Postprint of: Niedźwiecki M., Meller M., Generalized adaptive comb filters/smoothers and their application to the identification of quasi-periodically varying systems and signals, Automatica, Vol. 49, Iss. 6 (2013), pp. 1601-1613, DOI: 10.1016/j.automatica.2013.02.037

Generalized adaptive comb filters/smoothers and their application to identification of quasi-periodically varying systems and signals *

Maciej Niedźwiecki [a], Michał Meller [a]

[a] *Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control, Gdańsk University of Technology, Narutowicza 11/12, 80-233 Gdańsk , Poland, Tel: + 48 58 3472519; fax: +48 58 3415821*

Abstract

The problem of both causal and noncausal identification of linear stochastic systems with quasi-harmonically varying parameters is considered. The quasi-harmonic description allows one to model nonsinusoidal quasi-periodic parameter changes. The proposed identification algorithms are called generalized adaptive comb filters/smoothers because in the special signal case they reduce down to adaptive comb algorithms used to enhance or suppress nonstationary harmonic signals embedded in noise. The paper presents a thorough statistical analysis of generalized adaptive comb algorithms, and demonstrates their statistical efficiency in the case where the fundamental frequency of parameter changes varies slowly with time according to the integrated random-walk model.

Key words: system identification, time-varying processes

1 Introduction

1.1 Problem statement

We will consider the problem of identification of quasi-periodically varying complex-valued systems governed by

$$y(t) = \sum_{i=1}^{n} \theta_i(t) u(t-i+1) + v(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t) + v(t) \tag{1}$$

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

$$\begin{aligned} \boldsymbol{\theta}(t) &= \sum_{k=1}^{K} \boldsymbol{\beta}_k(t) e^{j\sum_{i=1}^{t}\omega_k(i)} \\ \boldsymbol{\beta}_k(t) &= [b_{k1}(t), \ldots, b_{kn}(t)]^{\mathrm{T}} \\ b_{ki}(t) &= a_{ki}(t) e^{j\nu_{ki}},\ i = 1, \ldots, n. \end{aligned} \tag{2}$$

The following three types of *real-valued* quantities are incorporated in (2): the instantaneous angular frequencies $\omega_k(t)$, the instantaneous amplitudes $a_{ki}(t)$, and the time-invariant phase shifts ν_{ki}. With a slight abuse of terminology, the complex-valued vectors $\boldsymbol{\beta}_k(t)$ will be further referred to as 'complex amplitudes'.

Under certain circumstances (in the presence of several strong reflectors) the model (1)–(2) can be used to describe rapidly fading mobile radio channels (Giannakis & Tepedelenlioğlu, 1998), (Bakkoury et al., 2000). In this case $y(t)$ denotes the sampled baseband signal received by the mobile unit, $\{u(t)\}$ denotes the sequence of transmitted symbols, and $v(t)$ denotes channel noise. We will assume that the frequencies $\omega_k(t)$ are harmonically related, namely

$$\omega_k(t) = m_k \omega_0(t), \quad k = 1, \ldots, K \tag{3}$$

where $\omega_0(t)$ denotes the slowly varying fundamental frequency and m_k are integer numbers. Such multiple frequencies, called harmonics, appear in the Fourier series expansions of periodic signals. For example, if parameter trajectory $\boldsymbol{\theta}(t)$ is periodic with period L, it admits the following Fourier representation: $\boldsymbol{\theta}(t) = \sum_{k=0}^{L-1} \boldsymbol{\beta}_k e^{jk\omega_0 t}$, $\omega_0 = (2\pi)/L$.

The notion of 'time-varying harmonics' can be regarded as a natural extension of the Fourier analysis to quasi-

* This work was supported by the National Science Center.

Email addresses: maciekn@eti.pg.gda.pl (Maciej Niedźwiecki), michal.meller@eti.pg.gda.pl (Michał Meller).

periodically varying systems, such as (1)–(2). The choice of the multipliers $m_k, k = 1, \ldots, K$, depends on our prior knowledge of the system time variation. When all harmonics are expected to be present, one should set $m_k = k$. In the presence of odd harmonics only, the natural choice is $m_k = 2k - 1$, etc.
In the special case where $n = 1$ and $\varphi(t) \equiv 1$, equations (1)–(3) describe a complex-valued harmonic signal $s(t) = \theta(t)$ buried in noise

$$y(t) = s(t) + v(t), \quad s(t) = \sum_{k=1}^{K} b_k(t) e^{j \sum_{i=1}^{t} \omega_k(i)}. \quad (4)$$

The problem of either elimination or extraction of harmonic signals buried in noise can be solved using adaptive comb filters (Nehorai & Porat, 1986), (Regalia, 1995). For this reason the system identification/tracking algorithm described below can be considered a generalized comb filter.

1.2 Contribution

The problem of causal identification (tracking) of single-mode quasi-periodically varying systems was studied in (Niedźwiecki & Kaczmarek, 2004), (Niedźwiecki & Kaczmarek, 2005a) and (Niedźwiecki & Kaczmarek, 2005b).
In the recent conference paper (Niedźwiecki & Meller, 2011a), the results presented earlier were extended to noncausal identification (smoothing). Additionally, a more sophisticated frequency estimation scheme was proposed, incorporating frequency rate tracking/smoothing and yielding better results in practice.
All papers published so far focus on identification of single-mode systems, i.e., systems with parameters that can be modeled as complex sinusoids (cisoids) with slowly varying amplitudes and a slowly varying instantaneous frequency.
This paper extends results presented in (Niedźwiecki & Meller, 2011a) to nonstationary systems with quasi-harmonically varying parameters, i.e., to systems with several frequency modes governed by the same slowly varying fundamental frequency. In practice such harmonic modes of variation often arise in oscillatory systems with nonlinear elements and/or loads (Neimark, 2003).
In principle, quasi-harmonically varying systems can be identified using the multiple-frequency versions of the algorithms mentioned above. Such algorithms are made up of several single-frequency sub-algorithms that work in parallel and are driven by the common prediction error. Since the estimated frequencies are in this case regarded as mutually unrelated quantities, the harmonic structure of the system/signal time variation is not exploited in any way. In this paper we present algorithms that take advantage of such a prior information, i.e., the algorithms that perform a coordinated frequency search. This allows one to improve estimation results considerably.

2 Generalized adaptive notch filter – overview of known results

Suppose that the identified nonstationary system has a single frequency mode ($K = 1$), i.e., it is governed by

$$y(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t) + v(t), \ \boldsymbol{\theta}(t) = \boldsymbol{\beta}(t) e^{j \sum_{i=1}^{t} \omega(i)} \quad (5)$$

where $\boldsymbol{\beta}(t) = [b_1(t), \ldots, b_n(t)]^{\mathrm{T}}$ and $\omega(t) \in (-\pi, \pi]$ are slowly varying quantities. Furthermore, suppose that:

(A1) The measurement noise $\{v(t)\}$ is a zero-mean circular white sequence with variance σ_v^2.

(A2) The sequence of regression vectors $\{\boldsymbol{\varphi}(t)\}$, independent of $\{v(t)\}$, is nondeterministic, wide-sense stationary and ergodic with known correlation matrix [1] $\boldsymbol{\Phi} = \mathrm{E}[\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)]$.

Denote by $\alpha(t) = \omega(t+1) - \omega(t)$ the rate of change of the instantaneous frequency $\omega(t)$. Under assumptions (A1)–(A2), identification of the system (5) can be carried out using the following generalized adaptive notch filtering (GANF) algorithm proposed in (Niedźwiecki & Meller, 2011a)

$$\begin{aligned}
\widehat{f}(t) &= e^{j[\widehat{\omega}(t-1)+\widehat{\alpha}(t-1)]} \widehat{f}(t-1) \\
\varepsilon(t) &= y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{f}(t)\widehat{\boldsymbol{\beta}}(t-1) \\
\widehat{\boldsymbol{\beta}}(t) &= \widehat{\boldsymbol{\beta}}(t-1) + \mu \boldsymbol{\Phi}^{-1} \boldsymbol{\varphi}^*(t) \widehat{f}^*(t) \varepsilon(t) \\
g(t) &= \frac{\mathrm{Im}\left[\varepsilon^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{f}(t)\widehat{\boldsymbol{\beta}}(t-1)\right]}{\widehat{\boldsymbol{\beta}}^{\mathrm{H}}(t-1)\boldsymbol{\Phi}\widehat{\boldsymbol{\beta}}(t-1)} \\
\widehat{\alpha}(t) &= \widehat{\alpha}(t-1) - \gamma_\alpha g(t) \\
\widehat{\omega}(t) &= \widehat{\omega}(t-1) + \widehat{\alpha}(t-1) - \gamma_\omega g(t) \\
\widehat{\boldsymbol{\theta}}(t) &= \widehat{f}(t)\widehat{\boldsymbol{\beta}}(t)
\end{aligned} \quad (6)$$

where $\widehat{f}(t)$ is an estimate of $f(t) = e^{j \sum_{i=1}^{t} \omega(i)}$, and $\mu > 0$, $\gamma_\omega > 0$, $\gamma_\alpha > 0$, such that $\gamma_\alpha \ll \gamma_\omega \ll \mu$, denote small adaptation gains determining the rate of amplitude adaptation, frequency adaptation and frequency rate adaptation, respectively.
The gradient search strategy, incorporated in (6) for the purpose of tracking $\omega(t)$ and $\alpha(t)$, is based on minimization of the following instantaneous measure of fit $J(t) = |\epsilon(t)|^2/2$, where $\epsilon(t) = y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t) f(t) \widehat{\boldsymbol{\beta}}(t-1)$. Note that

$$\begin{aligned}
\frac{\partial J(t)}{\partial \omega(t)} &= \mathrm{Re}\left[\epsilon(t) \, \frac{\partial \epsilon^*(t)}{\partial \omega(t)}\right] \\
&= -\mathrm{Re}\left[j\epsilon(t)\boldsymbol{\varphi}^{\mathrm{H}}(t) f^*(t) \widehat{\boldsymbol{\beta}}^*(t-1)\right] \\
&= \mathrm{Im}\left[\epsilon^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) f(t) \widehat{\boldsymbol{\beta}}(t-1)\right].
\end{aligned} \quad (7)$$

[1] Hereinafter the symbol $*$ will denote complex conjugation, and the symbol H – Hermitian (conjugate) transpose.

Therefore, the term $g(t)$ in (6) can be interpreted as a normalized estimate of the gradient (7). Normalization makes the algorithm scale-invariant. When it is not applied, the tracking properties of the GANF algorithm depend not only on the user-defined adaptation gains γ_ω and γ_α, but also on the system-related variables, which is inconvenient from the practical viewpoint.
Tracking properties of the GANF algorithm (6) were analyzed in (Niedźwiecki & Meller, 2011a) in the case where the vector of 'amplitudes' $\boldsymbol{\beta}(t)$ is unknown but constant:

(A3*) $\boldsymbol{\beta}(t) \equiv \boldsymbol{\beta}$, i.e., $\boldsymbol{\theta}(t) = e^{j\omega(t)}\boldsymbol{\theta}(t-1), \ \ \forall t.$

Note that under (A3*) the normalization term $\widehat{\boldsymbol{\beta}}^{\mathrm{H}}(t-1)\boldsymbol{\Phi}\widehat{\boldsymbol{\beta}}(t-1)$ can be regarded as an estimate of the power of the noiseless system output

$$b^2 = \boldsymbol{\beta}^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta} = \mathrm{E}\left[\boldsymbol{\beta}^{\mathrm{H}}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\beta}\right] = \mathrm{E}\left[|\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t)|^2\right].$$

Using the approximating linear filter (ALF) technique – the stochastic linearization approach proposed in (Tichavský & Händel, 1995) – one can show that the frequency and frequency rate estimation errors $\Delta\widehat{\omega}(t) = \omega(t) - \widehat{\omega}(t)$, $\Delta\widehat{\alpha}(t) = \alpha(t) - \widehat{\alpha}(t)$, can be approximately expressed in the form

$$\Delta\widehat{\omega}(t) \cong G_1(q^{-1})e(t) + G_2(q^{-1})\delta(t) \tag{8}$$

$$\Delta\widehat{\alpha}(t) \cong H_1(q^{-1})e(t) + H_2(q^{-1})\delta(t) \tag{9}$$

where $\{e(t)\}$, $e(t) = -\mathrm{Im}[\boldsymbol{\beta}^{\mathrm{H}}\boldsymbol{\varphi}^*(t)f^*(t)v(t)/b^2]$, is a zero-mean white noise with variance $\sigma_e^2 = \sigma_v^2/(2b^2)$, $\delta(t) = \alpha(t) - \alpha(t-1)$ denotes the one-step change of the frequency rate, and

$$\begin{aligned}
G_1(q^{-1}) &= (1-q^{-1})[\gamma_\omega + (\gamma_\alpha - \gamma_\omega)q^{-1}]/D(q^{-1}) \\
G_2(q^{-1}) &= q^{-1}[1 - \gamma_\omega - (1-\mu)q^{-1}]/D(q^{-1}) \\
H_1(q^{-1}) &= \gamma_\alpha(1-q^{-1})^2/D(q^{-1}) \\
H_2(q^{-1}) &= [1 + (\mu + \gamma_\omega - 2)q^{-1} + (1-\mu)q^{-2}]/D(q^{-1})
\end{aligned}$$

where $D(q^{-1}) = 1 + d_1q^{-1} + d_2q^{-2} + d_3q^{-3}$, $d_1 = \mu + \gamma_\omega + \gamma_\alpha - 3$, $d_2 = 3 - 2\mu - \gamma_\omega$, $d_3 = \mu - 1$. All filters are asymptotically stable if adaptation gains fulfill the following (sufficient) stability conditions: $0 < \mu < 1$, $0 < \gamma_\omega < 1$, $0 < \gamma_\alpha < 1$ and $\mu(\gamma_\omega + \gamma_\alpha) > \gamma_\alpha$.
In spite of its simplicity, the gradient frequency tracking mechanism adopted in (6) has very good statistical properties – as shown in (Niedźwiecki & Meller, 2011a), when the instantaneous frequency drifts according to the Gaussian integrated random-walk (IRW) model, namely

(A4) $\{\delta(t)\}$, independent of $\{v(t)\}$ and $\{\boldsymbol{\varphi}(t)\}$, is a zero-mean white sequence with variance σ_δ^2.

(A5) The sequences $\{v(t)\}$ and $\{\delta(t)\}$ are normally distributed.

the optimally tuned GANF algorithm (6) is statistically efficient, i.e., it reaches the Cramér-Rao-type lower frequency and frequency rate tracking bounds.
Note that $\alpha(t) - \alpha(t-1) = \delta(t)$ implies $(1-q^{-1})^2\omega(t) = \delta(t-1)$. Since $(1-q^{-1})^2\omega(t) = 0$ entails $\omega(t) = \gamma_1 + \gamma_2 t$, where γ_1 and γ_2 denote arbitrary constants, the IRW model can be regarded as a perturbed linear growth/decay model – for small perturbations ($\sigma_\delta \ll \sigma_v/b$) the corresponding frequency changes will be further referred to as quasi-linear.

3 Multiple-frequency GANF

Denote by $y_k(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}_k(t) + v(t)$, where $\boldsymbol{\theta}_k(t) = f_k(t)\boldsymbol{\beta}_k(t)$, and $\widehat{f}_k(t)$ is an estimate of $f_k(t) = e^{j\sum_{i=1}^{t}\omega_k(i)}$, the output of this subsystem of (1) which is associated with the frequency ω_k. If the signals $y_1(t), \ldots, y_K(t)$ were measurable, one could design K independent GANF algorithms of the form (6), each taking care of a particular subsystem. Since it holds that $\boldsymbol{\theta}(t) = \sum_{k=1}^{K}\boldsymbol{\theta}_k(t)$, the final parameter estimate could be easily obtained by combining the partial estimates.
Even though the outputs $y_k(t)$ are not available, one can replace them with the surrogate (estimated) outputs obtained from

$$\widehat{y}_k(t) = y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\sum_{\substack{i=1 \\ i\neq k}}^{K}\widehat{\boldsymbol{\theta}}_i(t|t-1)$$

where $\widehat{\boldsymbol{\theta}}_k(t|t-1) = \widehat{f}_k(t)\widehat{\boldsymbol{\beta}}_k(t-1)$ denotes the one-step-ahead prediction of $\boldsymbol{\theta}_k(t)$. Note that after replacing $y(t)$ with $\widehat{y}_k(t)$ in (6), one obtains $\varepsilon_k(t) = \widehat{y}_k(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}_k(t|t-1) = y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\sum_{k=1}^{K}\widehat{f}_k(t)\widehat{\boldsymbol{\beta}}_k(t-1) = \varepsilon(t)$, $\forall k$, which means that all sub-algorithms should be driven by the same 'global' prediction error. Such an approach was used, with good results, to design multiple-frequency algorithms in (Niedźwiecki & Kaczmarek, 2004). When applied to (6) it yields

$$\begin{aligned}
\widehat{f}_k(t) &= e^{j[\widehat{\omega}_k(t-1) + \widehat{\alpha}_k(t-1)]}\widehat{f}_k(t-1) \\
\varepsilon(t) &= y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\sum_{k=1}^{K}\widehat{f}_k(t)\widehat{\boldsymbol{\beta}}_k(t-1) \\
\widehat{\boldsymbol{\beta}}_k(t) &= \widehat{\boldsymbol{\beta}}_k(t-1) + \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^\star(t)\widehat{f}_k^\star(t)\varepsilon(t) \\
g_k(t) &= \frac{\mathrm{Im}\left[\varepsilon^\star(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{f}_k(t)\widehat{\boldsymbol{\beta}}_k(t-1)\right]}{\widehat{\boldsymbol{\beta}}_k^{\mathrm{H}}(t-1)\boldsymbol{\Phi}\widehat{\boldsymbol{\beta}}_k(t-1)} \\
\widehat{\alpha}_k(t) &= \widehat{\alpha}_k(t-1) - \gamma_\alpha g_k(t) \\
\widehat{\omega}_k(t) &= \widehat{\omega}_k(t-1) + \widehat{\alpha}_k(t-1) - \gamma_\omega g_k(t) \\
\widehat{\boldsymbol{\theta}}_k(t) &= \widehat{f}_k(t)\widehat{\boldsymbol{\beta}}_k(t) \\
k &= 1, \ldots, K \\
\widehat{\boldsymbol{\theta}}(t) &= \sum_{k=1}^{K}\widehat{\boldsymbol{\theta}}_k(t)
\end{aligned} \tag{10}$$

In (Niedźwiecki & Kaczmarek, 2005b) it was shown that the number of frequency modes, as well as all initial conditions needed to smoothly start (start without initialization transients) the GANF algorithm, can be inferred from nonparametric DFT-based analysis of a short startup fragment of the input-output data. The tool that can be used for this purpose was termed generalized (system) periodogram, as in the signal case it reduces to the classical periodogram.

When applied to identification of the system (1)–(3), the GANF algorithm (10) has two serious drawbacks.

First, it does not take into consideration the harmonic structure (3), i.e., the estimated frequencies are regarded as mutually unrelated quantities, while the true harmonics vary in a coordinated way. Hence, even though such an unconstrained multiple-frequency generalized adaptive notch filter can be used to identify the multiharmonic system/signal, its tracking characteristics will be generally inferior to those offered by solutions that incorporate the harmonic constraints.

Second, the algorithm (10) is not robust to incorrect frequency matching. While the strong frequency components, i.e., those characterized by large values of the signal-to-noise ratio $\mathrm{SNR}_k(t) = \| \boldsymbol{\beta}_k(t) \|^2_{\boldsymbol{\Phi}} / \sigma_v^2$ are usually tracked successfully, the weak ones may be difficult to follow – even if the initial frequency assignment is correct, the sub-algorithms tracking such weak components may, after some time, lock onto the neighboring, stronger components, corresponding to higher or lower frequencies. Moreover, when the system/signal is nonstationary, the 'strength' of different harmonic components may vary with time, which further complicates the picture.

4 Generalized adaptive comb filter

In order to arrive at the algorithm which performs coordinated search of the instantaneous fundamental frequency $\omega_0(t)$, one should minimize $J(t)$ for

$$\epsilon(t) = y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t) \sum_{k=1}^{K} f_k(t) \widehat{\boldsymbol{\beta}}_k(t-1)$$

where $f_k(t) = e^{j \sum_{i=1}^{t} \omega_k(i)} = e^{j m_k \sum_{i=1}^{t} \omega_0(i)}$. Note that

$$\begin{aligned} \frac{\partial J(t)}{\partial \omega_0(t)} &= -\mathrm{Re}\left[j \epsilon(t) \boldsymbol{\varphi}^{\mathrm{H}}(t) \sum_{k=1}^{K} m_k f_k^*(t) \widehat{\boldsymbol{\beta}}_k^*(t-1) \right] \\ &= \sum_{k=1}^{K} m_k \, \mathrm{Im}\left[\epsilon^*(t) \boldsymbol{\varphi}^{\mathrm{T}}(t) f_k(t) \widehat{\boldsymbol{\beta}}_k(t-1) \right]. \end{aligned} \tag{11}$$

This leads to the following recursive estimation scheme which will be further referred to as a generalized adaptive comb filter (GACF)

$$\begin{aligned} \widehat{f}_k(t) &= e^{j m_k [\widehat{\omega}_0(t-1) + \widehat{\alpha}_0(t-1)]} \widehat{f}_k(t-1) \\ \varepsilon(t) &= y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t) \sum_{k=1}^{K} \widehat{f}_k(t) \widehat{\boldsymbol{\beta}}_k(t-1) \\ \widehat{\boldsymbol{\beta}}_k(t) &= \widehat{\boldsymbol{\beta}}_k(t-1) + \mu \boldsymbol{\Phi}^{-1} \boldsymbol{\varphi}^{\star}(t) \widehat{f}_k^{\star}(t) \varepsilon(t) \\ \widehat{\boldsymbol{\theta}}_k(t) &= \widehat{f}_k(t) \widehat{\boldsymbol{\beta}}_k(t) \\ k &= 1, \ldots, K \\ g(t) &= \frac{\sum_{k=1}^{K} m_k \, \mathrm{Im}\left[\varepsilon^{\star}(t) \boldsymbol{\varphi}^{\mathrm{T}}(t) \widehat{f}_k(t) \widehat{\boldsymbol{\beta}}_k(t-1) \right]}{\sum_{k=1}^{K} m_k^2 \, \widehat{\boldsymbol{\beta}}_k^{\mathrm{H}}(t-1) \boldsymbol{\Phi} \widehat{\boldsymbol{\beta}}_k(t-1)} \\ \widehat{\alpha}_0(t) &= \widehat{\alpha}_0(t-1) - \gamma_\alpha g(t) \\ \widehat{\omega}_0(t) &= \widehat{\omega}_0(t-1) + \widehat{\alpha}_0(t-1) - \gamma_\omega g(t) \\ \widehat{\boldsymbol{\theta}}(t) &= \sum_{k=1}^{K} \widehat{\boldsymbol{\theta}}_k(t) \end{aligned} \tag{12}$$

One can show, using the ALF technique, that under (A1)–(A2) and the following assumption

(A3) $\boldsymbol{\beta}_k(t) \equiv \boldsymbol{\beta}_k$, i.e., $\boldsymbol{\theta}_k(t) = e^{j\omega_k(t)} \boldsymbol{\theta}_k(t-1), k = 1, \ldots, K, \ \forall t.$

which is a multi-frequency variant of (A3*), the frequency and frequency rate estimation errors $\Delta\widehat{\omega}_0(t) = \omega_0(t) - \widehat{\omega}_0(t)$, $\Delta\widehat{\alpha}_0(t) = \alpha_0(t) - \widehat{\alpha}_0(t)$, can be approximately expressed in the form (see Appendix 1)

$$\Delta\widehat{\omega}_0(t) \cong G_1(q^{-1}) e_0(t) + G_2(q^{-1}) \delta(t) \tag{13}$$

$$\Delta\widehat{\alpha}_0(t) \cong H_1(q^{-1}) e_0(t) + H_2(q^{-1}) \delta(t) \tag{14}$$

where the transfer functions $G_1(q^{-1}), G_2(q^{-1}), H_1(q^{-1})$ and $H_2(q^{-1})$ are identical with those appearing in (8)–(9), and $e_0(t) = \sum_{k=1}^{K} m_k e_k(t)$, $e_k(t) = -\mathrm{Im}\left[\boldsymbol{\beta}_k^{\mathrm{H}} \boldsymbol{\varphi}^*(t) f_k^*(t) v(t) / b_0^2 \right]$, $b_0^2 = \sum_{k=1}^{K} m_k^2 b_k^2$ and $b_k^2 = \boldsymbol{\beta}_k^{\mathrm{H}} \boldsymbol{\Phi} \boldsymbol{\beta}_k$. Note that the normalizing term $\sum_{k=1}^{K} m_k^2 \, \widehat{\boldsymbol{\beta}}_k^{\mathrm{H}}(t-1) \boldsymbol{\Phi} \widehat{\boldsymbol{\beta}}_k(t-1)$, which appears in the expression for $g_k(t)$ in (12), can be regarded as an estimate of b_0^2.

Furthermore, one can show that $\{e_k(t)\}, k = 1, \ldots, K$, are zero-mean white noise sequences with cross-correlation functions given by (see Appendix 2)

$$\mathrm{E}\left[e_k(t) e_l(s) \right] = \begin{cases} \rho_{kl}(t) & \text{for } t = s \\ 0 & \text{for } t \neq s \end{cases} \tag{15}$$

where

$$\begin{aligned} \rho_{kl}(t) = \frac{\sigma_v^2}{2 b_0^4} \big\{ & \mathrm{Re}[\boldsymbol{\beta}_k^{\mathrm{H}} \boldsymbol{\Phi} \boldsymbol{\beta}_l] \cos[\phi_k(t) - \phi_l(t)] \\ & - \mathrm{Im}[\boldsymbol{\beta}_k^{\mathrm{H}} \boldsymbol{\Phi} \boldsymbol{\beta}_l] \sin[\phi_k(t) - \phi_l(t)] \big\} \end{aligned} \tag{16}$$

and $\phi_k(t) = \sum_{i=1}^{t} \omega_k(i) = m_k \sum_{i=1}^{t} \omega_0(i)$.
Setting $l = k$ in (16), one obtains

$$\sigma_{e_k}^2 = \mathrm{E}[|e_k(t)|^2] = \frac{\sigma_v^2 b_k^2}{2b_0^4} \tag{17}$$

where $b_k^2 = \boldsymbol{\beta}_k^{\mathrm{H}} \boldsymbol{\Phi} \boldsymbol{\beta}_k$ is the power of the k-th component of the noiseless system output.
When frequency changes are sufficiently slow, so that the functions $f_k(t), k = 1, \ldots, K$, can be regarded as locally almost periodic, the sequences $\{e_k(t)\}$ are mutually orthogonal in the sense that

$$\langle \rho_{kl}(t) \rangle_T \cong \langle \rho_{kl}(t) | \omega_0(t) \equiv \omega_0 \rangle_\infty = 0, \quad \forall k \neq l$$

where $\langle x(t) \rangle_T = (1/T) \sum_{i=0}^{T-1} x(t-i)$ denotes the local average of $x(t)$, and $T \gg T_0 = 2\pi/\omega_0$. Using this result, one obtains

$$\left\langle \sigma_{e_0}^2(t) \right\rangle_T = \left\langle \mathrm{E}[|e_0(t)|^2] \right\rangle_T \cong \sum_{k=1}^{K} m_k^2 \sigma_{e_k}^2 = \frac{\sigma_v^2}{2b_0^2} \tag{18}$$

Suppose that the assumption (A4) holds true. Then, using standard results from the linear filtering theory, one obtains

$$\begin{aligned} &\left\langle \mathrm{E}\left\{ [\widehat{\omega}_0(t) - \omega_0(t)]^2 \right\} \right\rangle_T = \left\langle \mathrm{E}\left\{ [\Delta\widehat{\omega}_0(t)]^2 \right\} \right\rangle_T \\ &\cong g[G_1(z^{-1})] \left\langle \sigma_{e_0}^2(t) \right\rangle_T + g[G_2(z^{-1})] \sigma_\delta^2 \end{aligned} \tag{19}$$

$$\begin{aligned} &\left\langle \mathrm{E}\left\{ [\widehat{\alpha}_0(t) - \alpha_0(t)]^2 \right\} \right\rangle_T = \left\langle \mathrm{E}\left\{ [\Delta\widehat{\alpha}_0(t)]^2 \right\} \right\rangle_T \\ &\cong g[H_1(z^{-1})] \left\langle \sigma_{e_0}^2(t) \right\rangle_T + g[H_2(z^{-1})] \sigma_\delta^2 \end{aligned} \tag{20}$$

where

$$g[X(z^{-1})] = \frac{1}{2\pi j} \oint X(z^{-1}) X(z) \frac{dz}{z}$$

is an integral evaluated along the unit circle in the z-plane and $X(z)$ denotes any stable proper rational transfer function.
The first term on the right hand side of (19) constitutes the variance component of the mean-squared frequency estimation error, and the second term – its bias component. The same remark applies to (20). According to (18), the variance components in (19) and (20) are inversely proportional to the quantity which will be further referred to as effective signal-to-noise ratio (ESNR)

$$\mathrm{ESNR} = \frac{b_0^2}{\sigma_v^2} = \frac{\sum_{k=1}^{K} m_k^2 b_k^2}{\sigma_v^2}$$

and which differs from the signal-to-noise ratio defined as

$$\mathrm{SNR} = \frac{\left\langle \mathrm{E}[|\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t)|^2] \right\rangle_T}{\sigma_v^2} \cong \frac{\sum_{k=1}^{K} b_k^2}{\sigma_v^2} \leq \mathrm{ESNR}$$

We note that equations (13)–(14) are identical with those derived earlier for the single frequency case – the only change needed to move from (8)–(9) to (13)–(14) is replacement of the noiseless output power b^2, which appears in the expression for σ_e^2, with the effective output power b_0^2, which appears in the expression for $< \sigma_{e_0}^2(t) >_T$. Since in the multi-frequency case ($K > 1$) it holds that $b_0^2 > \sum_{k=1}^{K} b_k^2$, the variance components of the mean-squared fundamental frequency and frequency rate tracking errors are smaller than the analogous errors observed, under the same SNR, in the single frequency case. This increased (compared to the unconstrained frequency estimation case) accuracy bonus is available due to incorporation in the estimation process prior information about the harmonic structure of the identified quasi-periodic phenomenon. The same qualitative effect can be observed in time-invariant frequency estimation schemes, such as the ones described in (Nehorai & Porat, 1986) and (James et al., 1994).

5 Generalized adaptive comb smoother

The important consequence of the fact that the approximate error equations (13)–(14) are identical with those derived in (Niedźwiecki & Meller, 2011a) for systems with a single frequency mode of parameter variation, is that the smoothing technique proposed there is directly applicable to the multiple-frequency case. Following (Niedźwiecki & Meller, 2011a), suppose that a prerecorded data block $\Omega(N) = \{y(i), \boldsymbol{\varphi}(i), i = 1, \ldots, N\}$ of length N is available, which is typical of off-line applications, i.e., those based on parameter/signal reconstruction, rather than tracking. The smoothed estimates of $\omega_0(t)$, $\alpha_0(t)$ and $\boldsymbol{\theta}(t)$, based on $\Omega(N)$, will be denoted by $\widetilde{\omega}_0(t)$, $\widetilde{\alpha}_0(t)$ and $\widetilde{\boldsymbol{\theta}}(t)$, respectively [2].
To obtain smoothed estimates, one can use a cascade of postprocessing filters derived in (Niedźwiecki & Meller, 2011a) and (Niedźwiecki & Meller, 2011b). The proposed fixed-interval generalized adaptive comb smoothing (GACS) procedure, listed in Table 1, is six-step

Step 1: The preliminary estimates $\widehat{\omega}_0(t)$ and $\widehat{\alpha}_0(t)$ are obtained using the pilot algorithm based on (12).

Step 2: To obtain the smoothed frequency rate estimates $\widetilde{\alpha}_0(t)$, the trajectory $\{\widehat{\alpha}_0(t), t = 1, \ldots, N\}$ is filtered, backward in time, using the anticausal filter $S(q)$: $\widetilde{\alpha}_0(t) = S(q)\widehat{\alpha}_0(t)$, where

$$S(q) = 1 - (1-q)H_2(q) = \gamma_\alpha q / D(q).$$

Step 3: To obtain the smoothed frequency estimates $\widetilde{\omega}_0(t)$, the trajectory $\{\widehat{\omega}_0(t), t = 1, \ldots, N\}$ is processed

[2] In the Kalman filtering/smoothing literature, the filtered and smoothed estimates of the state vector $\mathbf{x}(t)$ are usually denoted by $\widehat{\mathbf{x}}(t|t)$ and $\widehat{\mathbf{x}}(t|N)$, respectively. In this paper a different notation is used to avoid false associations.

using the noncausal filter $S(q)T(q^{-1})$, where

$$T(q^{-1}) = \frac{a_1 q^{-1}}{1 + a_2 q^{-1}}\ , \quad a_1 = \frac{\gamma_\alpha}{\gamma_\omega}\ , \quad a_2 = \frac{\gamma_\alpha - \gamma_\omega}{\gamma_\omega}\ .$$

This can be achieved by means of backward-time processing of the prefiltered trajectory: $\widetilde{\omega}_0(t) = q^{-1}S(q)\bar{\omega}_0(t)$, where $\bar{\omega}_0(t) = qT(q^{-1})\widehat{\omega}_0(t)$.

Step 4: The amplitude coefficients are re-estimated using the frequency-guided version of the pilot algorithm, obtained by replacing in (12) the causal frequency estimates $\widehat{\omega}_0(t)$ with their noncausal (smoothed) counterparts $\widetilde{\omega}_0(t)$, evaluated at Step 3.

Step 5: To obtain the smoothed amplitude estimates $\widetilde{\boldsymbol{\beta}}_k(t), k = 1, \ldots, K$, the re-estimated amplitude trajectories $\{\bar{\boldsymbol{\beta}}_k(t), t = 1, \ldots, N\}$ are filtered, backward in time, using the anticausal filter $F(q)$: $\widetilde{\boldsymbol{\beta}}_k(t) = F(q)\bar{\boldsymbol{\beta}}_k(t), k = 1, \ldots, K$, where $F(q) = \mu/[1 - (1-\mu)q]$.

Step 6: To obtain the smoothed partial parameter estimates $\widetilde{\boldsymbol{\theta}}_k(t)$, the smoothed amplitude estimates $\widetilde{\boldsymbol{\beta}}_k(t)$ are combined with the smoothed phase estimates $\widetilde{\phi}_k(t) = m_k \sum_{i=1}^{t} \widetilde{\omega}_0(t)$. The smoothed parameter estimate $\widetilde{\boldsymbol{\theta}}(t)$ is evaluated as a sum of its harmonic components.

Remark: We note that Step 2 above is optional – if all that is needed is estimation of $\boldsymbol{\theta}(t)$, the frequency rate smoothing part of the algorithm can be skipped.

Denote by $\Delta\widetilde{\omega}_0(t) = \omega_0(t) - \widetilde{\omega}_0(t)$ and $\Delta\widetilde{\alpha}_0(t) = \alpha_0(t) - \widetilde{\alpha}_0(t)$ the frequency and frequency rate smoothing errors, respectively. Under assumptions (A1) – (A3), the approximate error equations can be obtained in the form

$$\Delta\widetilde{\omega}_0(t) \cong I_1(q^{-1})e_0(t) + I_2(q^{-1})\delta(t) \qquad (21)$$

$$\Delta\widetilde{\alpha}_0(t) \cong J_1(q^{-1})e_0(t) + J_2(q^{-1})\delta(t) \qquad (22)$$

where

$$\begin{aligned} I_1(q^{-1}) &= S(q)T(q^{-1})G_1(q^{-1}) \\ I_2(q^{-1}) &= \frac{1 - S^2(q)S^2(q^{-1})}{1 - q^{-1}} \\ J_1(q^{-1}) &= S(q)H_1(q^{-1}) \\ J_2(q^{-1}) &= \frac{1 - S(q)S(q^{-1})}{1 - q^{-1}}. \end{aligned}$$

Denote by $X^+(q^{-1}) = \big[X(q^{-1})X(q)\big]_+$ the stable factor of a rational transfer function $X(q^{-1})X(q)$. Based on (21) - (22), the mean-squared estimation errors can be evaluated in a way similar to (19) - (20)

$$\begin{aligned} &\big\langle \mathrm{E}\left\{[\widetilde{\omega}_0(t) - \omega_0(t)]^2\right\}\big\rangle_T = \big\langle \mathrm{E}\left\{[\Delta\widetilde{\omega}_0(t)]^2\right\}\big\rangle_T \\ &\cong g[I_1^+(z^{-1})]\left\langle \sigma_{e_0}^2(t)\right\rangle_T + g[I_2^+(z^{-1})]\sigma_\delta^2 \end{aligned} \qquad (23)$$

$$\begin{aligned} &\big\langle \mathrm{E}\left\{[\widetilde{\alpha}_0(t) - \alpha_0(t)]^2\right\}\big\rangle_T = \big\langle \mathrm{E}\left\{[\Delta\widetilde{\alpha}_0(t)]^2\right\}\big\rangle_T \\ &\cong g[J_1^+(z^{-1})]\left\langle \sigma_{e_0}^2(t)\right\rangle_T + g[J_2^+(z^{-1})]\sigma_\delta^2. \end{aligned} \qquad (24)$$

Table 1. Generalized adaptive comb smoother

pilot filter :

$$\begin{aligned} \widehat{f}_k(t) &= e^{jm_k[\widehat{\omega}_0(t-1) + \widehat{\alpha}_0(t-1)]}\widehat{f}_k(t-1) \\ \varepsilon(t) &= y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\textstyle\sum_{k=1}^{K}\widehat{f}_k(t)\widehat{\boldsymbol{\beta}}_k(t-1) \\ \widehat{\boldsymbol{\beta}}_k(t) &= \widehat{\boldsymbol{\beta}}_k(t-1) + \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^{\star}(t)\widehat{f}_k^{\star}(t)\varepsilon(t) \\ g(t) &= \frac{\sum_{k=1}^{K} m_k\ \mathrm{Im}[\varepsilon^{\star}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{f}_k(t)\widehat{\boldsymbol{\beta}}_k(t-1)]}{\sum_{k=1}^{K} m_k^2\ \widehat{\boldsymbol{\beta}}_k^{\mathrm{H}}(t-1)\boldsymbol{\Phi}\widehat{\boldsymbol{\beta}}_k(t-1)} \\ \widehat{\alpha}_0(t) &= \widehat{\alpha}_0(t-1) - \gamma_\alpha g(t) \\ \widehat{\omega}_0(t) &= \widehat{\omega}_0(t-1) + \widehat{\alpha}_0(t-1) - \gamma_\omega g(t) \\ t &= 1, \ldots, N, \quad k = 1, \ldots, K \end{aligned}$$

frequency rate smoother [optional] :

$$\begin{aligned} [\ \widetilde{\alpha}_0(N) &= \widetilde{\alpha}_0(N+1) = \widetilde{\alpha}_0(N+2) = \widehat{\alpha}_0(N)\ ] \\ \widetilde{\alpha}_0(t) &= -d_1\widetilde{\alpha}_0(t+1) - d_2\widetilde{\alpha}_0(t+2) \\ &\quad - d_3\widetilde{\alpha}_0(t+3) + \gamma_\alpha\widehat{\alpha}_0(t+1) \\ t &= N-1, \ldots, 1 \end{aligned}$$

frequency smoother :

$$\begin{aligned} [\ \bar{\omega}_0(0) &= \widehat{\omega}_0(1)\ ] \\ \bar{\omega}_0(t) &= -a_2\bar{\omega}_0(t-1) + a_1\widehat{\omega}_0(t) \\ t &= 1, \ldots, N \end{aligned}$$

$$\begin{aligned} [\ \widetilde{\omega}_0(N+1) &= \widetilde{\omega}_0(N+2) = \widetilde{\omega}_0(N+3) = \widehat{\omega}_0(N)\ ] \\ \widetilde{\omega}_0(t) &= -d_1\widetilde{\omega}_0(t+1) - d_2\widetilde{\omega}_0(t+2) \\ &\quad - d_3\widetilde{\omega}_0(t+3) + \gamma_\alpha\bar{\omega}_0(t) \\ t &= N, \ldots, 1 \end{aligned}$$

frequency−guided filter :

$$\begin{aligned} \widetilde{f}_k(t) &= e^{jm_k\widetilde{\omega}_0(t)}\widetilde{f}_k(t-1) \\ \bar{\varepsilon}(t) &= y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\textstyle\sum_{k=1}^{K}\widetilde{f}_k(t)\bar{\boldsymbol{\beta}}_k(t-1) \\ \bar{\boldsymbol{\beta}}_k(t) &= \bar{\boldsymbol{\beta}}_k(t-1) + \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^{*}(t)\widetilde{f}_k^{*}(t)\bar{\varepsilon}(t) \\ t &= 1, \ldots, N, \quad k = 1, \ldots, K \end{aligned}$$

amplitude smoother :

$$\begin{aligned} [\ \widetilde{\boldsymbol{\beta}}_k(N+1) &= \bar{\boldsymbol{\beta}}_k(N)\ ] \\ \widetilde{\boldsymbol{\beta}}_k(t) &= (1-\mu)\widetilde{\boldsymbol{\beta}}_k(t+1) + \mu\bar{\boldsymbol{\beta}}_k(t) \\ t &= N, \ldots, 1, \quad k = 1, \ldots, K \end{aligned}$$

output filter :

$$\begin{aligned} \widetilde{\boldsymbol{\theta}}_k(t) &= \widetilde{f}_k(t)\widetilde{\boldsymbol{\beta}}_k(t), \quad \widetilde{\boldsymbol{\theta}}(t) = \textstyle\sum_{k=1}^{K}\widetilde{\boldsymbol{\theta}}_k(t) \\ t &= 1, \ldots, N \end{aligned}$$

6 Optimization and Cramér-Rao bounds

Consider a system (1)–(3) with pseudo-linear frequency changes. In order to achieve the best tracking/smoothing results, the adaptation gains of the GACF/GACS algorithms should be chosen so as to trade-off the bias and variance components in (19)–(20) and (21)–(22). Such optimal settings depend *exclusively* on the balance be-

tween the bias and variance error components, determined by the scalar coefficient

$$\kappa = \frac{\mathrm{E}[w^2(t)]}{2\left\langle \mathrm{E}[e_0^2(t)]\right\rangle_T} \cong \frac{b_0^2\sigma_\delta^2}{\sigma_v^2} = \mathrm{ESNR}\cdot\sigma_\delta^2$$

further referred to as the rate of nonstationarity of the analyzed system [in signal analysis a similar concept was introduced earlier in (Tichavský & Händel, 1995)].
Using residue calculus (Jury, 1964), one can easily derive analytical expressions quantifying the mean-squared estimation errors in terms of μ, γ_ω and γ_α. Unfortunately, these expressions (not listed here) are too complicated to enable minimization of the MSE scores in an explicit, analytical form. For this reason the optimal values of adaptation gains were searched numerically. The optimal settings that minimize the frequency and frequency rate, tracking and smoothing, errors were found to be in all four cases identical – the corresponding values, obtained for several nonstationarity rates κ, are listed in Table 2.
Our next step was to establish the Cramér-Rao-type lower tracking bounds (LTB_{ω_0}, LTB_{α_0}) and lower smoothing bounds (LSB_{ω_0}, LSB_{α_0}), which set the upper limits for the frequency and frequency rate estimation accuracy using *any* causal/noncausal identification algorithm – see Table 2. Note the performance gains that can be achieved when tracking is replaced with smoothing. Even though such analysis has mainly a theoretical value, it allows one to evaluate tracking/smoothing performance of the proposed algorithms in absolute, rather than relative, terms. The LTBs and LSBs for the system governed by (1)–(3) were obtained under assumptions (A1) – (A5) and, to make the analysis easier, under some technical constraints imposed on the initial conditions – see Appendix 3. Again, rather than providing the closed-form analytical formulas, we show how LTBs and LSBs can be established numerically for a given value of κ.
It was found out that there is a perfect agreement between the lower tracking and smoothing bounds and the MSE values obtained by minimizing (19)–(20) and (21)–(22), respectively – in some cases the computed values agreed up to the six decimal place. This means that, at least theoretically, the optimally tuned GACF/GACS algorithms should be statistically efficient frequency and frequency rate trackers/smoothers. In the next section we will verify this statement using computer simulations.
According to the results given above, in the constant-amplitude case the frequency estimates are unambiguous – in spite of the fact that the corresponding amplitude estimates are complex-valued quantities and, as such, could potentially create nonidentifiability problems. Although the same was also observed in the time-varying-amplitude case, some caution is needed in interpreting the quantities $\widehat{\omega}_0(t)$, $\widetilde{\omega}_0(t)$ and $\widehat{\boldsymbol{\beta}}_k(t)$, $\widetilde{\boldsymbol{\beta}}_k(t)$ unless identifiability is formally proved. Note, however, that this potential nonidentifiability problem does not extend to estimation of $\boldsymbol{\theta}(t)$, which is our main interest here.

7 Simulation and experimental results

7.1 ALF-based analysis

To check the validity of the analytical expressions (19)–(20), based on the approximating linear filter equations (13)–(14), the following two-tap FIR system (inspired by channel equalization applications) was simulated

$$y(t) = \theta_1(t)u(t) + \theta_2(t)u(t-1) + v(t) \qquad (25)$$

where $u(t)$ denotes a white 4-QAM [quadrature amplitude modulation – see e.g. (Giannakis & Tepedelenlioğlu, 1998)] input sequence ($u(t) = \pm 1 \pm j$, $\sigma_u^2 = 2$) and $v(t)$ denotes a complex-valued Gaussian measurement noise.
Each of $n = 2$ impulse response coefficients had $K = 4$ modes of variation – system parameters varied according to

$$\boldsymbol{\theta}(t) = \begin{bmatrix}\theta_1(t)\\ \theta_2(t)\end{bmatrix} = \sum_{k=1}^{4} \boldsymbol{\beta}_{k0} e^{j\left[\nu_k + m_k \sum_{\tau=1}^{t}\omega_0(\tau)\right]}$$
$$\boldsymbol{\beta}_{10} = [2-j, 1+j2]^{\mathrm{T}}, \boldsymbol{\beta}_{20} = [1-j, 1]^{\mathrm{T}}$$
$$\boldsymbol{\beta}_{30} = [1, 1]^{\mathrm{T}}, \boldsymbol{\beta}_{40} = [0.5, j0.5]^{\mathrm{T}}$$

where $m_k = k$, the phase shifts $\nu_1, \ldots, \nu_4$ were drawn independently from the uniform distribution on $[0, 2\pi)$, and the fundamental frequency $\omega_0(t)$ was governed by the integrated random-walk model [obeying (A4) and (A5)], starting from $\omega_0(0) = \pi/4$. Note that in this case $\boldsymbol{\varphi}(t) = [u(t), u(t-1)]^{\mathrm{T}}$, $\boldsymbol{\Phi} = \mathbf{I}_2\sigma_u^2$, $b_1^2 = 20$, $b_2^2 = 6$, $b_3^2 = 4$, $b_4^2 = 1$ and $b_0^2 = 96$ (since $\sum_{k=1}^{4} b_k^2 = 31$, there was a noticeable discrepancy between SNR and effective SNR).
The approximations (19)–(20) were checked for 3 values of the signal-to-noise ratio: SNR=0 dB ($\sigma_v = 5.5678$), SNR=10 dB ($\sigma_v = 1.7607$) and SNR=20 dB ($\sigma_v = 0.5568$), for 2 values of the nonstationarity rate: $\kappa = 10^{-10}$ and $\kappa = 10^{-9}$, and for 10 values of the adaptation gain μ, ranging from 0.01 to 0.1. To reduce the number of design degrees of freedom, the two other gains adopted for GACF/GACS algorithms were set to: $\gamma_\omega = \mu^2/2$ and $\gamma_\alpha = \mu^3/8$ – in agreement with the general tendency observed, under different levels of effective SNR, for the optimal settings (this rule of thumb was found to work quite well in practice). The mean-squared frequency and frequency rate estimation errors were evaluated (for the optimally tuned GACF algorithm) by means of joint time and ensemble averaging. First, for each realization of the measurement noise sequence and each realization of the frequency trajectory, the mean-squared errors were computed from 1000 iterations of the GACF filter (after the algorithm has reached its steady-state). The obtained results were next averaged over 50 realizations of $\{\delta(t), v(t)\}$ and $\nu_1, \ldots, \nu_4$.
Figs. 1 and 2 show comparison of theoretical curves and the time-averaged values of the mean-squared fre-

Table 2. Optimal GACF/GACS settings and the corresponding normalized lower tracking bounds.

κ	μ^{opt}	$\gamma_\omega^{\text{opt}}$	$\gamma_\alpha^{\text{opt}}$	$\text{LTB}_{\omega_0}/\sigma_\delta^2$	$\text{LTB}_{\alpha_0}/\sigma_\delta^2$	$\text{LSB}_{\omega_0}/\sigma_\delta^2$	$\text{LSB}_{\alpha_0}/\sigma_\delta^2$
10^{-10}	0.0472	0.00113	0.0000138	$2.05 \cdot 10^5$	$8.21 \cdot 10^1$	$1.18 \cdot 10^4$	$1.38 \cdot 10^1$
$5 \cdot 10^{-10}$	0.0613	0.00192	0.0000306	$9.09 \cdot 10^4$	$6.28 \cdot 10^1$	$5.28 \cdot 10^3$	$1.05 \cdot 10^1$
10^{-9}	0.0685	0.00241	0.0000432	$6.39 \cdot 10^4$	$5.58 \cdot 10^1$	$3.73 \cdot 10^3$	9.39
$5 \cdot 10^{-9}$	0.0886	0.00407	0.0000955	$2.82 \cdot 10^4$	$4.26 \cdot 10^1$	$1.67 \cdot 10^3$	7.18
10^{-8}	0.0990	0.00509	0.000134	$1.97 \cdot 10^4$	$3.79 \cdot 10^1$	$1.18 \cdot 10^3$	6.40
$5 \cdot 10^{-8}$	0.127	0.00852	0.000295	$8.66 \cdot 10^3$	$2.89 \cdot 10^1$	$5.28 \cdot 10^2$	4.89
10^{-7}	0.142	0.0106	0.000414	$6.06 \cdot 10^3$	$2.57 \cdot 10^1$	$3.73 \cdot 10^2$	4.36
$5 \cdot 10^{-7}$	0.181	0.0177	0.000905	$2.63 \cdot 10^3$	$1.95 \cdot 10^1$	$1.67 \cdot 10^2$	3.33
10^{-6}	0.201	0.0219	0.00126	$1.83 \cdot 10^3$	$1.73 \cdot 10^1$	$1.18 \cdot 10^2$	2.96

quency and frequency rate estimation errors obtained via simulation. Note the good agreement between theoretical evaluations and the actual algorithm's performance, which can be observed for SNR $\geq$ 10 dB. Generally, the degree of fit improves with decreasing κ and increasing SNR, which is consistent with the operating range of the approximating linear filter technique. Similar results, not reported here, were obtained for the GACS algorithm.

7.2 Statistical efficiency

Fig. 3 shows comparison of the theoretical values of the lower frequency tracking bound LTB_{ω_0} and the lower frequency rate tracking bound LTB_{α_0} with experimental results obtained for optimally tuned GACF algorithm designed for the system described in Section 7.1. In agreement with the results of theoretical analysis presented in Section 6, for small rates of system nonstationarity the proposed GACF algorithm is statistically efficient, i.e., under the conditions specified earlier, it can't be outperformed by any other tracking algorithm. The same conclusion can be drawn after inspection of the plots shown in Fig. 4, illustrating behavior of the optimally tuned GACS algorithm.

7.3 Performance

The aim of this simulation experiment was to compare performance of the proposed GACF/GACS algorithms with that yielded by the unconstrained multiple-frequency versions of the GANF/GANS algorithms.
The simulated two-tap FIR system (25) was governed by

$$\boldsymbol{\theta}(t) = \begin{bmatrix} \theta_1(t) \\ \theta_2(t) \end{bmatrix} = \sum_{k=1}^{4} \mathbf{C}(t)\boldsymbol{\beta}_{k0} e^{jm_k \sum_{\tau=1}^{t} \omega_0(\tau)}$$

where $\mathbf{C}(t) = \text{diag}\{\sin(\pi t/1000), \cos(\pi t/1000)\}$ and $\boldsymbol{\beta}_{10} = [-1, -j]^{\text{T}}$, $\boldsymbol{\beta}_{20} = [-0.112, j0.112]^{\text{T}}$, $\boldsymbol{\beta}_{30} = [-0.0402, -j0.0402]^{\text{T}}$, $\boldsymbol{\beta}_{40} = [-0.0207, j0.0207]^{\text{T}}$.
The fundamental frequency was changing sinusoidally according to

$$\omega_0(t) = \frac{\pi}{10}\left[1 + \frac{1}{3}\sin\left(\frac{\pi t}{1000}\right)\right].$$

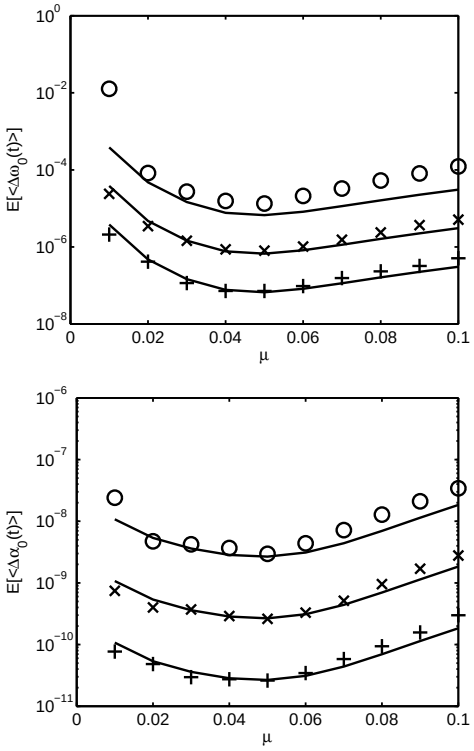

Fig. 1. Average variance of the frequency estimation error (upper figure) and the frequency rate estimation error (lower figure) for a nonstationary FIR system with 4 frequency modes governed by the integrated random-walk model. The theoretical results (solid lines) are compared with simulation results obtained for the rate of nonstationarity $\kappa = 10^{-10}$ and 10 different values of μ ($\gamma_\omega = \mu^2/2$, $\gamma_\alpha = \mu^3/8$). The corresponding signal-to-noise ratios were equal to: SNR=0 dB (○), SNR=10 dB (×) and SNR=20 dB (+).

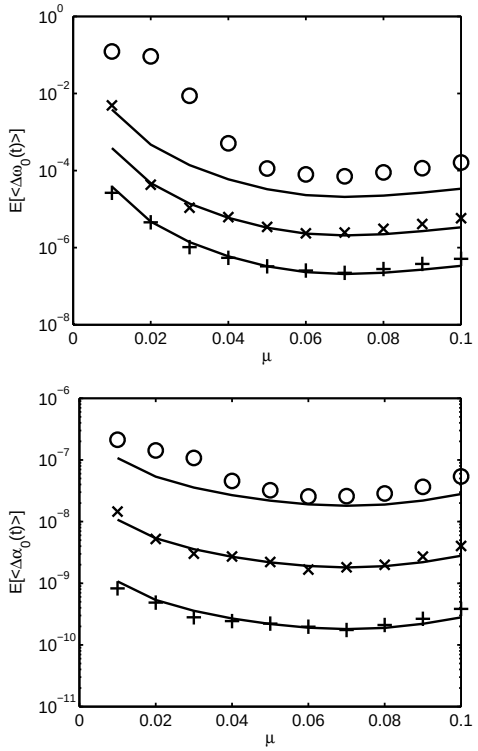

Fig. 2. Average variance of the frequency estimation error (upper figure) and the frequency rate estimation error (lower figure) for a nonstationary FIR system with 4 frequency modes governed by the integrated random-walk model. The theoretical results (solid lines) are compared with simulation results obtained for the rate of nonstationarity $\kappa = 10^{-9}$ and 10 different values of μ ($\gamma_\omega = \mu^2/2$, $\gamma_\alpha = \mu^3/8$). The corresponding signal-to-noise ratios were equal to: SNR=0 dB (○), SNR=10 dB (×) and SNR=20 dB (+).

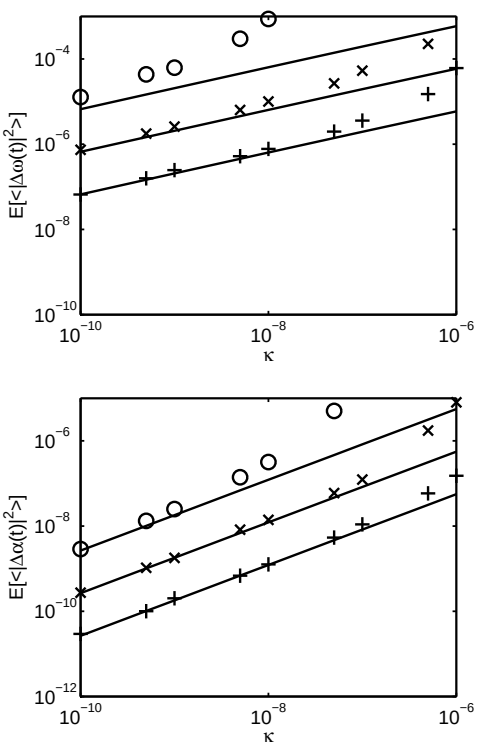

Fig. 3. Comparison of the theoretical values of the lower frequency (upper figure) and frequency rate (lower figure) tracking bounds (solid lines) with experimental results obtained for the system with quasi-linear frequency changes for 3 different SNR values: SNR=0 dB (○), SNR=10 dB (×), SNR=20 dB (+), and 9 different values of the rate of system nonstationarity κ.

Only the odd harmonics were present: $m_k = 2k - 1$, $k = 1, \ldots, 4$. Similarly as in the previous experiments, the system was excited with a 4-QAM sequence.

Fig. 7 shows comparison of the mean-squared parameter tracking/smoothing errors yielded by the proposed GACF/GACS algorithms. All MSE values were obtained by means of joint time averaging (the evaluation interval [1001,3000] was placed inside a wider analysis interval [1,4000]), and ensemble averaging (50 realizations of measurement noise were used). The standard deviation of noise was equal to $\sigma_v = 0.04$. For each value of μ, the values of adaptation gains γ_ω and γ_α were chosen using the rule of thumb described in Section 7.1. Note that the GACS algorithm yields uniformly better results than its GACF counterpart. Both GACF and GACS algorithms perform better than the correspnding multifrequency GANF and GANS algorithms, respectively.

7.4 Estimation of MRI noise

Magnetic resonance imaging (MRI) equipment is used to visualize internal structures of the human body without exposing subjects to harmful radiation. It is utilized in many medical institutions for diagnostic purposes and, quite recently, as an aid during some operations – in the latter case it works in the nearly real-time mode (Kurumi et al., 2007).

MRI devices generate very loud harmonic noise (with intensity exceeding 100 dB) caused by vibration – owing to the Lorentz force – of the gradient coil. Exposure to this noise is very annoying both for the patients and for the medical staff. MRI noise can be reduced using the active noise control (ANC) techniques, the better the more accurately one can track the underlying multiharmonic signal ($K > 30$). Since this signal is nonstationary – both the amplitudes and the fundamental frequency change over time – its estimation is a challenging task. Fig. 8 shows the time plots and periodograms of the original MRI noise (recorded in the axial mode), as well as the time plots and periodograms of prediction errors $\varepsilon(t) = y(t) - \widehat{s}(t|t-1) = y(t) - \sum_{k=1}^{K} \widehat{f}_k(t)\widehat{b}_k(t-1)$ yielded by the proposed adaptive comb filter (ACF) and by the multiple-frequency version of the adaptive notch filter (ANF). All results were obtained for $\mu = 0.01$, $\gamma_\omega = \mu^2/2$ and $\gamma_\alpha = \mu^3/8$. The complex-valued version of the MRI signal was obtained using the discrete Hilbert transform.

While the time plots obtained for ANF and ACF look similar, the corresponding periodograms differ significantly. It is clear that the ACF is much more effective

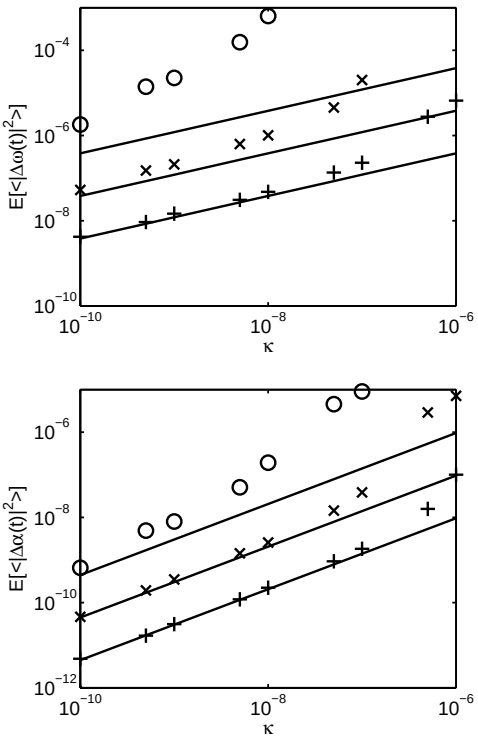

Fig. 4. Comparison of the theoretical values of the lower frequency (upper figure) and frequency rate (lower figure) smoothing bounds (solid lines) with experimental results obtained for the system with quasi-linear frequency changes, for 3 different SNR values: SNR=0 dB (○), SNR=10 dB (×), SNR=20 dB (+), and 9 different values of the rate of system nonstationarity κ.

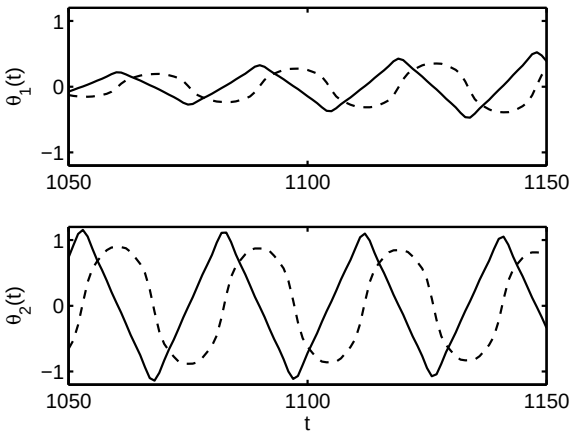

Fig. 5. Real parts (solid lines) and imaginary parts (broken lines) of system parameters observed in a short time interval.

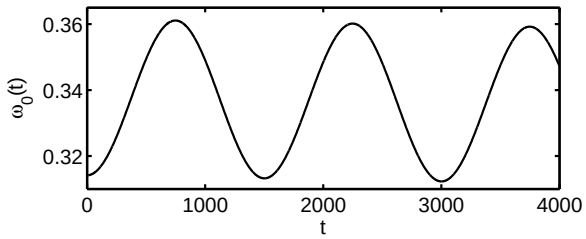

Fig. 6. Evolution of the instantaneous fundamental frequency

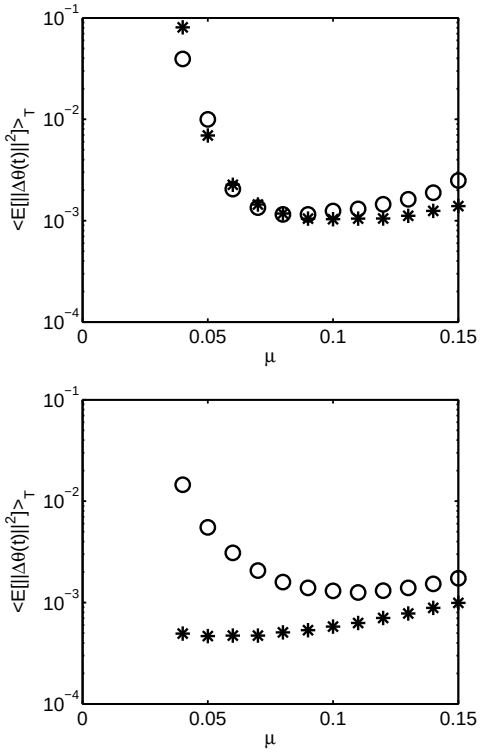

Fig. 7. Comparison of the mean-squared parameter tracking (○) and parameter smoothing (∗) errors yielded by the generalized adaptive notch algorithms (upper figure) and generalized adaptive comb algorithms (lower figure) for the two-tap FIR system with sinusoidally varying fundamental frequency of parameter variation and sinusoidally varying amplitudes of all harmonics.

in suppressing signal harmonics than ANF, even though both algorithms used the same starting values. The failure of the ANF algorithm can be explained by its poor frequency matching capabilities – after some initial period the algorithm locks on dominant harmonics, leaving the remaining ones unattenuated.

8 Conclusion

The problem of identification of linear stochastic systems with quasi-harmonically varying parameters was considered. Both causal and noncausal identification algorithms were derived, referred to as generalized adaptive comb filters (GACFs) and generalized adaptive comb smoothers (GACSs), respectively. In both cases the frequency and frequency rate estimation

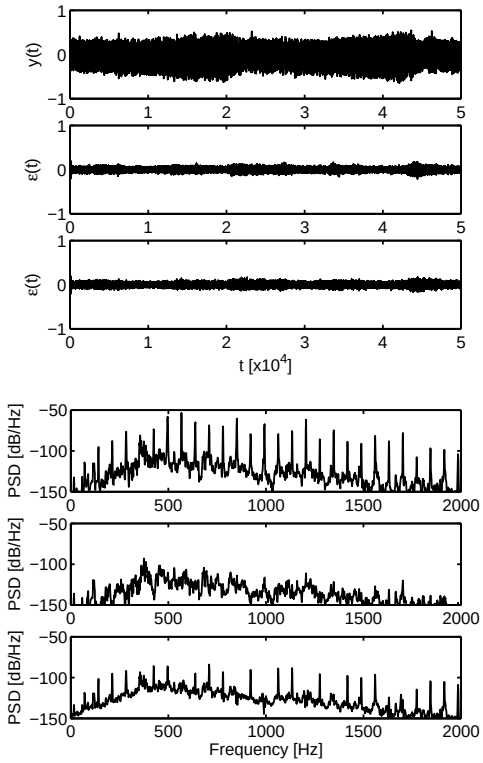

Fig. 8. Time plots (upper figures) and periodograms (lower figures) of the: MRI noise (two top plots), prediction errors yielded by the ACF algorithm (two middle plots), and prediction errors yielded by the multiple-frequency ANF algorithm (two bottom plots).

properties of the proposed algorithms were analyzed using the method of approximating linear filter. It was shown, and later confirmed by means of computer simulations, that when the fundamental frequency of parameter changes varies slowly with time according to the integrated random-walk model, the optimally tuned GACF/GACS algorithms are (under Gaussian assumptions) statistically efficient frequency and frequency rate trackers/smoothers, i.e., they reach the Cramér-Rao-type lower tracking/smoothing bounds – expressions allowing one to evaluate these bounds were also derived in the paper.

References

Bakkoury, J., Roviras, D., Ghogho, M. & Castanie, F. (2000). Adaptive MLSE receiver over rapidly fading channels. *Signal Processing*, 80, 1347–1360.

Giannakis, G.B. & Tepedelenlioğlu, C. (1998). Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels. *Proc. IEEE*, 86, 1969–1986.

Haykin, S. (1996). *Adaptive Filter Theory*. Englewood Cliffs NJ: Prentice Hall.

James, B., Anderson B.D.O. & Williamson, R.C. (1994). Conditional mean and maximum likelihood approaches to multiharmonic frequency estimation. *IEEE Trans. Signal Process.*, 42, 1366–1375.

Jury, M. (1964). *Theory and Application of the $\mathcal{Z}$-transform Method*. New York: Wiley.

Kurumi, Y., Tani T., Naka S., Shiomi H., Shimizu T., Abe K., Endo Y. & Morikawa S. (2007). MR-guided microwave ablation for malignancies. *International Journal of Clinical Oncology*, 12, 85–93.

Nehorai A. & Porat B. (1986). Adaptive comb filtering for harmonic signal enhancement. *IEEE Trans. Acoust. Speech, Signal Process.*, 34, 1124-1138.

Neimark, J.I. (2003). *Mathematical Models in Natural Sciences and Engineering*. Springer.

Niedźwiecki, M. & Kaczmarek, P. (2004). Generalized adaptive notch filters. *Proc. 2004 IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, Montreal, Canada, II-657–II-660.

Niedźwiecki, M. & Kaczmarek, P. (2005a). Estimation and tracking of quasi-periodically varying systems. *Automatica*, 41, 1503-1516.

Niedźwiecki, M. & Kaczmarek, P. (2005b). Identification of quasi-periodically varying systems using the combined nonparametric/parametric approach. *IEEE Trans. Signal Process.*, 53, 4588–4598.

Niedźwiecki, M. & Meller, M. (2011a). Identification of quasi-periodically varying systems with quasi-linear frequency changes. *Proc. 18th IFAC World Congress*, Milano, Italy, pp. 9070-9078.

Niedźwiecki, M. & Meller, M. (2011b). New algorithms for adaptive notch smoothing. *IEEE Trans. Signal Process.*, 59, 2024-2037.

Olkin, I. & Pratt, J. (1958). A multivariate Tchebycheff inequality. *Annals of Mathematical Statistics*, 29, 226–234.

Regalia, P.A. (1995). *Adaptive IIR Filtering in Signal Processing and Control*. New York: Marcel Dekker.

Tichavský, P. & Händel, P. (1995). Two algorithms for adaptive retrieval of slowly time-varying multiple cisoids in noise. *IEEE Trans. on Signal Process.*, 43, 1116-1127.

Tsatsanis, M.K. & Giannakis, G.B. (1996). Modeling and equalization of rapidly fading channels. *Int. J. Adaptive Contr. Signal Processing*, 10, 159–176.

Widrow, B. & Stearns, S.D. (1985). *Adaptive Signal Processing*. Prentice-Hall.

van Trees, H. (1968). *Detection, Estimation and Modulation Theory*. New York: Wiley.

APPENDIX 1
Derivation of (13) and (14)

Denote by $\Delta\widehat{\boldsymbol{\theta}}_k(t) = \boldsymbol{\theta}_k(t) - \widehat{\boldsymbol{\theta}}_k(t)$ the parameter estimation error and let $\Delta\widehat{x}_k(t) = \mathrm{Im}[\boldsymbol{\theta}_k^{\mathrm{H}}(t)\boldsymbol{\Phi}\ \Delta\widehat{\boldsymbol{\theta}}_k(t)/b_0^2]$. According to (Tichavský & Händel, 1995), when carrying ALF analysis, one should neglect all terms of order higher than one in $\Delta\widehat{\omega}_k(t)$, $\Delta\widehat{\alpha}_k(t)$, $\Delta\widehat{\boldsymbol{\theta}}_k(t)$, $\delta(t)$ and $v(t)$, including all cross-terms.

To derive recursion for $\Delta\widehat{x}(t) = \sum_{k=1}^{K}\Delta\widehat{x}_k(t)$, note that

$$\widehat{\boldsymbol{\theta}}_k(t) = \boldsymbol{\zeta}_k(t) + \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\varepsilon(t)$$
$$\varepsilon(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\sum_{k=1}^{K}\boldsymbol{\theta}_k(t) + v(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\sum_{k=1}^{K}\boldsymbol{\zeta}_k(t)$$

where $\boldsymbol{\zeta}_k(t) = e^{j[\widehat{\omega}_k(t-1)+\widehat{\alpha}_k(t-1)]}\widehat{\boldsymbol{\theta}}_k(t-1)$ and

$$\widehat{\alpha}_k(t) = m_k\widehat{\alpha}_0(t), \quad \widehat{\omega}_k(t) = m_k\widehat{\omega}_0(t) \qquad (26)$$

Therefore

$$\widehat{\boldsymbol{\theta}}_k(t) = \left[\, \mathbf{I} - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) \,\right]\boldsymbol{\zeta}_k(t) + \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}_k(t) + \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)v(t) + \boldsymbol{\xi}_k(t)$$

where

$$\boldsymbol{\xi}_k(t) = \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\sum_{\substack{i=1\\ i\neq k}}^{K}\left[\, \boldsymbol{\theta}_i(t) - \boldsymbol{\zeta}_i(t) \,\right]$$

Since in the case considered $\boldsymbol{\theta}_k(t) = e^{j\omega_k(t)}\boldsymbol{\theta}_k(t-1)$, the last relationship leads to

$$\begin{aligned}\Delta\widehat{\boldsymbol{\theta}}_k(t) = {} & \left[\, \mathbf{I} - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) \,\right]\boldsymbol{\theta}_k(t)\\ & - \left[\, \mathbf{I} - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) \,\right]\boldsymbol{\zeta}_k(t)\\ & - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)v(t) - \boldsymbol{\xi}_k(t). \end{aligned} \tag{27}$$

Note that $\boldsymbol{\zeta}_k(t)$ can be rewritten in the form $\boldsymbol{\zeta}_k(t) = e^{j\omega_k(t)}e^{-j\Delta\widehat{\omega}_k(t-1)}e^{-j\Delta\widehat{\alpha}_k(t-1)}[\boldsymbol{\theta}_k(t-1) - \Delta\widehat{\boldsymbol{\theta}}_k(t-1)]$. Using the following approximations $e^{-j\Delta\widehat{\omega}_k(t-1)} \cong 1 - j\Delta\widehat{\omega}_k(t-1)$, $e^{-j\Delta\widehat{\alpha}_k(t-1)} \cong 1 - j\Delta\widehat{\alpha}_k(t-1)$, that hold for small frequency and frequency rate errors, respectively, and applying ALF rules, one arrives at

$$\begin{aligned}&\boldsymbol{\zeta}_k(t) = e^{j\omega_k(t)}e^{-j\Delta\widehat{\omega}_k(t-1)}e^{-j\Delta\widehat{\alpha}_k(t-1)}\times\\ &\times[\boldsymbol{\theta}_k(t-1) - \Delta\widehat{\boldsymbol{\theta}}_k(t-1)] \cong e^{j\omega_k(t)}[\,1 - j\Delta\widehat{\omega}_k(t-1)]\times\\ &\times[1 - j\Delta\widehat{\alpha}_k(t-1)][\boldsymbol{\theta}_k(t-1) - \Delta\widehat{\boldsymbol{\theta}}_k(t-1)]\\ &\cong e^{j\omega_k(t)}[\,1 - j\Delta\widehat{\omega}_k(t-1) - j\Delta\widehat{\alpha}_k(t-1)]\times\\ &\times[\boldsymbol{\theta}_k(t-1) - \Delta\widehat{\boldsymbol{\theta}}_k(t-1)] \cong \boldsymbol{\theta}_k(t) - e^{j\omega_k(t)}\Delta\widehat{\boldsymbol{\theta}}_k(t-1)\\ &- j[\Delta\widehat{\omega}_k(t-1) + \Delta\widehat{\alpha}_k(t-1)]\boldsymbol{\theta}_k(t). \end{aligned} \tag{28}$$

Combining (27) with (28), one obtains

$$\begin{aligned}\Delta\widehat{\boldsymbol{\theta}}_k(t) \cong {} & \left[\, \mathbf{I} - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) \,\right]e^{j\omega_k(t)}\Delta\widehat{\boldsymbol{\theta}}_k(t-1)\\ & + j\left[\, \mathbf{I} - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) \,\right][\Delta\widehat{\omega}_k(t-1)\\ & + \Delta\widehat{\alpha}_k(t-1)]\boldsymbol{\theta}_k(t) - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)v(t) - \boldsymbol{\xi}_k(t). \end{aligned} \tag{29}$$

Let $\Delta\widetilde{\boldsymbol{\theta}}_k(t) = \Delta\widehat{\boldsymbol{\theta}}_k(t)f_k^*(t)$. After multiplying both sides of (29) with $f_k^*(t)$, one arrives at

$$\begin{aligned}\Delta\widetilde{\boldsymbol{\theta}}_k(t) \cong {} & \left[\, \mathbf{I} - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) \,\right]\Delta\widetilde{\boldsymbol{\theta}}_k(t-1)\\ & + j\left[\, \mathbf{I} - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) \,\right][\Delta\widehat{\omega}_k(t-1)\\ & + \Delta\widehat{\alpha}_k(t-1)]\boldsymbol{\beta}_k - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)f_k^*(t)v(t) - \boldsymbol{\xi}_k(t)f_k^*(t). \end{aligned} \tag{30}$$

For small values of adaptation gains μ, γ_ω and γ_α, the quantities $\Delta\widetilde{\boldsymbol{\theta}}_k(t)$, $\Delta\widehat{\omega}_k(t)$ and $\Delta\widehat{\alpha}_k(t)$ change slowly compared to $\boldsymbol{\varphi}(t)$ and $f_k(t), k = 1, \ldots, K$, i.e., (30) can be regarded as a two-scale difference equation. When solving such equation for slowly varying quantities, one is allowed to replace some functionals of the fast varying quantities with their time averages. This is usually referred to as a deterministic averaging technique.

Denote by $\langle x(t)\rangle_T = (1/T)\sum_{i=0}^{T-1} x(t-i)$ the local average of $x(t)$, and by $\langle x(t)\rangle_\infty = \lim_{T\mapsto\infty}\langle x(t)\rangle_T$ the corresponding limiting value (provided it exists).

We will exploit the fact that for any process $\{\boldsymbol{\varphi}(t)\}$ obeying (A2), and for sufficiently large values of T, it holds that $\left\langle \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\right\rangle_T \cong \left\langle \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\right\rangle_\infty = \boldsymbol{\Phi}$ and $\left\langle \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)e^{j\omega t}\right\rangle_T \cong \left\langle \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)e^{j\omega t}\right\rangle_\infty = \mathbf{0},\ \forall\omega \neq 0$. According to the first relationship, the data-dependent matrix $\mathbf{I} - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)$ in (30) can be replaced with $\mathbf{I} - \mu\boldsymbol{\Phi}^{-1}\left\langle \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\right\rangle_T \cong (1-\mu)\mathbf{I}$. Similarly, according to the second relationship, since for sufficiently slow frequency variations $f_k(t), \widehat{f}_k(t)$, $k = 1, \ldots, K$, are locally almost periodic functions of time, it holds that $\langle\boldsymbol{\xi}_k(t)f_k^*(t)\rangle_T \cong 0$, allowing one to neglect the last term on the right hand side of (30). Hence, using the averaging technique, one arrives at the following approximate error equation

$$\begin{aligned}\Delta\widetilde{\boldsymbol{\theta}}_k(t) \cong {} & (1-\mu)\Delta\widetilde{\boldsymbol{\theta}}_k(t-1) + j(1-\mu)[\Delta\widehat{\omega}_k(t-1)\\ & + \Delta\widehat{\alpha}_k(t-1)]\boldsymbol{\beta}_k - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)f_k^*(t)v(t). \end{aligned} \tag{31}$$

Multiplying both sides of (31) with $\boldsymbol{\beta}_k^{\mathrm{H}}\boldsymbol{\Phi}$, one obtains

$$\begin{aligned}\boldsymbol{\beta}_k^{\mathrm{H}}\boldsymbol{\Phi}\Delta\widetilde{\boldsymbol{\theta}}_k(t) \cong {} & \lambda\boldsymbol{\beta}_k^{\mathrm{H}}\boldsymbol{\Phi}\Delta\widetilde{\boldsymbol{\theta}}_k(t-1)\\ & + j\lambda[\Delta\widehat{\omega}_k(t-1) + \Delta\widehat{\alpha}_k(t-1)]\boldsymbol{\beta}_k^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_k\\ & - \mu\boldsymbol{\beta}_k^{\mathrm{H}}\boldsymbol{\varphi}^*(t)f_k^*(t)v(t) \end{aligned} \tag{32}$$

where $\lambda = 1 - \mu$. Finally, dividing both sides of equation (32) by b_0^2, taking imaginary parts, and noting that $\boldsymbol{\beta}_k^{\mathrm{H}}\boldsymbol{\Phi}\Delta\widetilde{\boldsymbol{\theta}}_k(t) = \boldsymbol{\theta}_k(t)^{\mathrm{H}}\boldsymbol{\Phi}\Delta\widehat{\boldsymbol{\theta}}_k(t)$ and $\boldsymbol{\beta}_k^{\mathrm{H}}f_k^*(t) = \boldsymbol{\theta}_k^{\mathrm{H}}(t)$, one arrives at

$$\begin{aligned}\Delta\widehat{x}_k(t) \cong {} & \lambda\Delta\widehat{x}_k(t-1)\\ & + \frac{\lambda b_k^2}{b_0^2}\left[\, \Delta\widehat{\omega}_k(t-1) + \Delta\widehat{\alpha}_k(t-1) \,\right] + \mu e_k(t) \end{aligned}$$

where $e_k(t) = -\mathrm{Im}\left[\boldsymbol{\beta}_k^{\mathrm{H}}\boldsymbol{\varphi}^*(t)f_k^*(t)v(t)/b_0^2\right]$. Note that $\Delta\widehat{x}(t) = \sum_{k=1}^K m_k\Delta\widehat{x}_k(t)$, $e_0(t) = \sum_{k=1}^K m_k e_k(t)$ and $b_0^2 = \sum_{k=1}^K m_k^2 b_k^2$. Hence, after incorporating (26), one obtains

$$\begin{aligned}\Delta\widehat{x}(t) \cong {} & \lambda\Delta\widehat{x}(t-1) + \lambda[\Delta\widehat{\omega}_0(t-1) + \Delta\widehat{\alpha}_0(t-1)]\\ & + \mu e_0(t). \end{aligned} \tag{33}$$

To derive recursions for $\Delta\widehat{\omega}_0(t)$ and $\Delta\widehat{\alpha}_0(t)$, note that in the tracking mode it holds that

$$g(t) = \frac{\sum_{k=1}^{K} m_k \, \mathrm{Im}\left[\varepsilon^\star(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{f}_k(t)\widehat{\boldsymbol{\beta}}_k(t-1)\right]}{\sum_{k=1}^{K} m_k^2 \, \widehat{\boldsymbol{\beta}}_k^{\mathrm{H}}(t-1)\boldsymbol{\Phi}\widehat{\boldsymbol{\beta}}_k(t-1)}$$
$$\cong \frac{\sum_{k=1}^{K} m_k \mathrm{Im}[\varepsilon^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\zeta}_k(t)]}{b_0^2}$$

and

$$\varepsilon(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}_k(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\zeta}_k(t) + \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\psi}_k(t) + v(t)$$
$$\boldsymbol{\psi}_k(t) = \sum_{\substack{i=1 \\ i\neq k}}^{K} [\, \boldsymbol{\theta}_i(t) + \boldsymbol{\zeta}_i(t) \,].$$

Furthermore, $\mathrm{Im}[\varepsilon^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\zeta}_k(t)] = \mathrm{Im}[z_k(t) - |\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\zeta}_k(t)|^2] = \mathrm{Im}[z_k(t)]$, where $z_k(t) = \boldsymbol{\theta}_k^{\mathrm{H}}(t)\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\zeta}_k(t) + \boldsymbol{\psi}_k^{\mathrm{H}}(t)\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\zeta}_k(t) + v^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}_k(t)$.
Using (28) and applying the ALF rules, one obtains the following approximation

$$\begin{aligned}
&\boldsymbol{\theta}_k^{\mathrm{H}}(t)\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\zeta}_k(t) \cong \boldsymbol{\theta}_k^{\mathrm{H}}(t)\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}_k(t) \\
&- \boldsymbol{\theta}_k^{\mathrm{H}}(t-1)\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\Delta\widehat{\boldsymbol{\theta}}_k(t-1) \\
&- j\boldsymbol{\theta}_k^{\mathrm{H}}(t)\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}_k(t)[\, \Delta\widehat{\omega}_k(t-1) + \Delta\widehat{\alpha}_k(t-1) \,] \\
&= \boldsymbol{\beta}_k^{\mathrm{H}}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\beta}_k - \boldsymbol{\beta}_k^{\mathrm{H}}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\Delta\widetilde{\boldsymbol{\theta}}(t-1) \\
&- j\boldsymbol{\beta}_k^{\mathrm{H}}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\beta}_k[\, \Delta\widehat{\omega}_k(t-1) + \Delta\widehat{\alpha}_k(t-1) \,]
\end{aligned}$$

which, after averaging, leads to

$$\begin{aligned}
z_k(t) \cong{}& \boldsymbol{\beta}_k^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_k - j\boldsymbol{\beta}_k^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_k[\, \Delta\widehat{\omega}_k(t-1) + \Delta\widehat{\alpha}_k(t-1) \,] \\
&- \boldsymbol{\beta}_k^{\mathrm{H}}\boldsymbol{\Phi}\Delta\widetilde{\boldsymbol{\theta}}_k(t-1) + v^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}_k(t).
\end{aligned}$$

Since $\boldsymbol{\beta}_k^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_k = b_k^2$, one arrives at

$$\begin{aligned}
g(t) = \mathrm{Im}[z(t)/b_0^2] \cong{}& -\Delta\widehat{x}(t-1) - \Delta\widehat{\omega}_0(t-1) \\
&- \Delta\widehat{\alpha}_0(t-1) + e_0(t).
\end{aligned}$$

Note that

$$\begin{aligned}
\Delta\widehat{\alpha}_0(t) &= \Delta\widehat{\alpha}_0(t-1) + \delta(t) + \gamma_\alpha g(t) \\
\Delta\widehat{\omega}_0(t) &= \Delta\widehat{\omega}_0(t-1) + \Delta\widehat{\alpha}_0(t-1) + \gamma_\omega g(t).
\end{aligned}$$

Combining the last three equations, one arrives at

$$\begin{aligned}
\Delta\widehat{\alpha}_0(t) \cong{}& (1-\gamma_\alpha)\Delta\widehat{\alpha}_0(t-1) + \delta(t) + \gamma_\alpha e_0(t) \\
&- \gamma_\alpha\Delta\widehat{\omega}_0(t-1) - \gamma_\alpha\Delta\widehat{x}(t-1) \qquad (34)
\end{aligned}$$

$$\begin{aligned}
\Delta\widehat{\omega}_0(t) \cong{}& (1-\gamma_\omega)\Delta\widehat{\omega}_0(t-1) + (1-\gamma_\omega)\Delta\widehat{\alpha}_0(t-1) \\
&+ \gamma_\omega e_0(t) - \gamma_\omega\Delta\widehat{x}(t-1). \qquad (35)
\end{aligned}$$

Finally, solving the set of linear equations (33), (34) and (35) for $\Delta\widehat{\omega}_0(t)$ and $\Delta\widehat{\alpha}_0(t)$, one obtains (13) and (14), respectively.

APPENDIX 2
Derivation of (16)

The relationship $\mathrm{E}\,[e_k(t)e_l(t)] = 0$, $\forall k \neq l$ stems from the fact that $\{v(t)\}$ is a sequence of zero-mean independent random variables, independent of $\{\boldsymbol{\varphi}(t)\}$. To arrive at the expression for $\rho_{kl}(t)$ we will introduce the following notation

$$v(t) = v_{\mathrm{R}}(t) + jv_{\mathrm{I}}(t), \;\; \eta_k(t) = \boldsymbol{\beta}_k^{\mathrm{T}}\boldsymbol{\varphi}(t) = \eta_{k\mathrm{R}}(t) + j\eta_{k\mathrm{I}}(t)$$

where the subscript R/I denotes the real/imaginary part of a complex variable.
Note that

$$\begin{aligned}
e_k(t) &= -\mathrm{Im}\left[\eta_k^*(t)e^{-j\phi_k(t)}v(t)/b_0^2\right] \\
&= \frac{1}{b_0^2}\big\{\, [\eta_{k\mathrm{R}}(t)v_{\mathrm{R}}(t) + \eta_{k\mathrm{I}}(t)v_{\mathrm{I}}(t)]\sin\phi_k(t) \\
&\quad -[\eta_{k\mathrm{R}}(t)v_{\mathrm{I}}(t) - \eta_{k\mathrm{I}}(t)v_{\mathrm{R}}(t)]\cos\phi_k(t)\big\}.
\end{aligned}$$

Since the sequence $\{v(t)\}$ is circular, it holds that $\mathrm{E}[v_{\mathrm{R}}^2(t)] = \mathrm{E}[v_{\mathrm{I}}^2(t)] = \sigma_v^2/2$ and $\mathrm{E}[v_{\mathrm{R}}(t)v_{\mathrm{I}}(t)] = 0$. Using these relationships, and the fact that the process $\{v(t)\}$ is independent of $\{\boldsymbol{\varphi}(t)\}$, one arrives at

$$\begin{aligned}
&\mathrm{E}_v\,[e_k(t)e_l(t)] \\
&= \frac{\sigma_v^2}{2b_0^4}\big\{\, [\eta_{k\mathrm{R}}(t)\eta_{l\mathrm{R}}(t) + \eta_{k\mathrm{I}}(t)\eta_{l\mathrm{I}}(t)]\cos\,[\phi_k(t) - \phi_l(t)] \\
&+ [\eta_{k\mathrm{R}}(t)\eta_{l\mathrm{I}}(t) + \eta_{k\mathrm{I}}(t)\eta_{l\mathrm{R}}(t)]\sin\,[\phi_k(t) - \phi_l(t)]\,\big\}. \qquad (36)
\end{aligned}$$

After elementary but tedious calculations, one can show that

$$\begin{aligned}
&\mathrm{E}_\varphi\,[\eta_{k\mathrm{R}}(t)\eta_{l\mathrm{R}}(t) + \eta_{k\mathrm{I}}(t)\eta_{l\mathrm{I}}(t)] = \mathrm{Re}[\boldsymbol{\beta}_k^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_l] \\
&\mathrm{E}_\varphi\,[\eta_{k\mathrm{R}}(t)\eta_{l\mathrm{I}}(t) + \eta_{k\mathrm{I}}(t)\eta_{l\mathrm{R}}(t)] = -\mathrm{Im}[\boldsymbol{\beta}_k^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_l]. \qquad (37)
\end{aligned}$$

Finally, combining (36) with (37), one obtains

$$\begin{aligned}
&\mathrm{E}\,[e_k(t)e_l(t)] = \mathrm{E}_\varphi\,\{\mathrm{E}_v\,[e_k(t)e_l(t)]\} \\
&= \frac{\sigma_v^2}{2b_0^4}\,\big\{\mathrm{Re}[\boldsymbol{\beta}_k^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_l]\cos[\phi_k(t) - \phi_l(t)] \\
&- \mathrm{Im}[\boldsymbol{\beta}_k^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_l]\sin[\phi_k(t) - \phi_l(t)]\big\} = \rho_{kl}(t).
\end{aligned}$$

APPENDIX 3
Computation of Lower Tracking/Smoothing Bounds

In this appendix, we will derive expressions for theoretical upper bounds that limit tracking/smoothing capabilities of *any* causal/noncausal frequency and frequency

rate estimation algorithms applied to quasi-periodically varying systems with quasi-linear frequency changes. The corresponding lower tracking bounds (LTBs) and lower smoothing bounds (LSBs) belong to the class of posterior (or Bayesian) Cramér-Rao bounds, applicable to signals/systems with random parameters [3].

Denote by $\boldsymbol{u}$ and $\mathbf{y}$ and $\mathbf{c}_0$ the vectors of system inputs (regarded as a known deterministic sequence, e.g. a particular realization of a stochastic process), noisy outputs, and fixed initial conditions, respectively, and let $\widehat{\mathbf{x}}(\mathbf{y}, \boldsymbol{u}, \mathbf{c}_0)$ be an estimator of a real-valued random parameter vector $\mathbf{x}$ based on $(\mathbf{y}, \boldsymbol{u}, \mathbf{c}_0)$. Then, under weak regularity conditions, one can show that (van Trees, 1968)

$$\mathrm{E}[(\widehat{\mathbf{x}}(\mathbf{y}, \boldsymbol{u}, \mathbf{c}_0) - \mathbf{x})(\widehat{\mathbf{x}}(\mathbf{y}, \boldsymbol{u}, \mathbf{c}_0) - \mathbf{x})^{\mathrm{T}} | \boldsymbol{u}, \mathbf{c}_0] \geq \mathbf{J}^{-1}(\boldsymbol{u}, \mathbf{c}_0)$$

where

$$\begin{aligned}\mathbf{J}(\boldsymbol{u}, \mathbf{c}_0) &= -\mathrm{E}\left[\frac{\partial^2 \log p(\mathbf{y}, \mathbf{x}|\boldsymbol{u}, \mathbf{c}_0)}{\partial \mathbf{x} \partial \mathbf{x}^{\mathrm{T}}}\right] \\ &= \mathrm{E}\left[\frac{\partial \log p(\mathbf{y}, \mathbf{x}|\boldsymbol{u}, \mathbf{c}_0)}{\partial \mathbf{x}} \frac{\partial \log p(\mathbf{y}, \mathbf{x}|\boldsymbol{u}, \mathbf{c}_0)}{\partial \mathbf{x}^{\mathrm{T}}}\right]\end{aligned}$$

and $p(\mathbf{y}, \mathbf{x}|\boldsymbol{u}, \mathbf{c}_0) = p(\mathbf{y}|\mathbf{x}, \boldsymbol{u}, \mathbf{c}_0) p(\mathbf{x})$ is the joint probability density function of the pair $(\mathbf{y}, \mathbf{x})$ given $\boldsymbol{u}$ and $\mathbf{c}_0$. When the input signal is a stochastic process, initial conditions are random, and averaging is extended to all realizations of $\boldsymbol{u}$ and $\mathbf{c}_0$, one obtains the following result

$$\begin{aligned}&\mathrm{E}[(\widehat{\mathbf{x}}(\mathbf{y}, \boldsymbol{u}, \mathbf{c}_0) - \mathbf{x})(\widehat{\mathbf{x}}(\mathbf{y}, \boldsymbol{u}, \mathbf{c}_0) - \mathbf{x})^{\mathrm{T}}] \geq \mathrm{E}[\mathbf{J}^{-1}(\boldsymbol{u}, \mathbf{c}_0)] \\ &\geq \{\mathrm{E}[\mathbf{J}(\boldsymbol{u}, \mathbf{c}_0)]\}^{-1} = \bar{\mathbf{J}}^{-1}\end{aligned}$$

where the second transition stems from the Jensen's inequality for matrices – see Olkin and Pratt (1958).

In the case considered, let $\mathbf{x}_t = [\alpha_0(1), \ldots, \alpha_0(t)]^{\mathrm{T}}$, $\mathbf{y}_t = [y(1), \ldots, y(t)]^{\mathrm{T}}$ and $\boldsymbol{u}_t = [\boldsymbol{\varphi}^{\mathrm{T}}(1), \ldots, \boldsymbol{\varphi}^{\mathrm{T}}(t)]^{\mathrm{T}}$.

To simplify further analysis, we will assume, in addition to (A1)–(A5), that the complex-valued 'amplitudes' can be written down in the form $\boldsymbol{\beta}_k = \boldsymbol{\beta}_{k0} e^{j\nu_k}$, $k = 1, 2, \ldots, K$, where $\boldsymbol{\beta}_{k0}$ are fixed (deterministic) complex-valued vectors and ν_k are mutually independent random phase shifts, distributed uniformly over the interval $[0,\ 2\pi)$. Note that under the last assumption it holds that $\mathrm{E}[\boldsymbol{\beta}_k \boldsymbol{\beta}_l^{\mathrm{H}}] = \mathbf{O}$, $\forall k \neq l$. Furthermore, we will assume that $\alpha_0(0)$ is uniformly distributed over $[\alpha_{\min}, \alpha_{\max}]$, and that $\omega_0(0)$ is a known deterministic quantity. Hence, the vector of initial conditions can be specified as $\mathbf{c}_0 = [\boldsymbol{\beta}_{10}, \ldots, \boldsymbol{\beta}_{K0},\ \nu_1, \ldots, \nu_K, \alpha_0(0), \omega_0(0)]^{\mathrm{T}}$, where the quantities $\boldsymbol{\beta}_{10}, \ldots, \boldsymbol{\beta}_{K0}$ and $\omega_0(0)$ are deterministic, and the quantities $\nu_1, \ldots, \nu_K$ and $\alpha_0(0)$ are stochastic.

[3] When the estimated quantities are stochastic variables, rather than unknown deterministic constants, the classical Cramér-Rao inequality does not apply.

First of all, note that

$$\log p(\mathbf{y}_t, \mathbf{x}_t|\boldsymbol{u}_t, \mathbf{c}_0) = \log p(\mathbf{y}_t|\mathbf{x}_t, \boldsymbol{u}_t, \mathbf{c}_0) + \log p(\mathbf{x}_t).$$

Since, under the assumptions listed above, the vectors $\mathbf{x}_t$ and $\mathbf{c}_0$ fully determine $\boldsymbol{\theta}(\tau)$, $\tau = 1, \ldots, t$: $\boldsymbol{\theta}(\tau) = \sum_{k=1}^{K} \boldsymbol{\beta}_k e^{jm_k \phi_0(\tau)}$, where

$$\phi_0(\tau) = \sum_{n=1}^{\tau} \omega_0(n) = \omega_0(0) + \sum_{n=1}^{\tau} \sum_{m=1}^{n-1} \alpha_0(m), \qquad (38)$$

one arrives at (for normally distributed $\{v(\tau)\}$)

$$\begin{aligned}&\log p(\mathbf{y}_t|\mathbf{x}_t, \boldsymbol{u}_t, \mathbf{c}_0) = \log p[\mathbf{y}_t|\boldsymbol{\theta}(1), \boldsymbol{\varphi}(1), \ldots, \boldsymbol{\theta}(t), \boldsymbol{\varphi}(t)] \\ &= c_1 - \frac{1}{\sigma_v^2} \sum_{\tau=1}^{t} |v(\tau)|^2 = c_1 - \frac{1}{\sigma_v^2} \sum_{\tau=1}^{t} |y(\tau) - \boldsymbol{\varphi}^{\mathrm{T}}(\tau) \boldsymbol{\theta}(\tau)|^2\end{aligned} \qquad (39)$$

where c_1 is a constant independent of $\mathbf{x}_t$.

Differentiating (39) with respect to $\alpha_0(m)$, one obtains

$$\begin{aligned}&\frac{\partial \log p(\mathbf{y}_t|\mathbf{x}_t, \boldsymbol{u}_t, \mathbf{c}_0)}{\partial \alpha_0(m)} \\ &= \frac{2}{\sigma_v^2} \sum_{\tau=1}^{t} \mathrm{Re}\left\{ [y(\tau) - \boldsymbol{\varphi}^{\mathrm{T}}(\tau) \boldsymbol{\theta}(\tau)]^* \boldsymbol{\varphi}^{\mathrm{T}}(\tau) \frac{\partial \boldsymbol{\theta}(\tau)}{\partial \alpha_0(m)} \right\}.\end{aligned}$$

and

$$\begin{aligned}&\frac{\partial^2 \log p(\mathbf{y}_t|\mathbf{x}_t, \boldsymbol{u}_t, \mathbf{c}_0)}{\partial \alpha_0(m) \partial \alpha_0(n)} \\ &= \frac{2}{\sigma_v^2} \sum_{\tau=1}^{t} \mathrm{Re}\left\{ -\frac{\partial \boldsymbol{\theta}^{\mathrm{H}}(\tau)}{\partial \alpha_0(n)} \boldsymbol{\varphi}^*(\tau) \boldsymbol{\varphi}^{\mathrm{T}}(\tau) \frac{\partial \boldsymbol{\theta}(\tau)}{\partial \alpha_0(m)} \right. \\ &\left. + [y(\tau) - \boldsymbol{\varphi}^{\mathrm{T}}(\tau) \boldsymbol{\theta}(\tau)]^* \boldsymbol{\varphi}^{\mathrm{T}}(\tau) \frac{\partial^2 \boldsymbol{\theta}(\tau)}{\partial \alpha_0(m) \partial \alpha_0(n)} \right\} \\ &= \frac{2}{\sigma_v^2} \sum_{\tau=1}^{t} \mathrm{Re}\left\{ -\frac{\partial \boldsymbol{\theta}^{\mathrm{H}}(\tau)}{\partial \alpha_0(n)} \boldsymbol{\varphi}^*(\tau) \boldsymbol{\varphi}^{\mathrm{T}}(\tau) \frac{\partial \boldsymbol{\theta}(\tau)}{\partial \alpha_0(m)} \right. \\ &\left. + v^*(t) \boldsymbol{\varphi}^{\mathrm{T}}(\tau) \frac{\partial^2 \boldsymbol{\theta}(\tau)}{\partial \alpha_0(m) \partial \alpha_0(n)} \right\}.\end{aligned}$$

where the last transition stems from $y(\tau) = \boldsymbol{\varphi}^{\mathrm{T}}(\tau) \boldsymbol{\theta}(\tau) + v(\tau)$.

This leads to

$$\begin{aligned}&\mathrm{E}\left[\frac{\partial^2 \log p(\mathbf{y}_t|\mathbf{x}_t, \boldsymbol{u}_t, \mathbf{c}_0)}{\partial \alpha_0(m) \partial \alpha_0(n)}\right] \\ &= -\frac{2}{\sigma_v^2} \sum_{\tau=1}^{t} \mathrm{Re}\left\{ \mathrm{E}\left[\frac{\partial \boldsymbol{\theta}^{\mathrm{H}}(\tau)}{\partial \alpha_0(n)} \boldsymbol{\varphi}^*(\tau) \boldsymbol{\varphi}^{\mathrm{T}}(\tau) \frac{\partial \boldsymbol{\theta}(\tau)}{\partial \alpha_0(m)}\right]\right\} \\ &= -\frac{2}{\sigma_v^2} \sum_{\tau=1}^{t} \mathrm{E}\left[\frac{\partial \boldsymbol{\theta}^{\mathrm{H}}(\tau)}{\partial \alpha_0(n)} \boldsymbol{\Phi} \frac{\partial \boldsymbol{\theta}(\tau)}{\partial \alpha_0(m)}\right].\end{aligned}$$

Using (38), one arrives at

$$\frac{\partial \boldsymbol{\theta}(\tau)}{\partial \alpha_0(m)} = j \sum_{k=1}^{K} m_k \boldsymbol{\beta}_k e^{j m_k \phi_0(\tau)} \sum_{s=1}^{k} \sum_{l=1}^{s-1} \delta_{\tau,l}$$
$$= j \sum_{k=1}^{K} m_k \boldsymbol{\beta}_k e^{j m_k \phi_0(\tau)} \max(\tau - m, 0)$$

where $\delta_{\tau,l} = \{0 \text{ if } \tau \neq l, 1 \text{ if } \tau = l\}$ denotes the Kronecker delta. This allows one to reach

$$\mathrm{E}\left[\frac{\partial \boldsymbol{\theta}^{\mathrm{H}}(\tau)}{\partial \alpha_0(n)} \boldsymbol{\Phi} \frac{\partial \boldsymbol{\theta}(\tau)}{\partial \alpha_0(m)}\right]$$
$$= -\sum_{k=1}^{K} m_k^2 \boldsymbol{\beta}_k^{\mathrm{H}} \boldsymbol{\Phi} \boldsymbol{\beta}_k \max(\tau - m, 0) \max(\tau - n, 0)$$
$$= -\sum_{k=1}^{K} m_k^2 b_k^2 \max(\tau - m, 0) \max(\tau - n, 0) \,,$$

where the crossterms $\boldsymbol{\beta}_k^{\mathrm{H}} \boldsymbol{\Phi} \boldsymbol{\beta}_l = e^{j(\nu_l - \nu_k)} \boldsymbol{\beta}_{k0}^{\mathrm{H}} \boldsymbol{\Phi} \boldsymbol{\beta}_{l0}$, $k \neq l$ average to zero due to independence of the phase shifts ν_k and ν_l.
Therefore

$$-\mathrm{E}\left[\frac{\partial^2 \log p(\mathbf{y}_t | \mathbf{x}_t, \boldsymbol{u}_t, \mathbf{c}_0)}{\partial \mathbf{x}_t \partial \mathbf{x}_t^{\mathrm{T}}}\right] = \frac{2b_0^2}{\sigma_v^2} \mathbf{A}_t \qquad (40)$$

where $[\mathbf{A}_t]_{mn} = \sum_{\tau=1}^{t} \max(\tau - m,\ 0) \cdot \max(\tau - n,\ 0)$.

In an analogous way, one can derive the second component of the generalized Fisher matrix. First, note that

$$\log p(\mathbf{x}_t)$$
$$= \log p[\ \alpha_0(1), \alpha_0(2) - \alpha_0(1), \ldots, \alpha_0(t) - \alpha_0(t-1)\ ]$$
$$= c_2 + c_3 - \frac{1}{2\sigma_\delta^2} \sum_{\tau=2}^{t} w^2(\tau)$$
$$= c_2 + c_3 - \frac{1}{2\sigma_\delta^2} \sum_{\tau=2}^{t} [\alpha_0(\tau) - \alpha_0(\tau-1)]^2 \qquad (41)$$

where $c_2 = \log[1/(\alpha_{\max} - \alpha_{\min})]$ and c_3 are constants independent of $\mathbf{x}_t$.
Differentiation of (41) results in

$$\frac{\log p(\mathbf{x}_t)}{\alpha_0(m)} = -\frac{1}{2\sigma_\delta^2} \sum_{\tau=2}^{t} \frac{\partial}{\partial \alpha_0(m)} [\alpha_0(\tau) - \alpha_0(\tau-1)]^2$$
$$= \frac{1}{\sigma_\delta^2} \begin{cases} \alpha_0(2) - \alpha_0(1) & \text{for} \quad m = 1 \\ \alpha_0(m+1) - 2\alpha_0(m) + \alpha_0(t-1) & \text{for} \ \ 1 < m < t \\ \alpha_0(t-1) - \alpha_0(t) & \text{for} \quad m = t \end{cases}$$
$$= \frac{1}{\sigma_\delta^2} \begin{cases} \delta(2) & \text{for} \quad m = 1 \\ \delta(m+1) - \delta(m) & \text{for} \ \ 1 < m < t \\ -\delta(t) & \text{for} \quad m = t \end{cases}$$

which leads to

$$\mathrm{E}\left[\frac{\partial \log p(\mathbf{x}_t)}{\partial \mathbf{x}_t} \frac{\partial \log p(\mathbf{x}_t)}{\partial \mathbf{x}_t^{\mathrm{T}}}\right] = \frac{1}{\sigma_\delta^2} \mathbf{B}_t \qquad (42)$$

where

$$\mathbf{B}_t = \begin{bmatrix} 1 & -1 & 0 & 0 & \ldots & 0 \\ -1 & 2 & -1 & 0 & \ldots & 0 \\ 0 & -1 & 2 & -1 & \ldots & 0 \\ & & & \ddots & & \\ 0 & \ldots & 0 & -1 & 2 & -1 \\ 0 & \ldots & 0 & 0 & -1 & 1 \end{bmatrix} .$$

Combining (40) with (42), one obtains

$$\bar{\mathbf{J}}_t = \frac{2b_0^2}{\sigma_v^2} \mathbf{A}_t + \frac{1}{\sigma_\delta^2} \mathbf{B}_t \ = \frac{1}{\sigma_\delta^2} \left[2\kappa \mathbf{A}_t + \mathbf{B}_t\right].$$

The asymptotic (steady-state) bounds on accuracy of frequency and frequency rate estimates can be obtained from

$$\mathrm{LTB}_{\omega_0} = \lim_{t \mapsto \infty} \inf_{\widehat{\omega}_0(\cdot)} \mathrm{E}\{\ [\omega_0(t) - \widehat{\omega}_0(t)]^2\} = \lim_{t \mapsto \infty} \mathbf{b}_t^{\mathrm{T}} \mathbf{J}_t^{-1} \mathbf{b}_t$$

$$\mathrm{LTB}_{\alpha_0} = \lim_{t \mapsto \infty} \inf_{\widehat{\alpha}_0(\cdot)} \mathrm{E}\{\ [\alpha_0(t) - \widehat{\alpha}_0(t)]^2\} = \lim_{t \mapsto \infty} \left[\mathbf{J}_t^{-1}\right]_{tt}$$

where $\mathbf{b}_t^{\mathrm{T}} = [\mathbf{1}_{t-1}^{\mathrm{T}}, 0]$, and $\mathbf{1}_t$ denotes the vector of ones of length t. The analogous expressions for lower smoothing bounds read

$$\mathrm{LSB}_{\omega_0} = \lim_{t \mapsto \infty} \inf_{\widetilde{\omega}_0(\cdot)} \mathrm{E}\{\ [\omega_0(t) - \widetilde{\omega}_0(t)]^2\} = \lim_{t \mapsto \infty} \mathbf{c}_t^{\mathrm{T}} \mathbf{J}_{2t}^{-1} \mathbf{c}_t$$

$$\mathrm{LSB}_{\alpha_0} = \lim_{t \mapsto \infty} \inf_{\widetilde{\alpha}_0(\cdot)} \mathrm{E}\{\ [\alpha_0(t) - \widetilde{\alpha}_0(t)]^2\} = \lim_{t \mapsto \infty} \left[\mathbf{J}_{2t}^{-1}\right]_{tt}$$

where $\mathbf{c}_t^{\mathrm{T}} = [\mathbf{1}_{t-1}^{\mathrm{T}}, \mathbf{0}_{t+1}^{\mathrm{T}}]$ and $\mathbf{0}_t$ denotes the vector of zeros of length t. The values of LTB and LSB, shown in Table 1, were computed numerically for t ranging from 100 to 600 (the convergence is slower for smaller values of κ).