset \Omega_N$ the set of frequencies corresponding to the “true” system model. One can show that, when $\Omega_o \subset \Omega$ —that is, when the true system model of order k_o belongs to the considered model class—it holds that (see Appendix 1)

$$E_Y[\sigma(k)] = N\sigma_v^2 \left(1 + \frac{kn}{N}\right). \quad (22)$$

Note that when the model is overparametrized ($k > k_o$) its predictive capability decreases—this is the price paid for estimation of superfluous parameters, which are in fact zero.

If the model is not underparametrized, one obtains (see Appendix 2)

$$E_Y[\rho(k)] = N\sigma_v^2 \left(1 - \frac{kn}{N}\right). \quad (23)$$

Combining (22) with (23), one arrives at the following unbiased estimate of $\sigma(k)$:

$$\hat{\sigma}(k) = \rho(N) \frac{1 + \frac{kn}{N}}{1 - \frac{kn}{N}} \quad (24)$$

which can be easily recognized as the Akaike’s final prediction error statistics. Even though for underparametrized model structures (24) loses its interpretation as an estimate of the “final prediction error,” it still reflects predictive ability of the evaluated models—see, e.g., Söderström and Stoica [7].

The residual sum of squares can be approximated with

$$\hat{\rho}(k) = \sum_{t=1}^N |y(t) - \varphi^T(t)\tilde{\theta}(t|N)|^2 \quad (25)$$

where

$$\tilde{\theta}(t|N) = \sum_{i=1}^k e^{j\omega_i t} \tilde{\alpha}_i(N), \quad t \in T. \quad (26)$$

This, finally, leads to the following criterion of model structure selection: choose the model that minimizes

$$\text{FPE}(k) = \hat{\rho}(k) \frac{1 + \frac{kn}{N}}{1 - \frac{kn}{N}}. \quad (27)$$

Remark 1: It is known that neither the FPE nor the (asymptotically equivalent) AIC estimates of the model order are consistent: when Akaike’s criteria are used the probability of model overfitting does not vanish to zero when $N \mapsto \infty$ —see, e.g., [7]. Since we are primarily interested in selecting the model with the best predictive capability, the property well reflected by the FPE statistics, inconsistency is a minor problem, if any.

Remark 2: So far the problem of detection and estimation of system frequencies has been solved using the methods of cyclostationary analysis.

Tsatsanis and Giannakis [1] proposed a solution based on joint analysis of the cyclic second-order and fourth-order cumulants of the output signal. The remarkable feature of this solution is that it does not require knowledge of the input sequence. On the negative side, the method described in [1] is fairly complicated, as it relies on simultaneous detection of pairs of re-

lated peaks in two cyclic spectra. Additionally, under certain circumstances it may suffer from nonidentifiability and nondetectability problems.

Bakkoury *et al.* [3] suggested a different framework, which bears some resemblance to our approach. It is based on localization of significant peaks in the cyclic cross-correlation spectra of system inputs and outputs. Unfortunately, no comparison can be made since the description in [3] is rather sketchy (the proposed frequency-selection procedure was not described in detail).

In both approaches mentioned above, frequency detection is based on cyclostationarity tests (see, e.g., [14]). The problem with such tests is that they are not directly related to any agreeable performance measure. In contrast with this, in our approach it is the predictive capability of the model that decides whether a specific frequency component should or should not be incorporated.

C. Refinements

If the true frequencies fall in between the grid nodes, the system periodogram becomes diffused. This means that not just one but several adjacent periodogram values can be attributed to a single frequency component. Quite obviously, in a situation like this, instead of examining the highest periodogram values one should apply the discrimination procedure to the highest periodogram peaks ($\hat{P}(\omega_i)$ is a periodogram peak if $\hat{P}(\omega_i \pm 2\pi/N) < \hat{P}(\omega_i)$).

The aim of the second modification is to provide more accurate frequency and amplitude estimates. The simplest way of achieving this is by means of zero-padding. When DFT is applied to a sequence extended with zeros, system periodogram can be evaluated and searched for local maxima on a finer grid of frequencies. For example, when the original sequence of length $N = 256$ is extended with $N_o = 768$ zeros, so that the new length is $N = 1024$, frequency quantization errors, which limit accuracy of the DFT-based estimation procedure, can be reduced by a factor of four. It should be stressed, however, that one is not allowed to extend the AIC-based selection rule to the zero-padded (higher resolution) periodogram. This is because for $N_o > 0$ the one-to-one correspondence between DFT and the method of least squares does not hold any more. All decisions concerning the number of system frequencies should be based on the unpadded periodogram.

Note that when zero padding is used to obtain the corrected frequency estimates $\tilde{\omega}_1, \dots, \tilde{\omega}_k$, it holds that

$$[\mathbf{F}(N)]_{mn} = \sum_{s=1}^N \delta_{mn}^s = \begin{cases} \frac{\delta_{mn}^{N+1} - \delta_{mn}}{N} & \text{for } m \neq n \\ \frac{\delta_{mn}^{N+1} - \delta_{mn}}{N-1} & \text{for } m = n \end{cases} \quad (28)$$

where $\delta_{mn} = e^{j(\tilde{\omega}_n - \tilde{\omega}_m)}$. It turns out that noticeable improvements can be achieved in the startup phase of the EWBF algorithm if (28), rather than (17), is used to initialize the matrix \mathbf{Q} in (11) (see below).

So far we have been assuming that signal frequencies and amplitudes can be regarded constant in the startup interval covering the first N input–output samples. In practice this is only approximately true, i.e., it is more realistic to assume that all quantities are slowly time-varying also in the startup interval. It is known

that, when applied to a nonstationary system, the least squares estimator based on N data points introduces an estimation delay of $(N+1)/2$ samples [6]. For even N this means that $\hat{\boldsymbol{\omega}}(N)$ and $\hat{\boldsymbol{\alpha}}(N)$ can be viewed as estimates of $\boldsymbol{\omega}(N/2)$ and $\boldsymbol{\alpha}(N/2)$, or of $\boldsymbol{\omega}(N/2+1)$ and $\boldsymbol{\alpha}(N/2+1)$, rather than of $\boldsymbol{\omega}(N)$ and $\boldsymbol{\alpha}(N)$, respectively. For this reason, to further reduce transients, it may be worthwhile to start the EWBF (GBF) algorithm at the instant $N/2$, rather than at the instant N .

D. Initialization Procedure

After introducing corrections described above, the initialization procedure can be summarized as follows:

1. Compute the sample estimate $\hat{\boldsymbol{\Phi}}(N)$ of the covariance matrix $\boldsymbol{\Phi}$.
2. Use DFT to compute system periodogram $\hat{P}(2\pi i/N)$, $i = 0, \dots, N-1$.
3. Pick the frequencies $\omega_i \in \Omega_N$, $i = 1, 2, \dots$, in the order of decreasing periodogram peaks, so that ω_1 corresponds to the maximum periodogram peak, ω_2 corresponds to the second maximum peak etc.
4. Select the number of frequency components k by minimizing the FPE statistics.
5. Use zero-padding to obtain the corrected frequency estimates $\tilde{\omega}_1, \dots, \tilde{\omega}_k$ (search the zero-padded periodogram for the maxima localized in the close neighborhood of the previously selected frequencies) and the corresponding DFT values: $\tilde{\mathbf{p}}_i(N) = [q_1(\tilde{\omega}_i), \dots, q_n(\tilde{\omega}_i)]$, $i = 1, \dots, k$.
6. Set initial conditions:
 $\tilde{\omega}_i(N/2) = \tilde{\omega}_i$, $\hat{f}_i(N/2) = e^{j\tilde{\omega}_i N/2}$, $i = 1, \dots, k$
 $\hat{\boldsymbol{\alpha}}_i(N/2) = 1/N \hat{\boldsymbol{\Phi}}^{-1}(N) \tilde{\mathbf{p}}_i(N)$, $i = 1, \dots, k$
 and (for the EWBF algorithm only):
 $\mathbf{Q}(N/2) = \mathbf{F}^{-1}(N) \otimes \hat{\boldsymbol{\Phi}}^{-1}(N)$.

Note that the same procedure can be used to restart the EWBF/GBF algorithm each time a frequency jump occurs and/or a new frequency component emerges. Both events can be detected by monitoring prediction errors $\varepsilon(t)$.

Remark: When the extended sequence length $N+N_o$ is a multiple of the basic sequence length N , the unpadded DFT, evaluated at Step 2, is a subset of the zero-padded DFT, required at Step 5. In cases like this there is no need to evaluate DFT twice.

E. Bandwidth Matching Conditions

1) *EWBF Algorithm:* To shed more light on the operation of the EWBF algorithm (6), consider the case where there is only one, time-invariant frequency component ω_i . If the input signal is wide-sense stationary and persistently exciting it holds that [see (7)]

$$\begin{aligned} \bar{\boldsymbol{\theta}}(t) &= \mathbb{E}[\hat{\boldsymbol{\theta}}(t)] \cong \mathbb{E}[\tilde{\boldsymbol{\theta}}(t)] \\ &= \frac{1-\lambda}{1-\lambda^t} e^{j\omega_i t} \boldsymbol{\Phi}^{-1} \mathbb{E} \left[\sum_{s=1}^t \lambda^{t-s} y(s) \boldsymbol{\psi}^*(s) \right] \end{aligned}$$

where the expectation is carried over $\{\boldsymbol{\varphi}(s), v(s), s \leq t\}$.

Since, in the case considered, $y(s) = \boldsymbol{\varphi}^T(s) \boldsymbol{\theta}(s) + v(s)$ and $\boldsymbol{\psi}(s) = e^{j\omega_i s} \boldsymbol{\varphi}(s)$, one obtains

$$\bar{\boldsymbol{\theta}}(t) \cong \frac{1-\lambda}{1-\lambda^t} \sum_{s=1}^t \lambda^{t-s} e^{j\omega_i(t-s)} \boldsymbol{\theta}(s)$$

and asymptotically, for sufficiently large t

$$\bar{\boldsymbol{\theta}}(t) \cong T_{\text{EWBF}}(q^{-1}) \boldsymbol{\theta}(t) \quad (29)$$

where

$$T_{\text{EWBF}}(q^{-1}) = \frac{1-\lambda}{1-\lambda e^{j\omega_i} q^{-1}}.$$

According to (29), the EWBF algorithm behaves, on the average, as a narrow-band extraction filter centered at the frequency ω_i . When the forgetting constant λ is close to one, the 3 dB bandwidth of this filter can be approximately expressed as

$$B_{\text{EWBF}} \cong 2(1-\lambda).$$

Observe that the mean parameter estimation error is given by

$$\boldsymbol{\theta}(t) - \bar{\boldsymbol{\theta}}(t) \cong (1 - T_{\text{EWBF}}(q^{-1})) \boldsymbol{\theta}(t) = N_{\text{EWBF}}(q^{-1}) \boldsymbol{\theta}(t)$$

where the filter

$$N_{\text{EWBF}}(q^{-1}) = \frac{\lambda(1 - e^{j\omega_i} q^{-1})}{1 - \lambda e^{j\omega_i} q^{-1}}$$

can be easily recognized as the notch filter centered at the frequency ω_i . This allows one to consider the EWBF algorithm a generalized notch filter.

2) *GBF Algorithm:* Analysis of the mean behavior of the GBF algorithm (8) resembles treatment of a classical least mean square (LMS) algorithm. Suppose that regression vectors form an independent identically distributed (i.i.d.) sequence (we will comment on this constraint later on) and that only one periodic component is present. Then it is straightforward to show that

$$\mathbb{E}[\hat{\boldsymbol{\alpha}}(t)] = (\mathbf{I}_n - \mu \boldsymbol{\Phi}) \mathbb{E}[\hat{\boldsymbol{\alpha}}(t-1)] + \mu e^{-j\omega_i t} \boldsymbol{\Phi} \boldsymbol{\theta}(t)$$

which leads to

$$\bar{\boldsymbol{\theta}}(t) = e^{j\omega_i t} (\mathbf{I}_n - \mu \boldsymbol{\Phi}) \bar{\boldsymbol{\theta}}(t-1) + \mu \boldsymbol{\Phi} \boldsymbol{\theta}(t). \quad (30)$$

To guarantee mean stability of the GBF algorithm, the stepsize μ must be chosen so as to fulfill the (well-known) constraint

$$\mu < \frac{1}{\lambda_{\max}(\boldsymbol{\Phi})}$$

where $\lambda_{\max}(\boldsymbol{\Phi})$ denotes the largest eigenvalue of $\boldsymbol{\Phi}$.

The i.i.d. assumption, imposed earlier on $\boldsymbol{\varphi}(t)$, is very restrictive. Using bounding techniques developed for the purpose of LMS analysis (see, e.g., [13]) one can show that (30) remains approximately true for m -dependent regressors (the process $x(t)$ is called m -dependent if for all t the sequences $\{x(i), i \leq t\}$ and $\{x(i), i \geq t+m\}$ are independent). Note that for white input signal the sequence of regression vectors in (1) is n -dependent, which means that (30) extends to the channel identification case.

Denote by \mathbf{D} a unitary matrix converting $\boldsymbol{\Phi}$ into a diagonal form

$$\mathbf{D}^H \mathbf{D} = \mathbf{D} \mathbf{D}^H = \mathbf{I}_n, \quad \mathbf{D}^H \boldsymbol{\Phi} \mathbf{D} = \boldsymbol{\Lambda}$$

where $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1(\boldsymbol{\Phi}), \dots, \lambda_n(\boldsymbol{\Phi})\}$ is a diagonal matrix made up of the eigenvalues of $\boldsymbol{\Phi}$.

Setting $\boldsymbol{\gamma}(t) = \mathbf{D}^H \boldsymbol{\theta}(t)$ and $\bar{\boldsymbol{\gamma}}(t) = \mathbf{D}^H \bar{\boldsymbol{\theta}}(t)$, one obtains the equation shown at the bottom of the next page or, equivalently

$$\bar{\boldsymbol{\theta}}(t) = \mathbf{D} \mathbf{T}_{\text{GBF}}(q^{-1}) \mathbf{D}^H \boldsymbol{\theta}(t). \quad (31)$$

Note that $\mathbf{T}_{\text{GBF}}(q^{-1})$ is a collection of notch filters with the same center frequency ω_i but with different bandwidths. Unlike the EWBF case, the bandwidths of the component filters depend not only on the user-defined quantities (stepsize μ) but also on the input data related factors (eigenvalues of the regression matrix Φ). This is a clear disadvantage of the GBF algorithm, especially in cases where the eigenvalue disparity index of Φ is large.

When the components of the regression vector are mutually uncorrelated, as in the channel identification application, it holds that $\Phi = \sigma_u^2 \mathbf{I}_n$. Since all eigenvalues of Φ are in this case identical and equal to σ_u^2 , one can rewrite (31) in a simpler form

$$\bar{\boldsymbol{\theta}}(t) = T_{\text{GBF}}(q^{-1})\boldsymbol{\theta}(t) \quad (32)$$

where

$$T_{\text{GBF}}(q^{-1}) = \frac{\mu\sigma_u^2}{1 - (1 - \mu\sigma_u^2)e^{j\omega_i}q^{-1}}, \quad B_{\text{GBF}} \cong 2\mu\sigma_u^2.$$

Note that in the special case discussed above notch filters associated with the GBF and EWBF algorithms become identical after setting $\lambda = 1 - \mu\sigma_u^2$. Finally, note that when the eigenvalues of Φ differ, the average 3 dB bandwidth of the filter bank $\mathbf{T}_{\text{GBF}}(q^{-1})$ can be expressed in the form

$$\bar{B}_{\text{GBF}} \cong \frac{1}{n} \sum_{i=1}^n 2\mu\lambda_i(\Phi) = \frac{2\mu}{n} \text{tr}\{\Phi\} = 2\mu\sigma_u^2 = B_{\text{GBF}}.$$

3) *Matching Conditions:* When the analysis carried out above is repeated for the BF algorithm (14), trimmed to the single frequency case, one obtains

$$\bar{\boldsymbol{\theta}}(N) \cong T_{\text{BF}}(q^{-1})\boldsymbol{\theta}(N) \quad (33)$$

where

$$T_{\text{BF}}(q^{-1}) = \frac{1 - e^{j\omega_i N}q^{-N}}{N(1 - e^{j\omega_i}q^{-1})}.$$

Similarly as $T_{\text{EWBF}}(q^{-1})$ and $T_{\text{GBF}}(q^{-1})$, the filter $T_{\text{BF}}(q^{-1})$ is centered at ω_i and narrow band. For sufficiently large N , its 3 dB bandwidth can be obtained from

$$B_{\text{BF}} \cong \frac{5.6}{N}.$$

Since $T_{\text{BF}}(q^{-1})$ is ‘‘switched’’ to $T_{\text{EWBF}}(q^{-1})$ (or to $T_{\text{GBF}}(q^{-1})$) in the initial phase of identification, it is reasonable to require that both filters have identical bandwidths. Setting $B_{\text{BF}} = B_{\text{EWBF}}$, one arrives at the following bandwidth matching conditions:

$$\lambda \cong 1 - \frac{2.8}{N}, \quad N \cong \frac{2.8}{1 - \lambda} \quad (34)$$

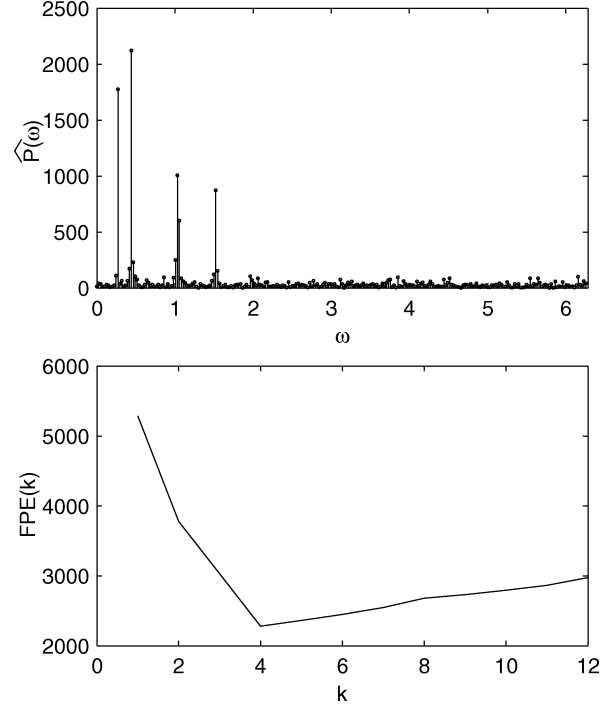


Fig. 1. Estimation of the number of basis frequencies using the nonparametric approach: (upper plot) the generalized periodogram and (lower plot) the FPE statistic.

which guarantee smooth transition from one filter to another at the switching point. The analogous conditions for the GBF algorithm can be obtained after setting $B_{\text{BF}} = \bar{B}_{\text{GBF}}$.

For the EWBF algorithm, the analysis carried out above can be easily extended to the multiple frequencies case provided that

$$|\omega_i - \omega_j| > \frac{1}{2} B_{\text{EWBF}} \cong \frac{1}{2} B_{\text{BF}}, \quad \forall i \neq j$$

i.e., provided that all frequencies are separable by the corresponding extraction filters. In the presence of eigenvalue disparity the analogous separability conditions for the GBF algorithm are more demanding.

IV. SIMULATION RESULTS

Figs. 1 and 2 show typical results obtained for a simulated time-varying communication channel with two impulse response coefficients $\theta_1(t)$ and $\theta_2(t)$ ($n = 2$), each of which was modeled as a linear combination of four complex exponentials ($k = 4$). The weighting coefficients in (2) had constant values

$$\begin{aligned} \boldsymbol{\alpha} &= [a_{11}, \dots, a_{41}, a_{12}, \dots, a_{42}] \\ &= [j-1, 1-j, -1+0.5j, 1-0.5j, 1, 0.5-1.5j, 1.2j+1, 1]. \end{aligned}$$

$$\begin{aligned} \bar{\boldsymbol{\gamma}}(t) &= e^{j\omega_i}(\mathbf{I}_n - \mu\boldsymbol{\Lambda})\bar{\boldsymbol{\gamma}}(t-1) + \mu\boldsymbol{\Lambda}\boldsymbol{\gamma}(t) = \mathbf{T}_{\text{GBF}}(q^{-1})\boldsymbol{\gamma}(t) \\ \mathbf{T}_{\text{GBF}}(q^{-1})\boldsymbol{\gamma}(t) &= \text{diag}\left\{ \frac{\mu\lambda_1(\Phi)}{1 - (1 - \mu\lambda_1(\Phi))e^{j\omega_1}q^{-1}}, \dots, \frac{\mu\lambda_n(\Phi)}{1 - (1 - \mu\lambda_n(\Phi))e^{j\omega_n}q^{-1}} \right\} \end{aligned}$$

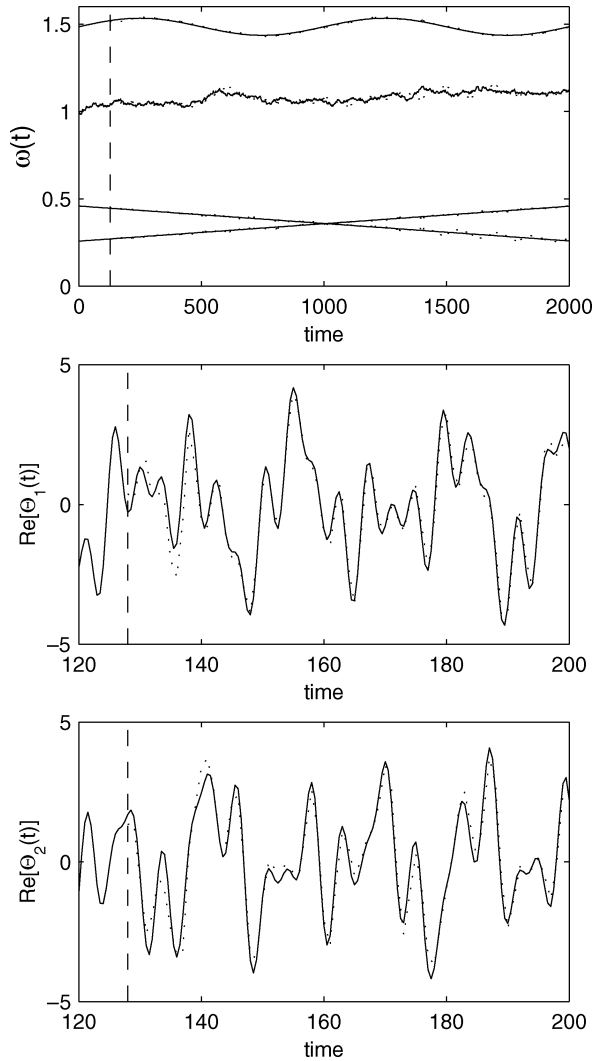


Fig. 2. (Upper plot) Frequency estimates and real parts of (middle plot) $\hat{\theta}_1(t)$ and (lower plot) $\hat{\theta}_2(t)$ yielded by the EWBF algorithm initialized at instant $t = 128$ in the way described in this paper. Solid lines depict true values and dotted lines show evolution of the corresponding estimates. Broken vertical lines show the moment of initialization of the recursive identification algorithm.

The white 4-QAM sequence was used as the input signal ($u(t) = \pm 1 \pm j$, $\sigma_u^2 = 2$) and the disturbance $v(t)$ was white and Gaussian with variance $\sigma_v^2 = 0.4$.

Fig. 1 shows the system periodogram, obtained for the first 256 input/output samples (upper plot) and the corresponding FPE statistic (lower plot). Quite clearly, generalized periodogram is a powerful tool that gives interesting insights into the structure of the investigated system. The FPE criterion indicates that there are four significant system frequency components, which is actually true.

Fig. 2 shows what happens when the adaptive EWBF algorithm (11) ($\lambda = 0.99$, $\eta = 0.0002$) is initialized using the results of nonparametric, periodogram-based system analysis ($N = 256$, $N_o = 768$). Note that when the EWBF filter is started in this way, its response is practically free of initialization transients. More importantly, simulation experiments indicate that without frequency preestimation, the adaptive EWBF algorithm on most occasions fails to lock onto correct frequencies. The

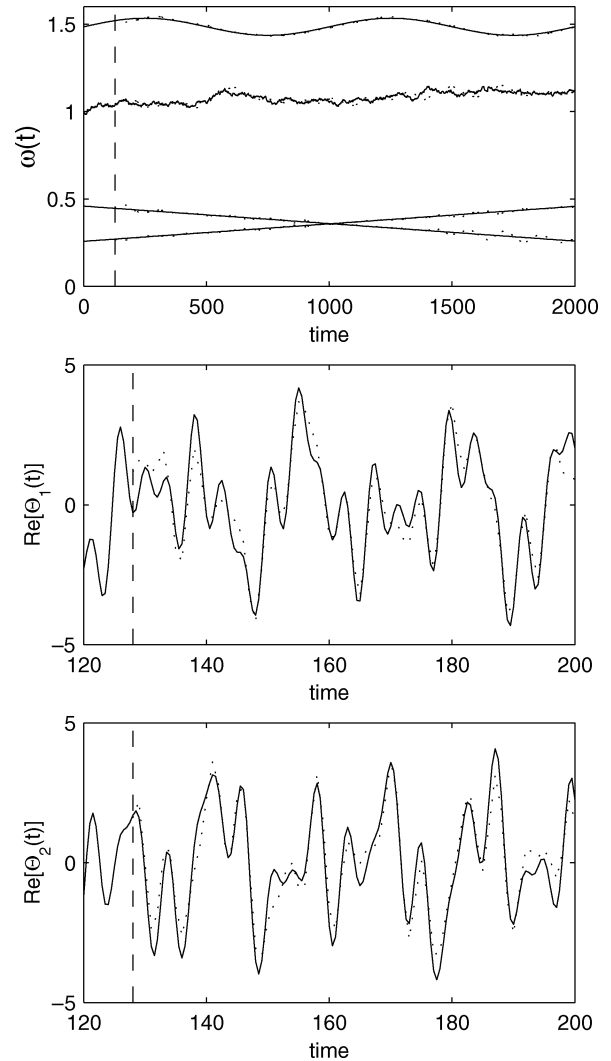


Fig. 3. (Upper plot) Frequency estimates and real parts of (middle plot) $\hat{\theta}_1(t)$ and (lower plot) $\hat{\theta}_2(t)$ yielded by the GBF algorithm initialized at instant $t = 128$ in the way described in this paper. Solid lines depict true values and dotted lines show evolution of the corresponding estimates. Broken vertical lines show the moment of initialization of the recursive identification algorithm.

analogous results obtained for the GBF algorithm ($\mu = 0.005$, $\eta = 0.0002$) are depicted in Fig. 3.

To test efficiency of the FPE-based model order selection rule, we performed tests for the system described above, subject to 100 different realizations of the measurement noise sequence and four different levels of the noise variance (in all cases the same input sequence was used). The results, gathered in Table I, confirm good properties of the proposed method. The tendency of FPE to overestimate the model order is not strongly emphasized unless the signal-to-noise ratio becomes very small.

Fig. 4 summarizes results of a Monte Carlo experiment, arranged to compare parameter tracking capabilities of the EWBF and GBF algorithms. According to the plots, showing the mean-squared norms of parameter estimation errors

$$\begin{aligned} E[\|\Delta\theta(t)\|^2] &= E[\|\hat{\theta}(t) - \theta(t)\|^2] \\ &= E[\|\hat{\theta}_1(t) - \theta_1(t)\|^2] + E[\|\hat{\theta}_2(t) - \theta_2(t)\|^2] \end{aligned}$$

TABLE I
NUMBER OF FREQUENCIES SELECTED BY THE FPE RULE FOR 100 REALIZATIONS OF THE NOISE SEQUENCE AND FOUR DIFFERENT VALUES OF NOISE VARIANCE

noise variance	number of frequencies			
	k=4	k=5	k=6	k=7
$\sigma_v^2=0.01$	96	4	0	0
$\sigma_v^2=0.1$	98	2	0	0
$\sigma_v^2=1$	98	2	0	0
$\sigma_v^2=10$	55	29	13	3

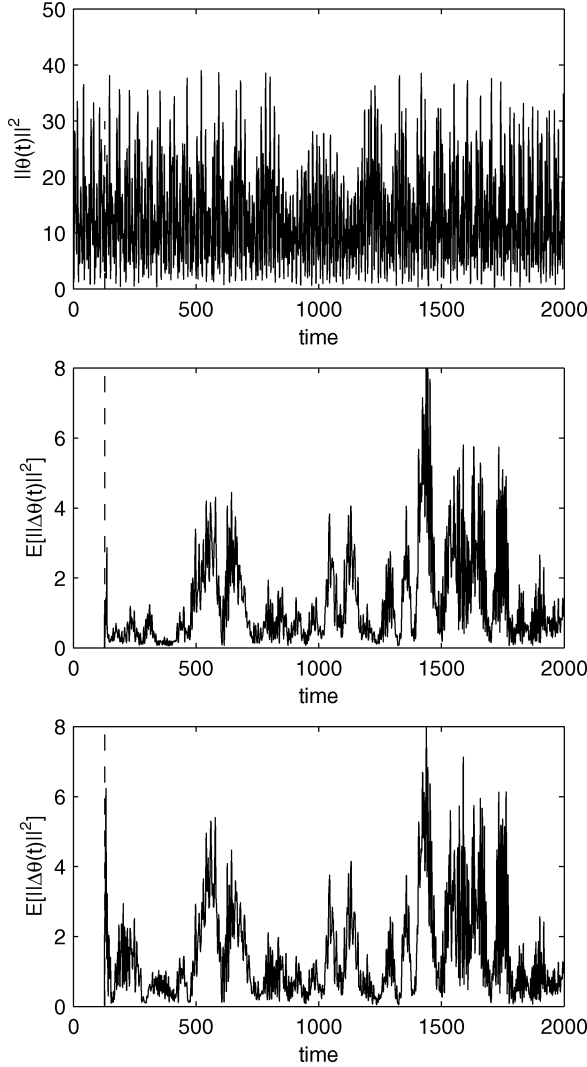


Fig. 4. Comparison of parameter tracking capabilities (evolution of the mean squared norms of parameter estimation errors) of the (middle plot) EWBF algorithm and (lower plot) GBF algorithm. The upper plot shows evolution of the squared norm of the true parameter vector. The middle and lower plots were obtained by averaging results of 100 simulation runs corresponding to different noise sequences and a fixed input sequence.

the compared algorithms yield almost identical results (the EWBF algorithm is slightly better in the startup phase). This is hardly a surprise since under the white noise excitation all eigenvalues of Φ are identical. When the eigenvalue disparity index of Φ is large ($\lambda_{\max}(\Phi)/\lambda_{\min}(\Phi) \gg 1$), the differences between EWBF and GBF become more pronounced. The error plots were obtained by averaging results of 100 simulation runs

(corresponding to different realizations of the noise sequence). As a reference, the plot of the squared norm of the true parameter vector $\|\theta(t)\|^2$ is also shown in Fig. 4.

V. CONCLUSION

The problem of identification of quasi-periodically varying finite impulse response systems was considered. Given that the number of system frequency modes is known a priori, and that good initial estimates of system frequencies are available, this problem can be solved using the model-based algorithms called generalized adaptive notch filters. We have shown that the number of frequency modes, as well as all initial conditions needed to smoothly start, or restart, generalized adaptive notch filters, can be inferred from nonparametric DFT-based analysis of a short startup fragment of input–output data. Such mixed nonparametric/parametric approach preserves advantages of both treatments. The resulting estimation procedure combines good tracking capabilities, typical of model-based solutions, with global frequency search, typical of DFT-based approach. Additionally, it allows one to almost completely eliminate startup transients.

APPENDIX I
DERIVATION OF (22)

Since we assumed that the model is not underparametrized, it holds that

$$\tilde{y}(t) = \varphi^T(t)\theta(t) + \tilde{v}(t) = \psi^T(t)\alpha + \tilde{v}(t)$$

where α denotes the vector of true system coefficients (when $k > k_o$ some of these coefficients are zero) and the noise sequence $\tilde{v}(t)$ is independent of $v(t)$, and hence independent of $\hat{\alpha}(N)$. Therefore $\tilde{y}(t) - \varphi^T(t)\hat{\theta}(t|N) = \psi^T(t)(\alpha - \hat{\alpha}(N)) + \tilde{v}(t)$ and

$$\sigma(k) = N\sigma_v^2 + (\hat{\alpha}(N) - \alpha)^H \mathbf{R}(N)(\hat{\alpha}(N) - \alpha). \quad (35)$$

Straightforward calculations lead to $\hat{\alpha}(N) = \alpha + \mathbf{R}^{-1}(N)\mathbf{q}(N)$ where $\mathbf{q}(N) = \sum_{t=1}^N v(t)\psi^*(t)$. Since the noise sequence is white it holds that $E[\mathbf{q}(N)\mathbf{q}^H(N)] = \sigma_v^2 \mathbf{R}(N)$, where the expectation is carried over $\{v(t), t \leq N\}$. Therefore

$$\begin{aligned} \text{cov}[\hat{\alpha}(N)] &= \mathbf{R}^{-1}(N)E[\mathbf{q}(N)\mathbf{q}^H(N)]\mathbf{R}^{-1}(N) \\ &= \sigma_v^2 \mathbf{R}^{-1}(N). \end{aligned}$$

Combining the last result with (35), one obtains

$$\begin{aligned} E[\sigma(k)] &= N\sigma_v^2 + \text{tr}\{\mathbf{R}(N)\text{cov}[\hat{\alpha}(N)]\} \\ &= N\sigma_v^2 + \sigma_v^2 \text{tr}\{\mathbf{I}_{kn}\} = N\sigma_v^2 \left(1 + \frac{kn}{N}\right). \end{aligned}$$

APPENDIX II
DERIVATION OF (24)

Observe that

$$\begin{aligned} y(t) - \varphi^T(t)\hat{\theta}(t|N) &= \psi^T(t)(\alpha - \hat{\alpha}(N)) + v(t) \\ &= v(t) - \psi^T(t)\mathbf{R}^{-1}(N)\mathbf{q}(N) \end{aligned}$$

which leads to

$$\begin{aligned} \rho(k) &= \sum_{t=1}^N |v(t)|^2 - 2\mathbf{q}^H(N)\mathbf{R}^{-1}(N) \sum_{t=1}^N v(t)\boldsymbol{\psi}^*(t) \\ &\quad + \mathbf{q}^H(N)\mathbf{R}^{-1}(N) \left(\sum_{t=1}^N \boldsymbol{\psi}^*(t)\boldsymbol{\psi}^T(t) \right) \mathbf{R}^{-1}(N)\mathbf{q}(N) \\ &= \sum_{t=1}^N |v(t)|^2 - \mathbf{q}^H(N)\mathbf{R}^{-1}(N)\mathbf{q}(N). \end{aligned}$$

Therefore

$$\begin{aligned} E[\rho(k)] &= N\sigma_v^2 - \text{tr}\{\mathbf{R}^{-1}(N)E[\mathbf{q}(N)\mathbf{q}^H(N)]\} \\ &= N\sigma_v^2 - \sigma_v^2 \text{tr}\{\mathbf{I}_{kn}\} = N\sigma_v^2 \left(1 - \frac{kn}{N}\right). \end{aligned}$$

REFERENCES

- [1] M. K. Tsatsanis and G. B. Giannakis, "Modeling and equalization of rapidly fading channels," *Int. J. Adapt. Contr. Signal Process.*, vol. 10, pp. 159–176, 1996.
- [2] G. B. Giannakis and C. Tepedelenlioglu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proc. IEEE*, vol. 86, pp. 1969–1986, 1998.
- [3] J. Bakkoury, D. Roviras, M. Ghogho, and F. Castanie, "Adaptive MLSE receiver over rapidly fading channels," *Signal Process.*, vol. 80, pp. 1347–1360, 2000.
- [4] P. Tichavský and A. Nehorai, "Comparative study of four adaptive frequency trackers," *IEEE Trans. Signal Process.*, vol. 45, pp. 1473–1484, 1997.
- [5] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1997.
- [6] M. Niedźwiecki, *Identification of Time-Varying Processes*. New York: Wiley, 2000.
- [7] T. Söderström and P. Stoica, *System Identification*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
- [8] M. Niedźwiecki and T. Kłaput, "Fast algorithms for identification of periodically varying systems," *IEEE Trans. Signal Process.*, vol. 51, pp. 3270–3279, 2003.
- [9] M. Niedźwiecki and P. Kaczmarek, "Estimation and tracking of quasiperiodically varying processes," in *Proc. 13th IFAC Symp. System Identification*, Rotterdam, The Netherlands, 2003, pp. 1102–1107.
- [10] G. B. Giannakis, "Cyclostationary signal analysis," in *The Digital Signal Processing Handbook*. Boca Raton, FL: CRC Press, 1998, pp. 17–1–17–31.
- [11] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, pp. 716–723, 1974.
- [12] —, "Statistical predictor identification," *Ann. Inst. Statist. Math.*, vol. 22, pp. 203–217, 1970.
- [13] O. Macchi, *Adaptive Processing*. New York: Wiley, 1995.
- [14] A. V. Dandawate and G. B. Giannakis, "Statistical tests for presence of cyclostationarity," *IEEE Trans. Signal Process.*, vol. 42, pp. 2355–2369, 1994.



Maciej Niedźwiecki was born in Poznań, Poland, in 1953. He received the M.Sc. and Ph.D. degrees from Gdańsk University of Technology, Gdańsk, Poland, and the Dr.Hab. (D.Sc.) degree from the Technical University of Warsaw, Warsaw, Poland, in 1977, 1981, and 1991, respectively.

He spent three years as a Research Fellow with the Department of Systems Engineering, Australian National University (1986–1989). From 1990 to 1993, he was a Vice Chairman of the Technical Committee on Theory, International Federation of Automatic Control (IFAC). He is the author of *Identification of Time-varying Processes* (New York: Wiley, 2000). He is a Professor and Head of the Department of Automatic Control, Faculty of Electronics, Telecommunications and Computer Science, Gdańsk University of Technology. His main areas of research interests include system identification, signal processing, and adaptive systems.



Piotr Kaczmarek received the M.Sc. degree in automatic control from Gdańsk University of Technology, Gdańsk, Poland, in 2000, where he is currently pursuing the Ph.D. degree. He graduated from the European Master Degree Course in Control and Management of Lean Manufacturing in Network Systems conducted in cooperation between Gdańsk University of Technology, Catholic University of Louvain, Belgium, and University of Karlsruhe, Germany.

Since 1998 he has been involved in many projects for the biggest Polish air conditioning systems manufacturers. His interests include system identification and adaptive filtering as well as optimization of production techniques.