Tracking Analysis of a Generalized Adaptive Notch Filter

Maciej Niedźwiecki and Piotr Kaczmarek

***Abstract*—The paper presents results of local performance analysis of a generalized adaptive notch filter (GANF). GANFs are used for identification/tracking of quasi-periodically varying dynamic systems and can be considered an extension, to the system case, of classical adaptive notch filters. The tracking properties of the algorithm are studied analytically using a direct averaging approach and an approximating linear filter technique. Even though restricted to a single-frequency case, the presented analysis provides valuable insights into the tracking mechanisms of GANF, including the associated speed/accuracy tradeoffs, the achievable performance bounds, and tracking limitations. In addition, it allows one to formulate some useful rules of thumb for choosing design parameters.**

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

I. Introduction

A. Problem Statement

GENERALIZED adaptive notch filters (GANFs) [1]–[3], were designed for the purpose of identification/tracking of quasi-periodically varying complex-valued systems, i.e., systems governed by

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

where $t=1,2,\dots$ denotes the normalized discrete time, $y(t)$ denotes the system output, $\boldsymbol{\varphi}(t)=[\varphi_1(t),\dots,\varphi_n(t)]^{\mathrm{T}}$ is the regression vector, $v(t)$ is an additive noise and $\boldsymbol{\theta}(t)=[\theta_1(t),\dots,\theta_n(t)]^{\mathrm{T}}$ denotes the vector of time varying coefficients, modeled as weighted sums of complex exponentials

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

All quantities in (1) and (2), except angular frequencies $\omega_1(t),\dots,\omega_k(t)$, are complex valued. Since the complex amplitudes $a_{li}(t)$ incorporate both magnitude and phase information, there is no explicit phase component in (2).

It will be assumed that both the amplitudes $a_{li}(t), l=1,\dots,n$ and frequencies $\omega_i(t)$ in (2) are slowly time varying and that $v(t)=v_R(t)+jv_I(t), \mathrm{E}[v_R^2(t)]=\mathrm{E}[v_r^2(t)]=\sigma_v^2/2$, $\mathrm{E}[v_R(y)v_I(t)]=0, \forall t$, is a complex white noise of variance σ_v^2, independent of the sequence of regression vectors $\boldsymbol{\varphi}(t)$.

One of the possible interesting applications, which admits such problem formulation, is identification of multipath (e.g., mobile radio) channels—see e.g., [4]–[6]. In this particular case the regression vector $\boldsymbol{\varphi}(t)$ is made up of past input (transmitted) symbols, $y(t)$ is the received baseband signal, $\boldsymbol{\theta}(t)$ is the vector of time varying impulse response coefficients of the channel, 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).

B. Generalized Adaptive Notch Filter

Denote by $\boldsymbol{\alpha}_i(t)=[a_{1i}(t),\dots,a_{ni}(t)]^{\mathrm{T}}$ the vector of system coefficients associated with a particular frequency ω_i. Similarly, let $\boldsymbol{\psi}_i(t)=f_i(t)\boldsymbol{\varphi}(t)$, where $f_i(t)=e^{j\sum_{s=1}^{t}\omega_i(s)}$, be the generalized regression vector associated with the ith frequency component. Using the short-hand notation introduced above, (1) and (2) can be rewritten in the form

$$y(t)=\sum_{i=1}^{k}\boldsymbol{\psi}_i^{\mathrm{T}}(t)\boldsymbol{\alpha}_i(t)+v(t)$$
$$\boldsymbol{\theta}(t)=\sum_{i=1}^{k}f_i(t)\boldsymbol{\alpha}_i(t).$$

From different algorithms capable of tracking both complex amplitudes and frequencies in a system governed by (1) and (2), we have chosen a relatively simple solution described in [3], which combines the exponentially weighted least squares approach to amplitude tracking with gradient search approach to frequency tracking

$$\begin{aligned}
\hat{f}_i(t)&=e^{j\hat{\omega}_i(t)}\hat{f}_i(t-1)\\
\hat{\boldsymbol{\psi}}_i(t)&=\hat{f}_i(t)\boldsymbol{\varphi}(t)\\
i&=1,\dots,k\\
\varepsilon(t)&=y(t)-\hat{\boldsymbol{\psi}}^{\mathrm{T}}(t)\hat{\boldsymbol{\alpha}}(t-1)\\
\mathbf{Q}(t)&=\frac{1}{\lambda}\left[\mathbf{Q}(t-1)-\frac{\mathbf{Q}(t-1)\hat{\boldsymbol{\psi}}(t)\hat{\boldsymbol{\psi}}^{\mathrm{H}}(t)\mathbf{Q}(t-1)}{\lambda+\hat{\boldsymbol{\psi}}^{\mathrm{H}}(t)\mathbf{Q}(t-1)\hat{\boldsymbol{\psi}}(t)}\right]\\
\mathbf{k}(t)&=\mathbf{Q}(t)\hat{\boldsymbol{\psi}}(t)\\
\hat{\boldsymbol{\alpha}}(t)&=\hat{\boldsymbol{\alpha}}(t-1)+\mathbf{k}^*(t)\varepsilon(t)\\
g_i(t)&=\mathrm{Im}\left\{\varepsilon^*(t)\hat{\boldsymbol{\psi}}_i^{\mathrm{T}}(t)\hat{\boldsymbol{\alpha}}_i(t-1)\right\}\\
\hat{\omega}_i(t+1)&=\hat{\omega}_i(t)-\eta g_i(t)\\
i&=1,\dots,k\\
\hat{\boldsymbol{\theta}}(t)&=\sum_{i=1}^{k}\hat{f}_i(t)\hat{\boldsymbol{\alpha}}_i(t)
\end{aligned} \quad (3)$$

Manuscript received January 12, 2005; revised March 11, 2005. 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. Jonathon Chambers.

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.2005.861087

where $\hat{\boldsymbol{\alpha}}(t) = [\hat{\boldsymbol{\alpha}}_1^{\mathrm{T}}(t), \ldots, \hat{\boldsymbol{\alpha}}_k^{\mathrm{T}}(t)]^{\mathrm{T}}$ and $\hat{\boldsymbol{\psi}}(t) = [\hat{\boldsymbol{\psi}}_1^{\mathrm{T}}(t), \ldots, \hat{\boldsymbol{\psi}}_k^{\mathrm{T}}(t)]^{\mathrm{T}}$.

In the above algorithm λ $(0 < \lambda < 1)$, usually set close to one, denotes the so-called forgetting constant, which controls the rate of amplitude adaptation, and $\eta > 0$, usually set close to zero, denotes the stepsize coefficient, which controls the rate of frequency adaptation.

The initial conditions for (3) should be set to $\hat{\boldsymbol{\alpha}}(0) = \mathbf{0}$ and $\mathbf{Q}(0) = c\mathbf{I}_{kn}$, where $\mathbf{I}_{kn}$ denotes the $kn \times kn$ identity matrix and c is a large positive constant—this is a standard initialization procedure for all recursive least-squares estimation algorithms [7].

The problem of identification of quasi-periodically varying systems can be considered a generalization, to the system case, of a classical signal processing task of either elimination or extraction of nonstationary sinusoidal signals buried in noise. Actually, note that when $n = 1$ and $\varphi(t) = 1, \forall t$ the model (1) and (2) becomes a description of a noisy nonstationary multifrequency signal $s(t) = \theta(t)$

$$y(t) = s(t) + v(t) = \sum_{i=1}^{k} a_i(t) e^{j\sum_{s=1}^{t} \omega_i(s)} + v(t). \qquad (4)$$

In this special case, the GANF (3) becomes an "ordinary" adaptive notch filter[1]

$$\begin{aligned}
\hat{f}_i(t) &= e^{j\hat{\omega}_i(t)} \hat{f}_i(t-1) \\
i &= 1, \ldots, k \\
\varepsilon(t) &= y(t) - \hat{\mathbf{f}}^{\mathrm{T}}(t)\hat{\boldsymbol{\alpha}}(t-1) \\
\mathbf{Q}(t) &= \frac{1}{\lambda}\left[\mathbf{Q}(t-1) - \frac{\mathbf{Q}(t-1)\hat{\mathbf{f}}(t)\hat{\mathbf{f}}^{\mathrm{H}}(t)\mathbf{Q}(t-1)}{\lambda + \hat{\mathbf{f}}^{\mathrm{H}}(t)\mathbf{Q}(t-1)\hat{\mathbf{f}}(t)}\right] \\
\mathbf{k}(t) &= \mathbf{Q}(t)\hat{\mathbf{f}}(t) \\
\hat{\boldsymbol{\alpha}}(t) &= \hat{\boldsymbol{\alpha}}(t-1) + \mathbf{k}^*(t)\varepsilon(t) \\
g_i(t) &= \mathrm{Im}\{\varepsilon^*(t)\hat{f}_i(t)\hat{a}_i(t-1)\} \\
\hat{\omega}_i(t+1) &= \hat{\omega}_i(t) - \eta g_i(t) \\
i &= 1, \ldots, k \\
\hat{s}(t) &= \sum_{i=1}^{k} \hat{f}_i(t)\hat{a}_i(t)
\end{aligned} \qquad (5)$$

where $\hat{\boldsymbol{\alpha}}(t) = [\hat{a}_1(t), \ldots, \hat{a}_k(t)]^{\mathrm{T}}$ and $\hat{\mathbf{f}}(t) = [\hat{f}_1(t), \ldots, \hat{f}_k(t)]^{\mathrm{T}}$.

The problem of elimination and extraction of complex sinusoidal signals (called cisoids) buried in noise was considered by many authors (see e.g., [8] and [9], and the references therein). Interestingly, the signal processing algorithm (5) does not coincide with any of the existing solutions.

C. Contribution and Novelty

The main purpose of our paper is analysis of tracking capabilities of the GANF (3) and its signal-oriented counterpart (5). We will relate performance of both algorithms to the system/signal characteristics, such as the signal-to-noise ratio (SNR) and the rate of frequency variation, as well as to the user-dependent quantities such as the forgetting constant λ and the step size η. These analytical results will allow us to rationalize the choice of design variables and to specify conditions under which the system/signal parameters can be tracked successfully.

[1]By adaptive notch filter, we mean here any notch filtering algorithm that is capable of tracking the signal frequency changes. In the notch filtering literature, the term "adaptive notch filter" (ANF) is often reserved for a special class of filters with constrained poles and zeros.

The paper is organized as follows. In Section II, we use the approximating linear filter technique, developed by Tichavský and Händel in their seminal paper [8], to analyze performance of the signal processing algorithm (5) in a neighborhood of its equilibrium state. Even though heuristic and restricted to a single-frequency case ($k = 1$), this analysis provides valuable insights into the tracking mechanisms, including the speed/accuracy tradeoffs, the achievable performance bounds, and tracking limitations. Based on these results, we formulate some useful tuning rules. The results of linear filter approximation allow us also to compare the algorithm (5) with several other known solutions to the problem of adaptive notch filtering. In Section III, we show that most of the conclusions reached for the signal processing algorithm (5) can be extended to the system identification algorithm (3). Section IV presents the results of simulation experiments, and Section V concludes.

II. Analysis of Signal Identification Algorithm

A. Reparametrized Algorithm

Before we start analyzing tracking properties of the signal identification algorithm (5), we will convert it into a more convenient "fixed-basis" form by applying the linear time-varying transformation

$$\begin{aligned}
\hat{\boldsymbol{\beta}}(t) &= \hat{\mathbf{F}}(t)\hat{\boldsymbol{\alpha}}(t), \quad \mathbf{l}(t) = \hat{\mathbf{F}}^*(t)\mathbf{k}(t) \\
\mathbf{P}(t) &= \hat{\mathbf{F}}^*(t)\mathbf{Q}(t)\hat{\mathbf{F}}(t)
\end{aligned} \qquad (6)$$

where $\hat{\mathbf{F}}(t) = \mathrm{diag}\{\hat{f}_1(t), \ldots, \hat{f}_k(t)\}$. (See [10] for more details on interpretation and advantages of the fixed-basis parametrization of nonstationary systems.)

Using (6) and $\hat{\mathbf{A}}(t) = \mathrm{diag}\{e^{j\hat{\omega}_1(t)}, \ldots, e^{j\hat{\omega}_k(t)}\}$, one can rewrite (5) in the following form:

$$\begin{aligned}
\varepsilon(t) &= y(t) - \mathbf{1}_k^{\mathrm{T}}\hat{\mathbf{A}}(t)\hat{\boldsymbol{\beta}}(t-1) \\
\mathbf{P}(t) &= \\
&\frac{1}{\lambda}\hat{\mathbf{A}}^*(t)\left[\mathbf{P}(t-1) - \frac{\mathbf{P}(t-1)\hat{\mathbf{A}}(t)\mathbf{1}_k\mathbf{1}_k^{\mathrm{T}}\hat{\mathbf{A}}^*(t)\mathbf{P}(t-1)}{\lambda + \mathbf{1}_k^{\mathrm{T}}\hat{\mathbf{A}}^*(t)\mathbf{P}(t-1)\hat{\mathbf{A}}(t)\mathbf{1}_k}\right]\hat{\mathbf{A}}(t) \\
\mathbf{l}(t) &= \mathbf{P}(t)\mathbf{1}_k \\
\hat{\boldsymbol{\beta}}(t) &= \hat{\mathbf{A}}(t)\hat{\boldsymbol{\beta}}(t-1) + \mathbf{l}^*(t)\varepsilon(t) \\
g_i(t) &= \mathrm{Im}\{\varepsilon^*(t)e^{j\hat{\omega}_i(t)}\hat{\beta}_i(t-1)\} \\
\hat{\omega}_i(t+1) &= \hat{\omega}_i(t) - \eta g_i(t) \\
i &= 1, \ldots, k \\
\hat{s}(t) &= \sum_{i=1}^{k} \hat{\beta}_i(t)
\end{aligned} \qquad (7)$$

where $\hat{\boldsymbol{\beta}}(t) = [\hat{\beta}_1(t), \ldots, \hat{\beta}_k(t)]^{\mathrm{T}}, \hat{\beta}_i(t) = \hat{f}_i(t)\hat{a}_i(t)$, $i = 1, \ldots, k$ and $\mathbf{1}_k = [\underbrace{1, \ldots, 1}_{k}]^{\mathrm{T}}$.

It should be emphasized that the algorithms (5) and (7) are strictly input–output equivalent, i.e., when started with the same initial conditions $(\boldsymbol{\beta}(0) = \boldsymbol{\alpha}(0), \mathbf{P}(0) = \mathbf{Q}(0))$, they yield identical signal estimates $\hat{s}(t)$.

As is straightforward to check, the algorithm (7) is almost identical with the algorithm known as multiple frequency tracker (MFT), proposed by Tichavský and Händel (see [8, (9)–(11)]). It turns out that the only difference lies in the frequency update mechanism, which in the case of MFT has the form (in our notation)

$$\begin{aligned} g_i(t) &= \mathrm{Arg}\left[\frac{\hat{\beta}_i(t)}{\hat{\beta}_i(t-1)e^{j\hat{\omega}_i(t)}}\right] \\ \hat{\omega}_i(t+1) &= \hat{\omega}_i(t) - \eta g_i(t) \\ i &= 1,\ldots,k. \end{aligned}$$

In the paper [9], which is a follow-up of [8], Tichavský and Nehorai analyzed four different frequency tracking algorithms: the multiple frequency tracker (summarized above), the adaptive notch filter with constrained poles and zeros (a variant of the algorithms presented, e.g., in [11] and [12], adapted for complex-valued signals), the hyperstable adaptive line enhancer [13]–[15], and the adaptive infinite-impulse response (IIR) filter described in [16]. Using the approximating linear filter (ALF) technique, introduced in [8], Tichavský and Nehorai have shown that local properties of all four algorithms are approximately the same. Approximating linear filters characterize the relation between the sequences of estimation errors and the sequences of measurement noise $v(t)$ and of the one-step changes of the true frequency $\omega(t+1)-\omega(t)$, provided that the analyzed algorithms operate in a neighborhood of their equilibrium state. Although the ALF-based analysis is *heuristic* (no strict mathematical conditions of its applicability were given in [8]), it provides results that are insightful and stay in very good agreement with simulation experiments. We will use the same tools to analyze (7).

B. Analysis for a Single Cisoid

Similarly as in [8], we will consider a single frequency case ($k = 1$) and steady state tracking conditions. Note that for $k = 1$, the scalar (1×1) counterpart of the matrix $\mathbf{P}(t)$, denoted by $\mathrm{p}(t)$, tends to a constant steady-state value $\mathrm{p}(\infty) = \lim_{t\mapsto\infty} \mathrm{p}(t) = 1-\lambda$. Hence, in the case considered, one can rewrite (7) in a much simpler form

$$\begin{aligned} \varepsilon(t) &= y(t) - e^{j\hat{\omega}(t)}\hat{\beta}(t-1) \\ \hat{\beta}(t) &= e^{j\hat{\omega}(t)}\hat{\beta}(t-1) + \mu\varepsilon(t) \\ g(t) &= \mathrm{Im}\{\varepsilon^*(t)e^{j\hat{\omega}(t)}\hat{\beta}(t-1)\} \\ \hat{\omega}(t+1) &= \hat{\omega}(t) - \eta g(t) \\ \hat{s}(t) &= \hat{\beta}(t) \end{aligned} \tag{8}$$

where $\mu = 1-\lambda$.

For MFT the analogous equations are identical with (8), except that

$$g(t) = \mathrm{Arg}\left[\frac{\hat{\beta}(t)}{\hat{\beta}(t-1)e^{j\hat{\omega}(t)}}\right]. \tag{9}$$

We will confine our study to a constant magnitude signal model

$$s(t) = \beta(t) = \beta(t-1)e^{j\omega(t)}, \quad |\beta(t)| = b. \tag{10}$$

Note that using (10), the noisy signal $y(t)$ can be expressed in the following nonrecursive form (cf. (4)):

$$y(t) = s(t) + v(t) = ae^{j\sum_{s=1}^{t}\omega_i(s)} + v(t)$$

where $a = \beta(0)$.

Denote by $\Delta\hat{\beta}(t) = \hat{\beta}(t) - \beta(t) = \hat{s}(t) - s(t)$ and $\Delta\hat{\omega}(t) = \hat{\omega}(t) - \omega(t)$ the signal estimation error and frequency estimation error, respectively. Let

$$\begin{aligned} \Delta\hat{\phi}(t) &= \beta^*(t)\Delta\hat{\beta}(t) = \Delta\hat{\phi}_R(t) + j\Delta\hat{\phi}_I(t) \\ e(t) &= \beta^*(t)v(t) = e_R(t) + je_I(t) \\ w(t+1) &= \omega(t+1) - \omega(t). \end{aligned} \tag{11}$$

It is straightforward to check that $|\Delta\hat{\phi}(t)|^2 = (\Delta\hat{\phi}_R(t))^2 + (\Delta\hat{\phi}_I(t))^2 = b^2|\Delta\hat{\beta}(t)|^2$ and that $e(t)$, similarly as $v(t)$, is a complex-valued white noise obeying $\sigma_e^2 = \mathrm{E}[|e(t)|^2] = b^2\sigma_v^2$, $\mathrm{E}[e_R^2(t)] = \mathrm{E}[e_I^2(t)] = \sigma_e^2/2$, $\mathrm{E}[e_R(t)e_I(t)] = 0$.

Using the technique proposed in [8], the following result can be proved.

Proposition 1: Assume that the sequences $\{e(t)\}$ and $\{w(t)\}$ are uniformly small so that one can neglect higher than first-order moments of their elements. Then, the algorithm (8) applied to signal

$$y(t) = \beta(t) + v(t), \quad \beta(t) = e^{j\omega(t)}\beta(t-1) \tag{12}$$

can be approximately described by the following linear filtering equations

$$\begin{aligned} \Delta\hat{\phi}_R(t) &= \lambda\Delta\hat{\phi}_R(t-1) + \mu e_R(t) \\ \Delta\hat{\phi}_I(t) &= \lambda\Delta\hat{\phi}_I(t-1) + \lambda b^2\Delta\hat{\omega}(t) + \mu e_I(t) \\ \Delta\hat{\omega}(t+1) &= \delta\Delta\hat{\omega}(t) - \eta\Delta\hat{\phi}_I(t-1) + \eta e_I(t) - w(t+1) \end{aligned} \tag{13}$$

where $\delta = 1-\eta b^2$.

Derivation: See Appendix I.

Solving the approximating linear equations (13) with respect to $\Delta\hat{\phi}_R(t)$, $\Delta\hat{\phi}_I(t)$, and $\Delta\hat{\omega}(t)$, one obtains

$$\begin{aligned} \Delta\hat{\phi}_R(t) &= F(q^{-1})e_R(t) \\ \Delta\hat{\phi}_I(t) &= G_1(q^{-1})e_I(t) + G_2(q^{-1})w(t) \\ \Delta\hat{\omega}(t) &= H_1(q^{-1})e_I(t) + H_2(q^{-1})w(t) \end{aligned} \tag{14}$$

where q^{-1} denotes the backward shift operator and

$$\begin{aligned} F(q^{-1}) &= \frac{1-\lambda}{1-\lambda q^{-1}} \\ G_1(q^{-1}) &= \frac{1-\lambda+(\lambda-\delta)q^{-1}}{1-(\lambda+\delta)q^{-1}+\lambda q^{-2}} \\ G_2(q^{-1}) &= -\frac{b^2\lambda}{1-(\lambda+\delta)q^{-1}+\lambda q^{-2}} \\ H_1(q^{-1}) &= \frac{(1-\delta)(1-q^{-1})q^{-1}}{b^2(1-(\lambda+\delta)q^{-1}+\lambda q^{-2})} \\ H_2(q^{-1}) &= -\frac{1-\lambda q^{-1}}{1-(\lambda+\delta)q^{-1}+\lambda q^{-2}}. \end{aligned} \tag{15}$$

It is easy to check that for any λ and δ from the interval (0,1), the poles of all transfer functions in (15) lie inside the unit circle in the complex plane. Hence, under the constraint mentioned above, the approximating linear filter associated with (8) is stable.

One of the main advantages of the ALF approach is that it gives interesting insights into tracking capabilities (and tracking limitations) of the analyzed algorithms. In addition, it allows one to compare different solutions to adaptive notch filtering.

Remark: The term "uniformly small sequences," used in formulation of Proposition 1, was borrowed from Tichavský and Händel and needs further clarification. Although the conditions under which one can neglect higher order moments in tracking analysis of adaptive filters were not explicitly formulated in [8], it is clear that sequences that are white, Gaussian and mutually uncorrelated satisfy this requirement. We have a rich simulation evidence which suggests that ALF analysis remains valid under considerably less stringent conditions imposed on the driving and measurement noise sequences.

C. Tracking Characteristics

Following many earlier tracking studies, we will assume that the frequency $\omega(t)$ evolves according to the random walk model, i.e., that the frequency increments $w(t)$ form a zero-mean white-noise sequence with variance σ_w^2, independent of $v(t)$. Then, using standard results from the linear filtering theory, one arrives at

$$\begin{aligned} \mathrm{E}[(\Delta\hat{\phi}_R(t))^2] &= I[F(z)]\mathrm{E}\left[e_R^2(t)\right] \\ \mathrm{E}[(\Delta\hat{\phi}_I(t))^2] &= I[G_1(z)]\mathrm{E}\left[e_I^2(t)\right] + I[G_2(z)]\mathrm{E}[w^2(t)] \\ \mathrm{E}[(\Delta\hat{\omega}(t))^2] &= I[H_1(z)]\mathrm{E}\left[e_I^2(t)\right] + I[H_2(z)]\mathrm{E}[w^2(t)] \end{aligned} \tag{16}$$

where

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

is an integral evaluated along the unit circle in the z-plane, and $X(z)$ denotes any stable proper rational transfer function.

By means of residue calculus (see, e.g., [17]), one obtains

$$\begin{aligned} I[F(z)] &= \frac{1-\lambda}{1+\lambda} \cong \frac{\mu}{2} \\ I[G_1(z)] &= \frac{1+\delta-\lambda-3\lambda\delta+2\lambda^2}{(1-\lambda)(1+2\lambda+\delta)} \cong \frac{\gamma}{2\mu}+\frac{\mu}{2} \\ I[G_2(z)] &= \frac{b^4\lambda^2(1+\lambda)}{(1-\lambda)(1-\delta)(1+2\lambda+\delta)} \cong \frac{b^4}{2\mu\gamma} \\ I[H_1(z)] &= \frac{2(1-\delta)^2}{b^4(1-\lambda)(1+2\lambda+\delta)} \cong \frac{\gamma^2}{2b^4\mu} \\ I[H_2(z)] &= \frac{(1-\lambda)^2(1+\lambda)+2\lambda(1-\delta)}{(1-\lambda)(1-\delta)(1+2\lambda+\delta)} \cong \frac{\mu}{2\gamma}+\frac{1}{2\mu} \end{aligned} \tag{17}$$

where $\gamma = 1-\delta = b^2\eta$ and all approximations hold for sufficiently small values of μ and γ.

Finally, combining (16) with (10), (11), and (17) one arrives at the following expressions for the steady-state mean-squared signal estimation and frequency estimation errors:

$$\mathrm{E}[|\hat{\beta}(t)-\beta(t)|^2] \cong \left[\frac{\gamma}{4\mu}+\frac{\mu}{2}\right]\sigma_v^2 + \frac{b^2}{2\mu\gamma}\sigma_w^2 \tag{18}$$

$$\mathrm{E}[(\hat{\omega}(t)-\omega(t))^2] \cong \frac{\gamma^2}{4b^2\mu}\sigma_v^2 + \left[\frac{\mu}{2\gamma}+\frac{1}{2\mu}\right]\sigma_w^2. \tag{19}$$

Even though the above variance (i.e., second-order) expressions were derived based on the first-order ALF approximations, and therefore without further statistical analysis they are not well justified from a theoretical viewpoint, they remain in perfect agreement with experimental results for a wide range of SNR rates (see Section IV. Observe that in each of the derived formulas there are terms proportional to the adaptation gains $\mu = 1-\lambda$ and $\gamma = 1-\delta$, and terms inversely proportional to μ and γ. This stays in agreement with the well-known fact in adaptive filtering: the adaptation gains should be chosen so as to compromise between the tracking speed of an adaptive filter (which increases with growing μ and γ) and its noise rejection capability (which decreases with growing μ and γ) [10].

Denote by μ_β and γ_β the values of μ and γ that minimize the mean-squared signal estimation error, and by μ_ω and γ_ω the analogous quantities that minimize the frequency estimation error. Straightforward calculations yield

$$\begin{gathered} \mu_\beta = \sqrt[4]{2\xi}, \quad \gamma_\beta = \sqrt{2\xi} \\ \mathrm{E}[|\hat{\beta}(t)-\beta(t)|^2|\mu_\beta,\gamma_\beta] \cong \sqrt[4]{2\xi}\sigma_v^2 \end{gathered} \tag{20}$$

and

$$\begin{gathered} \mu_\omega = \sqrt[4]{8\xi}, \quad \gamma_\omega = \sqrt{2\xi} \\ \mathrm{E}[(\hat{\omega}(t)-\omega(t))^2|\mu_\omega,\gamma_\omega] \cong \sqrt[4]{2\xi^{-1}}\sigma_w^2 \end{gathered} \tag{21}$$

where

$$\xi = \frac{b^2\sigma_w^2}{\sigma_v^2}. \tag{22}$$

Note that the optimal values of design parameters and the best achievable performance are functions of a scalar coefficient ξ—a product of the SNR b^2/σ_v^2 and the variance of frequency changes σ_w^2. The coefficient ξ can be regarded a measure of signal nonstationarity and plays an important role in analysis of tracking capabilities of the algorithm (8).

Beyond any doubt, performance of an adaptive filter can be regarded entirely satisfactory if it does not significantly differ from performance achievable under exact knowledge of signal parameters. Using (20), one can easily establish conditions under which the algorithm (8) fulfills this requirement. Demanding that the mean-squared signal estimation error $\mathrm{E}[|\hat{\beta}(t)-\beta(t)|^2]$, which for small adaptation gains is approximately equal to the excess prediction error $\mathrm{E}[\epsilon^2(t)] - \sigma_v^2$, be much smaller than the minimum attainable mean-squared prediction error, equal to σ_v^2, one arrives at the following condition:

$$\sqrt[4]{2\xi} \ll 1 \tag{23}$$

which guarantees that the performance "overhead" due to adaptation will be small. Assuming that the term "much smaller" can be interpreted as "at least ten times smaller," one can rewrite (23) in a more explicit form as $\xi < 10^{-5}$. Quite obviously, the analyzed algorithm will not cease to work if condition (23) is not fulfilled, but its performance under such high degree of signal nonstationarity may be far from satisfactory.

Our second remark will concern the frequency tracking bound (21). According to Tichavský [18], under Gaussian assumptions, the limiting value of the minimum achievable frequency tracking error (called posterior Cramér–Rao bound in

[18]) for a single noisy cisoid with frequency varying according to the random walk model is given in the form

$$\lim_{t\mapsto\infty} \inf \mathrm{E}[(\hat{\omega}(t)-\omega(t))^2] = f(z)\sigma_w^2 \tag{24}$$

where $z = \xi + \sqrt{\xi^2+8\xi}$ and $f(z) = \sqrt{1+4z^{-1}}$. Note that when $\sqrt{\xi} \ll 1$, which is a less stringent condition than (23), it holds that $z \cong \sqrt{8\xi}$ and $f(z) \cong \sqrt{4z^{-1}} \cong \sqrt[4]{2\xi^{-1}}$. A comparison of the resulting posterior Cramér–Rao bound $\lim_{t\mapsto\infty} \inf \mathrm{E}[(\hat{\omega}(t)-\omega(t))^2] = \sqrt[4]{2\xi^{-1}}\sigma_w^2$ with (21) implies that the analyzed algorithm (similarly as the other frequency trackers, examined in [8] and [9]) is, in the range of applicability of the ALF approximation, a *statistically efficient* procedure for estimation/tracking of a slowly varying frequency $\omega(t)$.

Our last comment will be devoted to the problem of choice of design variables μ and γ (or equivalently λ and δ). First of all, recall that γ, equal to $b^2\eta$, is a function of a signal power $b^2 = |\beta(t)|^2$. Therefore, unless $|\beta(t)|$ is constant (which we have been assuming so far) and known *a priori*, the user does not have full control over the adaptation gain γ. This obvious drawback can be eliminated by replacing the correction term $g(t)$ in (8) with the normalized correction term

$$\bar{g}(t) = \frac{g(t)}{\hat{b}^2(t)} = \mathrm{Im}\left[\frac{\varepsilon^*(t)e^{j\hat{\omega}(t)}\hat{\beta}(t-1)}{\hat{b}^2(t)}\right] \tag{25}$$

where $\hat{b}^2(t)$ denotes a local estimate of $b^2 = |\beta(t)|^2$, for example

$$\hat{b}^2(t) = \lambda_o \hat{b}^2(t-1) + (1-\lambda_o)|\hat{\beta}(t)|^2$$

where $0 \le \lambda_o < 1$ is the local averaging coefficient (e.g., $\lambda_o = 0.9$). Careful analysis shows that such modification does not change equations of the approximating linear filter associated with (8), provided that δ is redefined as $\delta = 1-\eta$. In this case, γ is equal to η, i.e., it is an entirely user-dependent quantity. The modifications described above can be easily extended to the multiple-frequency case.

Even though our optimization study was not based on realistic assumptions (the random walk model of signal frequency variation can be criticized as rather naive), its results, summarized in (20) and (21), have some practical relevance as they suggest useful tuning rules. Observe that the optimal settings of γ (γ_β and γ_ω) are proportional to the squares of the corresponding settings of μ (μ_β and μ_ω, respectively). Therefore, to make tuning easier it may be worthwhile to set $\gamma = \mu^2$ (when signal tracking is the main objective) or $\gamma = \mu^2/2$ (if frequency tracking is the main goal). The problem is then reduced to selection of a single design parameter μ. Optimization of μ can be performed either sequentially (e.g., by setting $\mu(t) = \mu[\hat{\xi}(t)]$, where $\hat{\xi}(t)$ is a continuously updated local estimate of the rate of signal nonstationarity—see, e.g., [19] and [20]), or using the parallel estimation approach (which combines in a rational way the results yielded by a bank of adaptive filters with different settings—see, e.g., [21] and [22]). Because of the lack of space such fully adaptive procedures will not be discussed here.

D. Comparison With Known Adaptive Notch Filtering Algorithms

We will start from presenting ALF results for the multiple frequency tracking algorithm based on (9). While the frequency tracking results summarized below are known (they are suitably modified versions of the results derived in [8]), the signal tracking analysis is new and seems to be presented for the first time.

Proposition 2: Assume that all conditions of ALF analysis are fulfilled. Then the MFT algorithm applied to the signal (12) can be approximately described by (14) with

$$\begin{aligned}
F(q^{-1}) &= \frac{1-\lambda}{1-\lambda q^{-1}}\\
G_1(q^{-1}) &= \frac{(1-\lambda)(1-\rho q^{-1})}{1-(2\lambda+\rho-\rho\lambda)q^{-1}+\lambda q^{-2}}\\
G_2(q^{-1}) &= -\frac{b^2\lambda}{1-(2\lambda+\rho-\rho\lambda)q^{-1}+\lambda q^{-2}}\\
H_1(q^{-1}) &= \frac{(1-\lambda)(1-\rho)(1-q^{-1})q^{-1}}{b^2(1-(2\lambda+\rho-\rho\lambda)q^{-1}+\lambda q^{-2})}\\
H_2(q^{-1}) &= -\frac{1-\lambda q^{-1}}{1-(2\lambda+\rho-\rho\lambda)q^{-1}+\lambda q^{-2}}
\end{aligned} \tag{26}$$

where $\rho = 1-\eta$.

Derivation: See Appendix II.

One can check that the substitution $\delta = \lambda+\rho-\rho\lambda$ (or equivalently $\gamma = \eta\mu$) converts transfer functions (15), derived in the previous subsection for the algorithm (8), into the transfer functions (26), characterizing local behavior of the MFT algorithm. It is therefore clear that for a single noisy cisoid, both algorithms have essentially the same signal tracking and frequency tracking properties, i.e., they achieve the same signal tracking and frequency tracking performance bounds, they show approximately the same sensitivity to the choice of design parameters, etc. Interestingly, the same conclusion was reached in [9], where the multiple frequency tracker was compared with yet another three adaptive notch filtering algorithms, mentioned at the beginning of Section II. The equivalence, pointed out previously, means also that the alternative tuning rule suggested in [8], which aims at maximization of the noise rejection rate for a fixed settling time of MFT, can be easily adapted to the GANF algorithm (7).

III. Analysis of System Identification Algorithm

Similarly as in the signal identification case, we will transform the original system identification algorithm (3) using the following mappings:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}(t) = \hat{\mathbf{F}}_n(t)\hat{\boldsymbol{\alpha}}(t), \quad \mathbf{l}(t) &= \hat{\mathbf{F}}_n^*(t)\mathbf{k}(t)\\
\mathbf{P}(t) &= \hat{\mathbf{F}}_n^*(t)\mathbf{Q}(t)\hat{\mathbf{F}}_n(t)
\end{aligned} \tag{27}$$

where $\hat{\mathbf{F}}_n(t) = \hat{\mathbf{F}}(t)\otimes\mathbf{I}_n$ and $\otimes$ denotes the Kronecker product of the corresponding matrices. Let $\hat{\mathbf{A}}_n(t) = \hat{\mathbf{A}}(t)\otimes\mathbf{I}_n$ and

$\varphi_k(t) = [\underbrace{\varphi^{\mathrm{T}}(t), \ldots, \varphi^{\mathrm{T}}(t)}_{k}]^{\mathrm{T}}$. Using (27), one can express the algorithm (3) in the following equivalent form:

$$\begin{aligned}
\varepsilon(t) &= y(t) - \varphi_k^{\mathrm{T}}(t)\hat{\mathbf{A}}_n(t)\hat{\boldsymbol{\beta}}(t-1)\\
\mathbf{P}(t) &= \frac{1}{\lambda}\hat{\mathbf{A}}_n^*(t)\\
&\quad \Bigg[\mathbf{P}(t-1)\\
&\quad -\frac{\mathbf{P}(t-1)\hat{\mathbf{A}}_n(t)\varphi_k(t)\varphi_k^{\mathrm{H}}(t)\hat{\mathbf{A}}_n^*(t)\mathbf{P}(t-1)}{\lambda + \varphi_k^{\mathrm{H}}(t)\hat{\mathbf{A}}_n^*(t)\mathbf{P}(t-1)\hat{\mathbf{A}}_n(t)\varphi_k(t)}\Bigg]\hat{\mathbf{A}}_n(t)\\
\mathbf{l}(t) &= \mathbf{P}(t)\varphi_k(t)\\
\hat{\boldsymbol{\beta}}(t) &= \hat{\mathbf{A}}_n(t)\hat{\boldsymbol{\beta}}(t-1) + \mathbf{l}^*(t)\varepsilon(t)\\
g_i(t) &= \mathrm{Im}\{\varepsilon^*(t)e^{j\hat{\omega}_i(t)}\varphi^{\mathrm{T}}(t)\hat{\boldsymbol{\beta}}_i(t-1)\}\\
\hat{\omega}_i(t+1) &= \hat{\omega}_i(t) - \eta g_i(t)\\
i &= 1, \ldots, k\\
\hat{\boldsymbol{\theta}}(t) &= \sum_{i=1}^{k}\hat{\boldsymbol{\beta}}_i(t).
\end{aligned} \tag{28}$$

Similarly, as in Section II, we will consider the single-frequency case ($k = 1$). Note that in this case, $\hat{\mathbf{A}}_n(t) = e^{j\hat{\omega}_i(t)}\mathbf{I}_n$, $\varphi_k(t) = \varphi(t)$ and the matrix $\mathbf{P}(t)$ can be written down in an explicit form

$$\mathbf{P}(t) = \left[\sum_{s=1}^{t}\lambda^{t-s}\varphi(s)\varphi^{\mathrm{H}}(s)\right]^{-1}. \tag{29}$$

If the sequence of regression vectors is wide-sense stationary and persistently exciting, and λ is close to 1, one can replace the matrix inverted in (29) with its expectation [23]. This results in the following steady-state approximation:

$$\mathbf{P}(t) \cong \left[\sum_{s=1}^{t}\lambda^{t-s}\boldsymbol{\Phi}^*\right]^{-1} \underset{t\mapsto\infty}{\longmapsto} (1-\lambda)(\boldsymbol{\Phi}^*)^{-1} \tag{30}$$

where $\boldsymbol{\Phi} = \mathrm{E}[\varphi^*(t)\varphi^{\mathrm{T}}(t)] > 0$.

Using this approximation, the GANF algorithm (28) can be, for a system with a single-frequency mode, rewritten in a simplified form, as follows:

$$\begin{aligned}
\varepsilon(t) &= y(t) - e^{j\hat{\omega}(t)}\varphi^{\mathrm{T}}(t)\hat{\boldsymbol{\beta}}(t-1)\\
\hat{\boldsymbol{\beta}}(t) &= e^{j\hat{\omega}(t)}\hat{\boldsymbol{\beta}}(t-1) + \mu\boldsymbol{\Phi}^{-1}\varphi^*(t)\varepsilon(t)\\
g(t) &= \mathrm{Im}\{\varepsilon^*(t)e^{j\hat{\omega}(t)}\varphi^{\mathrm{T}}(t)\hat{\boldsymbol{\beta}}(t-1)\}\\
\hat{\omega}(t+1) &= \hat{\omega}(t) - \eta g(t)\\
\hat{\boldsymbol{\theta}}(t) &= \hat{\boldsymbol{\beta}}(t)
\end{aligned} \tag{31}$$

which will be a subject of our further analysis.

A. Tracking Characteristics

Let

$$\begin{aligned}
\Delta\hat{\boldsymbol{\beta}}(t) &= \hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\\
\Delta\hat{\phi}(t) &= \boldsymbol{\beta}^{\mathrm{H}}(t)\boldsymbol{\Phi}\Delta\hat{\boldsymbol{\beta}}(t) = \Delta\hat{\phi}_R(t) + j\Delta\hat{\phi}_I(t)\\
e(t) &= \boldsymbol{\beta}^{\mathrm{H}}(t)\varphi^*(t)v(t) = e_R(t) + je_I(t).
\end{aligned}$$

Note that $e(t)$ is a complex-valued white noise with variance $\sigma_e^2 = \mathrm{E}[|e(t)|^2] = \boldsymbol{\beta}^{\mathrm{H}}(0)\boldsymbol{\Phi}\boldsymbol{\beta}(0)\sigma_v^2$. One can show that $\mathrm{E}[e_R^2(t)] = \mathrm{E}[e_I^2(t)] = \sigma_e^2/2$ and $\mathrm{E}[e_R(t)e_I(t)] = 0$.

Using the notation introduced above, one can prove

Proposition 3: Assume that all conditions of ALF analysis are fulfilled and that the sequence of regression vectors $\varphi(t)$, independent of $v(t)$ and $w(t)$, is wide-sense stationary and persistently exciting. Then, the GANF algorithm (31) applied to the system governed by

$$y(t) = \varphi^{\mathrm{T}}(t)\boldsymbol{\beta}(t) + v(t), \quad \boldsymbol{\beta}(t) = e^{j\omega(t)}\boldsymbol{\beta}(t-1) \tag{32}$$

can be approximately described by (13) with $b^2 = \boldsymbol{\beta}_o^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_o$, where $\boldsymbol{\beta}_o = \boldsymbol{\beta}(0)$.

Derivation: See Appendix III.

Since ALF associated with the GANF is the same as that derived in Section II for the signal case, many of the conclusions drawn there extend to the system case. In particular, analysis of the frequency tracking error for random walk frequency variations (19), and the corresponding optimal settings (21) remain valid in the system identification case. Similarly, as in the signal case, the optimal values of design parameters are functions of a scalar coefficient

$$\xi = \frac{\boldsymbol{\beta}_o^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_o\sigma_w^2}{\sigma_v^2} \tag{33}$$

which is a product of the SNR $\boldsymbol{\beta}_o^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_o/\sigma_v^2$ and the variance of frequency changes σ_w^2, and which can be regarded a measure of system nonstationarity. In addition, similarly as in the signal case, one can make the tracking capability of the analyzed algorithm (nearly) independent of the system-related statistics by replacing the correction term $g(t)$ in (31) with the normalized correction term

$$\bar{g}(t) = \frac{g(t)}{\hat{b}^2(t)} = \mathrm{Im}\left[\frac{\varepsilon^*(t)e^{j\hat{\omega}(t)}\varphi^{\mathrm{T}}(t)\hat{\boldsymbol{\beta}}(t-1)}{\hat{b}^2(t)}\right] \tag{34}$$

where $\hat{b}^2(t)$ denotes a local estimate of $\boldsymbol{\beta}_o^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_o = \mathrm{E}[|\varphi^{\mathrm{T}}(t)\boldsymbol{\beta}(t)|^2]$, for example

$$\hat{b}^2(t) = \lambda_o\hat{b}^2(t-1) + (1-\lambda_o)|\varphi^{\mathrm{T}}(t)\hat{\beta}(t)|^2.$$

Unfortunately, the parameter tracking results cannot be easily generalized to the system case. The estimate of the noiseless system output $s(t) = \varphi^{\mathrm{T}}(t)\boldsymbol{\beta}(t)$ is given by $\hat{s}(t) = \varphi^{\mathrm{T}}(t)\hat{\boldsymbol{\beta}}(t)$, yielding the system tracking error $\varepsilon_s(t) = \hat{s}(t) - s(t) = \varphi^{\mathrm{T}}(t)\Delta\hat{\boldsymbol{\beta}}(t)$. If the adaptation gains μ and γ are sufficiently small (the slow adaptation case) and the sequence of regression vectors $\varphi(t)$ is wide-sense stationary, the variance of the system tracking error (approximately equal to the variance of the one-step-ahead excess prediction error) is given by $\mathrm{E}[|\varepsilon_s(t)|^2] = \mathrm{E}[\Delta\hat{\boldsymbol{\beta}}^{\mathrm{H}}(t)\varphi^*(t)\varphi^{\mathrm{T}}(t)\Delta\hat{\boldsymbol{\beta}}(t)] \cong \mathrm{E}[\Delta\hat{\boldsymbol{\beta}}^{\mathrm{H}}(t)\boldsymbol{\Phi}\Delta\hat{\boldsymbol{\beta}}(t)]$. Hence, the quantity $\mathrm{E}[\Delta\hat{\boldsymbol{\beta}}^{\mathrm{H}}(t)\boldsymbol{\Phi}\Delta\hat{\boldsymbol{\beta}}(t)]$ is a natural extension, to the system case, of the signal tracking statistic $\mathrm{E}[|\Delta\hat{\beta}(t)|^2]$. The ALF-based analysis of (31) allows one to evaluate $\mathrm{E}[|\Delta\hat{\phi}(t)|^2] = \mathrm{E}[|\boldsymbol{\beta}^{\mathrm{H}}(t)\boldsymbol{\Phi}\Delta\hat{\boldsymbol{\beta}}(t)|^2]$. Unfortunately, unlike the signal case, the above quantity cannot be explicitly related to the system error

$\mathrm{E}[\Delta\hat{\boldsymbol{\beta}}^{\mathrm{H}}(t)\boldsymbol{\Phi}\Delta\hat{\boldsymbol{\beta}}(t)]$. The only exception is a single-parameter system ($\dim\boldsymbol{\beta} = \dim\boldsymbol{\varphi} = n = 1$) for which it holds that $\mathrm{E}[|\Delta\hat{\phi}(t)|^2] = |\beta_o|^2\sigma_\varphi^4\mathrm{E}[|\Delta\hat{\beta}(t)|^2] = b^2\mathrm{E}[|\varepsilon_s(t)|^2]$, where $b^2 = |\beta_o|^2\sigma_\varphi^2$. Therefore, in this simple case, the signal-oriented tracking results (18) and (20) remain valid for the system identification algorithm (31).

B. Statistical Efficiency

Consider a system governed by (32) with the frequency $\omega(t)$ evolving according to the random walk model. Suppose that the initial value $\boldsymbol{\beta}(0) = \boldsymbol{\beta}_o$ is known, that the prior distribution of $\omega(1)$ is noninformative (i.e., $\pi(\omega(1)) = 1/(2\pi)$ for $\omega(1) \in (-\pi, \pi]$) and that the white-noise sequences $\{v(t)\}$ and $\{w(t)\}$ are mutually independent and Gaussian. In addition, suppose that the driving sequence $\{\boldsymbol{\varphi}(t)\}$ is wide-sense stationary and independent of $\{v(t)\}$ and $\{w(t)\}$.

Denote by $Y(t) = \{y(1), \dots, y(t)\}$ the history of system output available at instant t, and by $U(t) = \{\boldsymbol{\varphi}(1), \dots, \boldsymbol{\varphi}(t)\}$—the analogous input history. Let $\hat{\boldsymbol{\omega}} = g[Y(t), U(t)] = [\hat{\omega}(1), \dots, \hat{\omega}(t)]^{\mathrm{T}}$ be any estimator (possibly biased) of the vector of instantaneous frequencies $\boldsymbol{\omega} = [\omega(1), \dots, \omega(t)]^{\mathrm{T}}$. Then, under some regularity conditions which can be easily verified in the Gaussian case, it holds that [24], [18]

$$\mathrm{E}[(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega})(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega})^{\mathrm{T}}] \geq \mathbf{J}_t^{-1} \tag{35}$$

where

$$\mathbf{J}_t^{-1} = -\mathrm{E}\left[\nabla^2_{\boldsymbol{\omega}\boldsymbol{\omega}} \log p(Y(t), U(t), \boldsymbol{\omega})\right]. \tag{36}$$

Since the density $p(Y(t), U(t), \boldsymbol{\omega})$ is a product of the likelihood function $p(Y(t), U(t)|\boldsymbol{\omega})$ and the prior density $p(\boldsymbol{\omega})$, the $t \times t$ matrix $\mathbf{J}_t$ is the sum of the standard Fisher information matrix (representing the information obtained from the data) and an additional *a priori* information matrix (representing the prior knowledge of the estimated parameters). Note that (35) implies that

$$\mathrm{E}[(\hat{\boldsymbol{\omega}}_t - \boldsymbol{\omega}_t)^2] = \mathrm{E}[(\hat{\omega}(t) - \omega(t))^2] \geq \left[\mathbf{J}_t^{-1}\right]_{tt}$$

where $\hat{\boldsymbol{\omega}}_i = \hat{\omega}(i)$ denotes the ith component of $\hat{\boldsymbol{\omega}}$. Hence, the limiting steady-state value of the mean-squared frequency estimation error (posterior Cramér–Rao bound) can be obtained by examining $\lim_{t\mapsto\infty}[\mathbf{J}_t^{-1}]_{tt}$, which leads to the following result.

Proposition 4: The limiting value of the posterior Cramér–Rao bound for the estimation of $\omega(t)$ in the quasi-periodically varying system described above is given by (24) and (33).

Proof: See Appendix IV. As argued in the previous subsection, the best achievable performance of the GANF algorithm (31), applied to a system with random-walk-type frequency changes, is characterized by (22) with ξ given by (33). Hence, using the same arguments as in Section 2A, one may conclude that, under the assumptions of the ALF approximation, the optimally tuned GANF is a statistically efficient estimator of a slowly drifting system frequency.

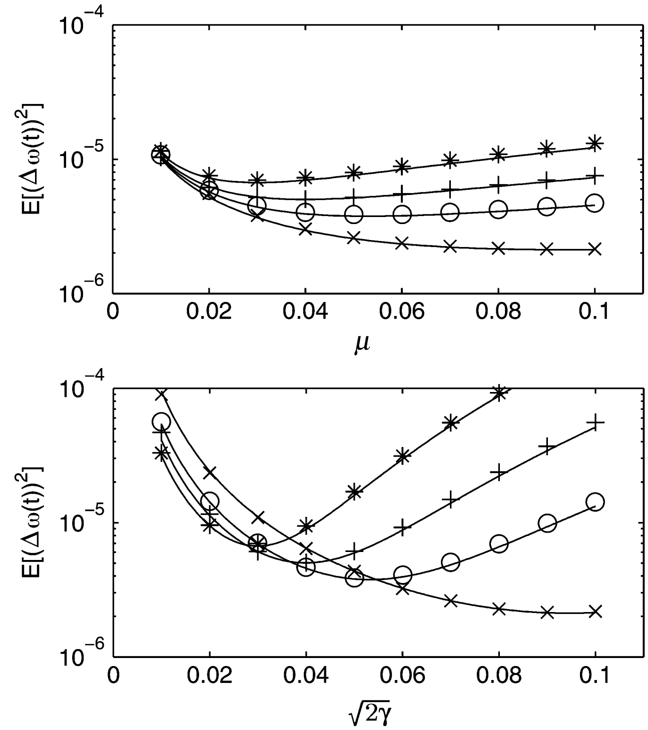

Fig. 1. Variance of the frequency estimation error $\Delta\hat{\omega}(t)$ for an FIR system with a single-frequency mode subject to a random walk drift. The theoretical results (solid lines) are compared with simulation results obtained for different values of μ given $\gamma = \gamma_\omega$ (upper plot) and different values of γ given $\mu = \mu_\omega$ (lower plot); the corresponding SNRs were 0 dB $(*)$, 5 dB $(+)$, 10 dB $(\circ)$, and 20 dB $(\times)$.

IV. Simulation Results

Several simulation experiments were performed to verify results of theoretical analysis presented in Section III (since the signal case was pretty well documented in [8] and [9], we focused exclusively on the system case). The results summarized below were obtained for a time-varying two-tap finite-impulse response (FIR) system governed by

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

where $u(t)$ denotes a white 4-QAM input sequence ($u(t) = \pm 1 \pm j, \sigma_u^2 = 2$), and $v(t)$ denotes a complex Gaussian measurement noise. The impulse response coefficients of the system were modeled as nonstationary cisoids $\theta_i(t) = a_i e^{j\psi(t)}, i = 1, 2, \psi(t) = \sum_{s=1}^{t}\omega(s)$, with time-invariant complex "amplitudes" $\boldsymbol{\alpha} = [a_1, a_2]^{\mathrm{T}} = [2 - j, 1 + 2j]^{\mathrm{T}}$. Note that in this case, $\boldsymbol{\beta}_o = \boldsymbol{\alpha}, \boldsymbol{\varphi}(t) = [u(t), u(t-1)]^{\mathrm{T}}$ and $\boldsymbol{\Phi} = \mathbf{I}_2\sigma_u^2$. The evolution of the frequency $\omega(t)$ was modeled as a random walk process with the variance of frequency increments set to $\sigma_w^2 = 10^{-7}$ and with the starting value set to $\omega(0) = \pi/2$. Four noise levels were considered ($\sigma_v^2 = 20, 2\sqrt{10}, 2$, and 0.2) to check tracking performance of the GANF algorithm under different SNR conditions (0, 5 10 and 20 dB, respectively). It should be emphasized that although the system described above was inspired by channel equalization applications, it is definitely too simple to have any practical relevance.

According to (21), to optimize frequency tracking, one should set μ to $\mu_\omega = \sqrt[4]{8\xi}$ and set γ to $\gamma_\omega = \sqrt{2\xi}$ (i.e., set η to $\eta_\omega = \gamma_\omega/\boldsymbol{\beta}_o^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_o$). Fig. 1 shows a comparison of theoretical evaluations, based on (21), with the results of computer simulations. For each SNR, the analysis was carried around the optimal

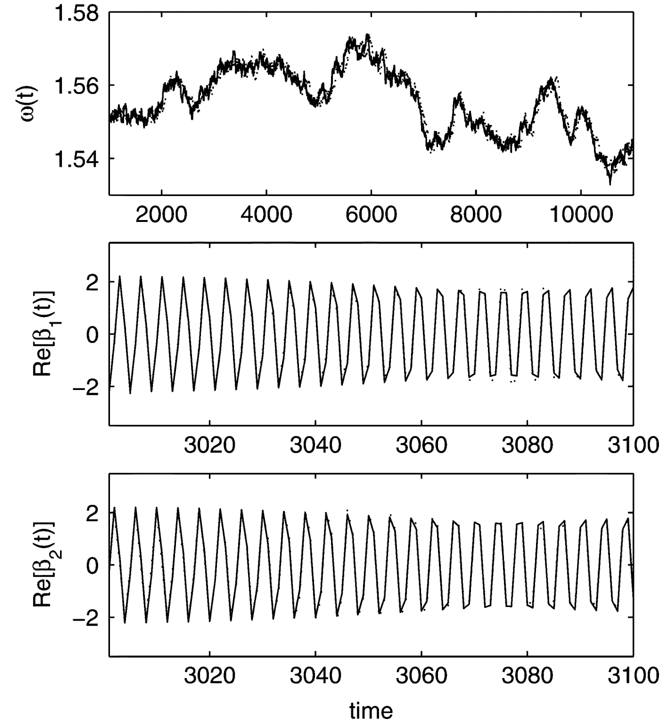

Fig. 2. Typical results of system frequency tracking (upper plot) and system parameter tracking (two lower plots). Solid lines depict true values, and dotted lines show the evolution of the corresponding estimates (SNR = 5 dB).

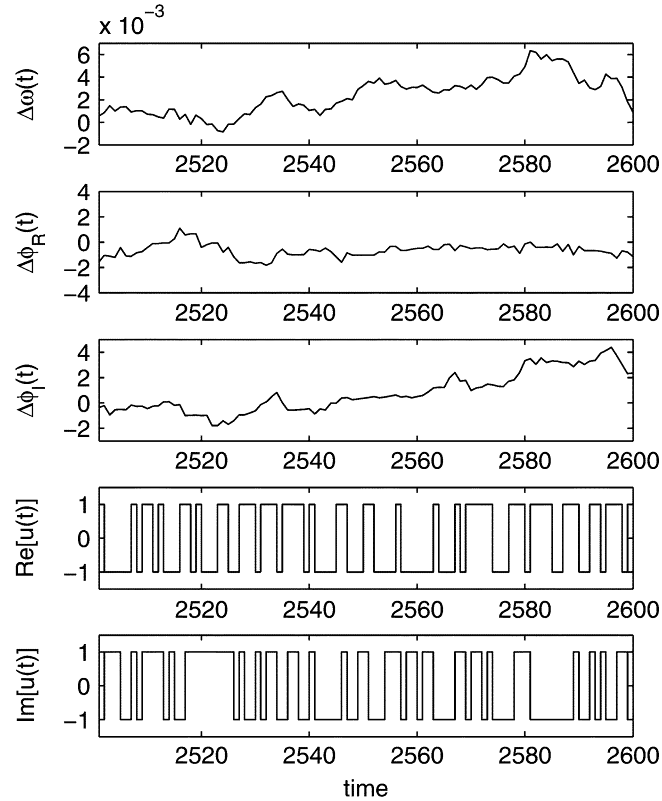

Fig. 3. Evolution of $\Delta\hat{\omega}(t)$, $\Delta\hat{\phi}(t)$ and $u(t)$ for a typical run of the optimally tuned GANF algorithm (SNR = 5 dB).

point $(\mu_\omega, \gamma_\omega)$. In the first experiment, γ was set to its optimal value γ_ω and μ was changed around μ_ω. In the second experiment, μ was set to μ_ω, and γ was changed around γ_ω. Both plots shown in Fig. 1 were obtained by double averaging. First, the mean-squared frequency estimation errors were computed for different pairs (μ, γ) and for a given frequency trajectory from 10 000 iterations of the GANF algorithm (after the algorithm has reached its steady state). The obtained results were next averaged over 50 realizations of $\{w(t)\}$, i.e., over 50 different frequency trajectories. Note very good agreement between the theoretical curves and the results of computer simulations.

Fig. 2 shows typical results of system parameter and frequency tracking.

Finally, Fig. 3 shows the evolution of $\Delta\hat{\omega}(t)$, $\Delta\hat{\phi}(t)$ and $u(t)$ for a typical run of the optimally tuned GANF algorithm ($\mu = \mu_\omega, \gamma = \gamma_\omega$). Note that, exactly as we assumed in Appendix III, the quantities $\Delta\hat{\phi}(t)$ (a linear combination of the elements of the modified error $\Delta\tilde{\boldsymbol{\beta}}(t)$, defined in Appendix III) and $\Delta\hat{\omega}(t)$ vary slowly compared with $u(t)$, i.e., compared with $\boldsymbol{\varphi}(t)$. This confirms validity of the averaging approach which was used to derive the approximating linear filter in the system case.

V. Conclusion

We have shown that performance of a generalized adaptive notch filter (GANF), proposed for identification/tracking of quasi-periodically varying systems, can be studied analytically using a direct averaging approach and an approximating linear filtering (ALF) technique. Such analysis provides valuable insights into the tracking capabilities of the algorithm and allows one to formulate useful rules of thumb for choosing its design parameters. We have also shown that, under the conditions of the ALF approximation, the optimally tuned GANF is a statistically efficient estimator of a slowly drifting system frequency.

Appendix I
Derivation of (13)

Since $y(t) = \beta(t) + v(t)$, it holds that

$$\varepsilon(t) = \beta(t) - e^{j\hat{\omega}(t)}\hat{\beta}(t-1) + v(t)$$
$$\Delta\hat{\beta}(t) = e^{j\hat{\omega}(t)}\hat{\beta}(t-1) - \beta(t) + \mu\varepsilon(t).$$

Combining the last two equations, one arrives at

$$\Delta\hat{\beta}(t) = \lambda[e^{j\hat{\omega}(t)}\hat{\beta}(t-1) - \beta(t)] + \mu v(t).$$

Note that

$$e^{j\hat{\omega}(t)}\hat{\beta}(t-1) = e^{j\omega(t)}e^{j\Delta\hat{\omega}(t)}[\Delta\hat{\beta}(t-1) + \beta(t-1)].$$

For small frequency errors, it holds that $e^{j\Delta\hat{\omega}(t)} \cong 1 + j\Delta\hat{\omega}(t)$. Using this approximation and neglecting all terms of order higher than one in $\Delta\hat{\omega}(t)$ and $\Delta\hat{\beta}(t-1)$, one obtains

$$e^{j\hat{\omega}(t)}\hat{\beta}(t-1) \cong \beta(t) + e^{j\omega(t)}\Delta\hat{\beta}(t-1) + j\beta(t)\Delta\hat{\omega}(t).$$

Therefore

$$\Delta\hat{\beta}(t) \cong \lambda e^{j\omega(t)}\Delta\hat{\beta}(t-1) + j\lambda\beta(t)\Delta\hat{\omega}(t) + \mu v(t)$$

and

$$\begin{aligned}\beta^*(t)\Delta\hat{\beta}(t) = {} & \lambda\beta^*(t-1)\Delta\hat{\beta}(t-1) \\ & + j\lambda b^2\Delta\hat{\omega}(t) + \mu\beta^*(t)v(t).\end{aligned}$$

The first two equations of (13) follow directly from the above result.

A similar technique can be used to cope with the frequency update in (8). Note that

$$\begin{aligned}\varepsilon^*(t)e^{j\hat{\omega}(t)}\hat{\beta}(t-1) &= e^{j\Delta\hat{\omega}(t)}\beta^*(t-1)\hat{\beta}(t-1)\\ &\quad -|\hat{\beta}(t-1)|^2+e^{j\hat{\omega}(t)}\hat{\beta}(t-1)v^*(t)\end{aligned}$$

and

$$\begin{aligned}e^{j\Delta\hat{\omega}(t)}\beta^*(t-1)\hat{\beta}(t-1) &\cong \beta^*(t-1)(1+j\Delta\hat{\omega}(t))(\Delta\hat{\beta}(t-1)+\beta(t-1))\\ &\cong \beta^*(t-1)\Delta\hat{\beta}(t-1)\\ &\quad +j|\beta(t-1)|^2\Delta\hat{\omega}(t)+|\beta(t-1)|^2.\end{aligned}$$

Combining the last two equations, and noting that

$$e^{j\hat{\omega}(t)}\hat{\beta}(t-1)v^*(t)\cong\beta(t)v^*(t)$$

(since under our first-order approximation the terms proportional to $v^*(t)\Delta\hat{\omega}(t)$ and $v^*(t)\Delta\hat{\beta}(t-1)$ can be neglected), gives

$$\begin{aligned}\varepsilon^*(t)e^{j\hat{\omega}(t)}\hat{\beta}(t-1) &\cong \beta^*(t-1)\Delta\hat{\beta}(t-1)+jb^2\Delta\hat{\omega}(t)\\ &\quad +|\beta(t-1)|^2-|\hat{\beta}(t-1)|^2+\beta(t)v^*(t)\\ \hat{\omega}(t+1) &\cong \hat{\omega}(t)-\eta b^2\Delta\hat{\omega}(t)-\eta\mathrm{Im}[\beta^*(t-1)\Delta\hat{\beta}(t-1)]\\ &\quad +\eta\mathrm{Im}[\beta^*(t)v(t)].\end{aligned}$$

Finally, since $\omega(t+1)=\omega(t)+w(t+1)$, the last equation is equivalent to

$$\begin{aligned}\Delta\hat{\omega}(t+1) \cong (1-\eta b^2)\Delta\hat{\omega}(t)&-\eta\mathrm{Im}[\beta^*(t-1)\Delta\hat{\beta}(t-1)]\\ &+\eta\mathrm{Im}[\beta^*(t)v(t)]-w(t+1)\end{aligned}$$

which is the third equation of (13).

All approximations hold for sufficiently high signal-to-noise ratio (SNR) and for a sufficiently low rate of frequency changes compared with 1/SNR (see further comments in Section III-A).

APPENDIX II
DERIVATION OF (27)

Since the first two equations of (13) remain valid for the multiple frequency tracker, all one needs to do is find a counterpart of the third equation for the modified gradient rule (9). We will use approximations proposed in [8] (with slight modifications due to notational differences). Note that

$$\begin{aligned}\mathrm{Arg}\left[\frac{\hat{\beta}(t)}{\hat{\beta}(t-1)e^{j\hat{\omega}(t)}}\right] &= \mathrm{Im}\ \log\left[\frac{\hat{\beta}(t)}{\hat{\beta}(t-1)e^{j\hat{\omega}(t)}}\right]\\ &= \mathrm{Im}\ \log\left[\frac{\hat{\beta}(t)}{\beta(t)}\right]-\mathrm{Im}\ \log\left[\frac{\hat{\beta}(t-1)}{\beta(t-1)}\right]\\ &\quad +\mathrm{Im}\ \log\left[\frac{e^{j\omega(t)}}{e^{j\hat{\omega}(t)}}\right]\end{aligned}$$

and

$$\frac{\hat{\beta}(t)}{\beta(t)}=1+\frac{\Delta\hat{\beta}(t)}{\beta(t)}=1+\frac{\beta^*(t)\Delta\hat{\beta}(t)}{b^2}.$$

Hence, in the range of small values of $\Delta\hat{\beta}(t)$, one obtains

$$g(t)\cong\frac{\Delta\hat{\phi}_I(t)}{b^2}-\frac{\Delta\hat{\phi}_I(t-1)}{b^2}-\Delta\hat{\omega}(t)$$

leading to

$$\begin{aligned}\Delta\hat{\omega}(t+1)\cong(1-\eta)\Delta\hat{\omega}(t)&\\ &+\frac{\eta(1-q^{-1})}{b^2}\Delta\hat{\phi}_I(t)-w(t+1).\end{aligned}$$

Transfer functions, given by (26), can be easily obtained by solving ALF equations with respect to $\Delta\hat{\phi}_R(t),\Delta\hat{\phi}_I(t)$ and $\Delta\hat{\omega}_I(t)$.

Remark: There are three differences between our transfer functions $H_1(q^{-1})$ and $H_2(q^{-1})$, given by (26), and the transfer functions derived in [8]

$$\begin{aligned}\Phi_1(q^{-1})&=\frac{(1-\rho)(1-\lambda)(1-q^{-1})}{1-(2\lambda+\rho-\rho\lambda)q^{-1}+\lambda q^{-2}}\\ \Phi_2(q^{-1})&=-\frac{\lambda+\rho-\rho\lambda-\lambda q^{-1}}{1-(2\lambda+\rho-\rho\lambda)q^{-1}+\lambda q^{-2}}.\end{aligned}$$

Note that Tichavský and Händel express frequency estimation errors in terms of normalized noise

$$\frac{v(t)}{\beta(t)}=\frac{\beta^*(t)v(t)}{|\beta(t)|^2}=\frac{e(t)}{b^2}.$$

This explains the lack of the multiplier b^2 in the denominator of $\Phi_1(q^{-1})$, which is the counterpart of our $H_1(q^{-1})$. The remaining two differences are caused by the fact that our frequency estimate $\hat{\omega}(t)$ corresponds to $\tilde{\omega}(t)=\hat{\omega}(t-1)$ in [8]. This single-step time advance has two consequences. First, it explains the presence of an additional term q^{-1} in the numerator of $H_1(q^{-1})$. Second, since $\Delta\hat{\omega}(t)=\Delta\tilde{\omega}(t+1)+w(t+1)$, the numerator of $\Phi_2(q^{-1})$ slightly differs from the numerator of $H_2(q^{-1})$. Note, however, that the numerator of $\Phi_2(q^{-1})$ can be rewritten in the form $1-(1-\lambda)(1-\rho)-\lambda q^{-1}=1-\lambda q^{-1}+0(\mu\gamma)$. It can be easily checked that, in the range of applicability of ALF, the asymptotic expressions for the mean-squared frequency estimation error, which incorporate $\Phi_2(q^{-1})$, are identical with the analogous expressions based on $H_2(q^{-1})$.

APPENDIX III
PROOF OF PROPOSITION 3

Note that

$$\varepsilon(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\left[\boldsymbol{\beta}(t)-e^{j\hat{\omega}(t)}\hat{\boldsymbol{\beta}}(t-1)\right]+v(t).$$

Therefore

$$\begin{aligned}\Delta\hat{\boldsymbol{\beta}}(t)&=e^{j\hat{\omega}(t)}\hat{\boldsymbol{\beta}}(t-1)-\boldsymbol{\beta}(t)+\mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\varepsilon(t)\\ &=(\mathbf{I}_n-\mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t))\left[e^{j\hat{\omega}(t)}\hat{\boldsymbol{\beta}}(t-1)-\boldsymbol{\beta}(t)\right]\\ &\quad +\mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)v(t)\end{aligned}$$

and (cf. Appendix I)

$$\begin{aligned}\Delta\hat{\boldsymbol{\beta}}(t)&\cong(\mathbf{I}_n-\mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t))e^{j\omega(t)}\Delta\hat{\boldsymbol{\beta}}(t-1)\\ &\quad +j(\mathbf{I}_n-\mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t))\boldsymbol{\beta}(t)\Delta\hat{\omega}(t)\\ &\quad +\mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)v(t).\end{aligned}$$

Multiplying both sides of the last equation with $f^*(t) = e^{-j\sum_{s=1}^{t}\omega(s)}$, one obtains

$$\begin{aligned}\Delta\tilde{\boldsymbol{\beta}}(t) \cong &(\mathbf{I}_n - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t))\Delta\tilde{\boldsymbol{\beta}}(t-1)\\ &+ j(\mathbf{I}_n - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t))\boldsymbol{\beta}_o\Delta\hat{\omega}(t)\\ &+ \mu\boldsymbol{\Phi}^{-1}f^*(t)\boldsymbol{\varphi}^*(t)v(t)\end{aligned}$$

where $\Delta\tilde{\boldsymbol{\beta}}(t) = f^*(t)\Delta\hat{\boldsymbol{\beta}}(t)$.

Assuming that the quantities $\Delta\tilde{\boldsymbol{\beta}}(t)$ and $\Delta\hat{\omega}(t)$ change slowly compared with $\boldsymbol{\varphi}(t)$, approximate analysis of the modified estimation error $\Delta\tilde{\boldsymbol{\beta}}(t)$ can be carried out using the direct averaging technique [25]. The averaging technique was proposed and used for analysis of slowly varying adaptive systems. Since the system (32) rapidly varies with time, some additional arguments are needed to justify the slow variation assumption mentioned above. Note that $\Delta\tilde{\boldsymbol{\beta}}(t) = f^*(t)\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_o$. It can be shown that, for small values of μ and γ, $f^*(t)\hat{\boldsymbol{\beta}}(t)$ varies slowly compared with $\hat{f}^*(t)\hat{\boldsymbol{\beta}}(t) = \hat{\boldsymbol{\alpha}}(t)$, which is itself a slowly varying quantity (in the case considered $\hat{\boldsymbol{\alpha}}(t)$ is a long-memory estimator of a time-invariant coefficient vector $\boldsymbol{\alpha}_o = \boldsymbol{\beta}_o$). Note, in particular, that variation of $\Delta\tilde{\boldsymbol{\beta}}(t)$ is much slower than variation of $\boldsymbol{\beta}(t)$ and $\hat{\boldsymbol{\beta}}(t)$. Similar analysis can be carried for $\Delta\hat{\omega}(t)$ (see Section IV for further comments on applicability of the averaging technique to the system analyzed in the paper). Using averaging, one obtains (with a slight abuse of the notation)

$$\Delta\tilde{\boldsymbol{\beta}}(t) \cong \lambda\Delta\tilde{\boldsymbol{\beta}}(t-1) + j\lambda\boldsymbol{\beta}_o\Delta\hat{\omega}(t) + \mu\boldsymbol{\Phi}^{-1}f^*(t)\boldsymbol{\varphi}^*(t)v(t).$$

Since $\boldsymbol{\beta}^{\mathrm{H}}(t)\boldsymbol{\Phi}\Delta\hat{\boldsymbol{\beta}}(t) = \boldsymbol{\beta}_o^{\mathrm{H}}\boldsymbol{\Phi}\Delta\tilde{\boldsymbol{\beta}}(t)$, one arrives at the relationship

$$\begin{aligned}\boldsymbol{\beta}^{\mathrm{H}}(t)\boldsymbol{\Phi}\Delta\hat{\boldsymbol{\beta}}(t) \cong &\lambda\boldsymbol{\beta}^{\mathrm{H}}(t-1)\boldsymbol{\Phi}\Delta\hat{\boldsymbol{\beta}}(t-1)\\ &+ j\lambda\boldsymbol{\beta}_o^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_o\Delta\hat{\omega}(t) + \mu\boldsymbol{\beta}^{\mathrm{H}}(t)\boldsymbol{\varphi}^*(t)v(t)\end{aligned}$$

from which the first two equations of (13) follow immediately.

Turning to the frequency update recursions, observe that

$$\begin{aligned}&\varepsilon^*(t)e^{j\hat{\omega}(t)}\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\beta}}(t-1)\\ &\quad= e^{j\Delta\hat{\omega}(t)}\boldsymbol{\beta}^{\mathrm{H}}(t-1)\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\beta}}(t-1)\\ &\qquad- |\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\beta}}(t-1)|^2 + e^{j\hat{\omega}(t)}\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\beta}}(t-1)v^*(t)\\ &\quad\cong \boldsymbol{\beta}^{\mathrm{H}}(t-1)\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\Delta\hat{\boldsymbol{\beta}}(t-1) + |\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\beta}(t-1)|^2\\ &\qquad- |\boldsymbol{\varphi}^{\mathrm{T}}(t)\hat{\boldsymbol{\beta}}(t-1)|^2 + j|\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\beta}(t-1)|^2\Delta\hat{\omega}(t)\\ &\quad+ \boldsymbol{\beta}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t)v^*(t)\end{aligned}$$

where the (first-order) approximation was carried in the same manner as in the signal identification case (see Appendix I). This leads to

$$\begin{aligned}\Delta\hat{\omega}(t+1) \cong &(1 - \eta\boldsymbol{\beta}_o^{\mathrm{H}}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\beta}_o)\Delta\hat{\omega}(t)\\ &- \eta\mathrm{Im}\left[\boldsymbol{\beta}_o^{\mathrm{H}}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\Delta\tilde{\boldsymbol{\beta}}(t-1)\right]\\ &+ \eta\mathrm{Im}\left[\boldsymbol{\beta}^{\mathrm{H}}(t)\boldsymbol{\varphi}^*(t)v(t)\right] - w(t+1).\end{aligned}$$

Using the averaging technique, one obtains

$$\begin{aligned}\Delta\hat{\omega}(t+1) \cong &(1 - \eta\boldsymbol{\beta}_o^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_o)\Delta\hat{\omega}(t)\\ &- \eta\mathrm{Im}\left[\boldsymbol{\beta}^{\mathrm{H}}(t-1)\boldsymbol{\Phi}\Delta\hat{\boldsymbol{\beta}}(t-1)\right]\\ &+ \eta\mathrm{Im}\left[\boldsymbol{\beta}^{\mathrm{H}}(t)\boldsymbol{\varphi}^*(t)v(t)\right] - w(t+1)\end{aligned}$$

which constitutes the third equation of (13).

APPENDIX IV
DERIVATION OF THE POSTERIOR CRAMÉR–RAO BOUND

First, observe that in the case considered

$$\begin{aligned}&p(Y(t), \boldsymbol{\omega}|U(t)) =\\ &\quad C\exp\left\{-\frac{1}{\sigma_v^2}\sum_{s=1}^{t}|y(s) - \boldsymbol{\varphi}^{\mathrm{T}}(s)\boldsymbol{\beta}_o f(s)|^2 - \frac{1}{2\sigma_w^2}\sum_{s=2}^{t}[\omega(s) - \omega(s-1)]^2\right\}\end{aligned}$$

where C is a constant which does not depend on $\boldsymbol{\omega}$.

Note that

$$\frac{\partial f(s)}{\partial\boldsymbol{\omega}_m} = \frac{\partial e^{j\sum_{i=1}^{s}\omega(i)}}{\partial\omega(m)} = \begin{cases} jf(s) & \text{for } m \le s\\ 0 & \text{for } s < m \le t.\end{cases}$$

Therefore

$$-\frac{\partial\log p(Y(t), \boldsymbol{\omega}|U(t))}{\partial\boldsymbol{\omega}_m} = -\frac{2}{\sigma_v^2}\sum_{s=m}^{t}\mathrm{Re}\left[jy^*(s)\boldsymbol{\varphi}^{\mathrm{T}}(s)\boldsymbol{\beta}_o f(s)\right] + \frac{c_m(\boldsymbol{\omega})}{\sigma_w^2}$$

where

$$c_m(\boldsymbol{\omega}) = \begin{cases}\omega(m) - \omega(m+1) & \text{for } m = 1\\ 2\omega(m) - \omega(m-1) - \omega(m+1) & \text{for } 1 < m < t\\ \omega(m) - \omega(m-1) & \text{for } m = t\end{cases}$$

and

$$-\frac{\partial^2\log p(Y(t), \boldsymbol{\omega}|U(t))}{\partial\boldsymbol{\omega}_m\partial\boldsymbol{\omega}_n} = \frac{2}{\sigma_v^2}\sum_{s=\max(m,n)}^{t}\mathrm{Re}[y^*(s)\boldsymbol{\varphi}^{\mathrm{T}}(s)\boldsymbol{\beta}_o f(s)] + \frac{g_{mn}}{\sigma_w^2}$$

where

$$g_{mn} = \begin{cases}1 & \text{for } m = n = 1 \text{ and } m = n = t\\ 2 & \text{for } 1 < m = n < t\\ -1 & \text{for } m = n \pm 1\\ 0 & \text{elsewhere}.\end{cases}$$

Since $y(s) = \boldsymbol{\varphi}^{\mathrm{T}}(s)\boldsymbol{\beta}_o f(s) + v(s)$, one arrives at

$$\mathrm{E}\left[\sum_{s=\max(m,n)}^{t}\mathrm{Re}[y^*(s)\boldsymbol{\varphi}^{\mathrm{T}}(s)\boldsymbol{\beta}_o f(s)]|U(t)\right] = \boldsymbol{\beta}_o^{\mathrm{H}}\left[\sum_{s=\max(m,n)}^{t}\boldsymbol{\varphi}^*(s)\boldsymbol{\varphi}^{\mathrm{T}}(s)\right]\boldsymbol{\beta}_o$$

which finally, after exploiting the wide-sense stationarity of $\{\boldsymbol{\varphi}(s)\}$, leads to

$$\begin{aligned}[\mathbf{J}_t]_{mn} &= -\mathrm{E}\left[\frac{\partial^2\log p(Y(t), U(t), \boldsymbol{\omega})}{\partial\boldsymbol{\omega}_m\partial\boldsymbol{\omega}_n}\right]\\ &= 2\frac{\boldsymbol{\beta}_o^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_o}{\sigma_v^2}h_{mn} + \frac{g_{mn}}{\sigma_w^2}, \quad m, n = 1, \ldots, t\end{aligned}$$

where $h_{mn} = t + 1 - \max(m, n)$.

Since the expression for $\mathbf{J}_t$ derived above is identical (up to some obvious notational differences) with the analogous formula obtained by Tichavský (see [18, eq. (18)]), the remaining steps of the proof are exactly the same as those in [18].

REFERENCES

[1] M. Niedźwiecki and P. Kaczmarek, "Estimation and tracking of quasi-periodically varying processes," in *Proc. 13th IFAC Symp. System Identification*, Rotterdam, The Netherlands, 2003, pp. 1102–1107.

[2] ——, "Generalized adaptive notch filters," in *Proc. 2004 IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Montreal, QC, Canada, 2004, pp. II-657–II-660.

[3] ——, "Identification of quasi-periodically varying systems using the combined nonparametric/parametric approach," *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 4588–4598, Dec. 2005.

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

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

[6] J. Bakkoury, D. Roviras, M. Ghogho, and F. Castanie, "Adaptive MLSE receiver over rapidly fading channels," *Signal Process.*, vol. 80, pp. 1347–1360, 2000.

[7] T. Söderström and P. Stoica, *System Identification*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[8] P. Tichavský and P. Händel, "Two algorithms for adaptive retrieval of slowly time-varying multiple cisoids in noise," *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1116–1127, May 1995.

[9] P. Tichavský and A. Nehorai, "Comparative study of four adaptive frequency trackers," *IEEE Trans. Signal Process.*, vol. 45, no. 6, pp. 1473–1484, Jun. 1997.

[10] M. Niedźwiecki, *Identification of Time-Varying Processes*. New York: Wiley, 2000.

[11] A. Nehorai, "A minimal parameter adaptive notch filter with constrained poles and zeros," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 33, no. 4, pp. 983–996, Aug. 1985.

[12] B. S. Chen, T. Y. Yang, and B. H. Lin, "Adaptive notch filter by direct frequency estimation," *Signal Process.*, vol. 27, pp. 161–167, 1992.

[13] M. Padmanabhan and K. Martin, "A second-order hyperstable adaptive filter for frequency estimation," *IEEE Trans. Circuits Syst.*, vol. 40, no. 6, pp. 398–403, Jun. 1993.

[14] M. Padmanabhan, "A hyperstable adaptive line enhancer for fast tracking of sinusoidal inputs," *IEEE Trans. Circuits Syst.*, vol. 43, no. 4, pp. 304–315, Apr. 1996.

[15] M. Padmanabhan and P. Tichavský, "Hyperstable adaptive line enhancer for tracking of multiple noisy cisoids," in *Proc. 1996 Eur. Signal Processing Conf. (EUSIPCO-96)*, Trieste, Italy, 1996, pp. 1271–1274.

[16] C. C. Ko and C. P. Li, "An adaptive IIR structure for the separation, enhancement, and tracking of multiple sinusoids," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2832–2834, Oct. 1994.

[17] M. Jury, *Theory and Application of the $\mathcal{Z}$-Transform Method*. New York, NY: Wiley, 1964.

[18] P. Tichavský, "Posterior Cramér–Rao bound for adaptive harmonic retrieval," *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1299–1302, May 1995.

[19] M. V. Dragoŝević and S. S. Stanković, "An adaptive notch filter with improved tracking properties," *IEEE Trans. Signal Process.*, vol. 43, no. 9, pp. 2068–2077, Sep. 1995.

[20] ——, "Fully adaptive constrained notch filter for tracking multiple frequencies," *Electron. Lett.*, vol. 31, pp. 1215–1217, 1995.

[21] M. Niedźwiecki, "Identification of nonstationary stochastic systems using parallel estimation schemes," *IEEE Trans. Autom. Control*, vol. 35, no. 3, pp. 329–334, Mar. 1990.

[22] ——, "Multiple-model approach to finite memory adaptive filtering," *IEEE Trans. Signal Process.*, vol. 40, no. 2, pp. 470–473, Feb. 1992.

[23] M. Niedźwiecki and L. Guo, "Nonasymptotic results for finite-memory WLS filters," *IEEE Trans. Autom. Control*, vol. 36, no. 2, pp. 198–206, Feb. 1991.

[24] H. L. van Trees, *Detection, Estimation and Modulation Theory*. New York, NY: Wiley, 1968.

[25] K. J. Kushner, *Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic System Theory*. Cambridge, MA: MIT Press, 1984.

Maciej Niedźwiecki was born in Poznań, Poland, in 1953. He received the M.Sc. and Ph.D. degrees from the 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.

From 1986 to 1989, he spent three years as a Research Fellow with the Department of Systems Engineering, Australian National University. From 1990 to 1993, he served as a Vice Chairman of Technical Committee on Theory of the International Federation of Automatic Control (IFAC). He is the author of the book *Identification of Time-varying Processes* (New York: Wiley, 2000). He currently works as 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 the Gdańsk University of Technology, Gdańsk, Poland, in 2000. He has also graduated from the European Master's Degree Course in control and management of lean manufacturing in network systems conducted in cooperation between Gdańsk University of Technology; the Catholic University of Louvain, Belgium; and the University of Karlsruhe, Germany. He is currently working toward the Ph.D. degree at the Gdańsk University of Technology.

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.