Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

On frequency tracking properties of a generalized adaptive notch filter

Maciej Niedźwiecki
Faculty of Electronics, Telecommunications and Computer Science
Department of Automatic Control, Gdańsk University of Technology
ul. Narutowicza 11/12, Gdańsk, Poland
maciekn@eti.pg.gda.pl

***Abstract*— The paper presents results of local performance analysis of a generalized adaptive notch filter (GANF). Generalized adaptive notch filters are used for identification/tracking of quasi-periodically varying dynamic systems and can be considered extension, to the system case, of classical adaptive notch filters. The frequency tracking properties of the algorithm are studied analytically using a direct averaging approach and an approximating linear filter (ALF) 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. Additionally, it allows one to formulate some useful rules of thumb for choosing design parameters. We show that under the conditions of the ALF approximation, the optimally tuned GANF is a statistically efficient estimator of a slowly drifting system frequency.**

I. INTRODUCTION

Generalized adaptive notch filters [1], [2], [3], were designed for the purpose of identification/tracking of quasi-periodically varying complex-valued systems, i.e. systems governed by

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

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

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

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 $v(t) = v_R(t) + jv_I(t)$, $\mathrm{E}[v_R^2(t)] = \mathrm{E}[v_I^2(t)] = \sigma_v^2/2$, $\mathrm{E}[v_R(y)v_I(t)] = 0$, $\forall t$, is a complex white noise of variance σ_v^2, independent of the sequence of regression vectors $\boldsymbol{\varphi}(t)$.

This work was supported by KBN under Grant 4 T11A 01225.

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

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

One of interesting applications, which admits such problem formulation, is identification of multipath (e.g. mobile radio) channels - see e.g. [4], [5] and [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).

The problem of identification of quasi-periodically varying systems can be considered generalization, to the system case, of a classical signal processing task of either elimination or extraction of nonstationary sinusoidal signals buried in noise. The problem of elimination and extraction of complex sinusoidal signals (called cisoids) buried in noise was considered by many authors - see e.g. [7], [8] and the references therein. From different algorithms capable of tracking both complex amplitudes and frequencies in a system governed by (1) - (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

$$\begin{aligned}
\widehat{f}_i(t) &= e^{j\widehat{\omega}_i(t)}\widehat{f}_i(t-1) \\
\widehat{\boldsymbol{\psi}}_i(t) &= \widehat{f}_i(t)\boldsymbol{\varphi}(t) \\
i &= 1, \dots, k \\
\varepsilon(t) &= y(t) - \widehat{\boldsymbol{\psi}}^{\mathrm{T}}(t)\widehat{\boldsymbol{\alpha}}(t-1) \\
\mathbf{Q}(t) &= \frac{1}{\lambda}\Big[\mathbf{Q}(t-1) \\
&\quad - \frac{\mathbf{Q}(t-1)\widehat{\boldsymbol{\psi}}(t)\widehat{\boldsymbol{\psi}}^{\mathrm{H}}(t)\mathbf{Q}(t-1)}{\lambda + \widehat{\boldsymbol{\psi}}^{\mathrm{H}}(t)\mathbf{Q}(t-1)\widehat{\boldsymbol{\psi}}(t)}\Big]
\end{aligned}$$

$$
\begin{aligned}
\mathbf{k}(t) &= \mathbf{Q}(t)\widehat{\boldsymbol{\psi}}(t) \\
\widehat{\boldsymbol{\alpha}}(t) &= \widehat{\boldsymbol{\alpha}}(t-1) + \mathbf{k}^*(t)\varepsilon(t) \\
g_i(t) &= \mathrm{Im}\{\varepsilon^*(t)\widehat{\boldsymbol{\psi}}_i^{\mathrm{T}}(t)\widehat{\boldsymbol{\alpha}}_i(t-1)\} \\
\widehat{\omega}_i(t+1) &= \widehat{\omega}_i(t) - \eta g_i(t) \\
i &= 1,\ldots,k \\
\widehat{\boldsymbol{\theta}}(t) &= \sum_{i=1}^{k} \widehat{f}_i(t)\widehat{\boldsymbol{\alpha}}_i(t)
\end{aligned} \tag{3}
$$

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

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 $\widehat{\boldsymbol{\alpha}}(0) = \mathbf{0}$ 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 RLS-type recursive estimation algorithms [9].

When examining tracking properties of the algorithm (3) we will rely on the approximating linear filter (ALF) technique, developed by Tichavský and Händel [7] for the purpose of analysis of performance of adaptive notch filters, and on the direct averaging technique [10], widely used in performance/convergence studies of adaptive systems. Even though 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 will formulate some useful tuning rules.

II. Tracking analysis

Before we start analyzing tracking properties of the algorithm (3) we will convert it into a more convenient form by applying the linear time-varying transformation

$$
\widehat{\boldsymbol{\beta}}(t) = \widehat{\mathbf{F}}_n(t)\widehat{\boldsymbol{\alpha}}(t), \quad \mathbf{l}(t) = \widehat{\mathbf{F}}_n^*(t)\mathbf{k}(t)
$$

$$
\mathbf{P}(t) = \widehat{\mathbf{F}}_n^*(t)\mathbf{Q}(t)\widehat{\mathbf{F}}_n(t) \tag{4}
$$

where $\widehat{\mathbf{F}}_n(t) = \widehat{\mathbf{F}}(t) \otimes \mathbf{I}_n$, $\widehat{\mathbf{F}}(t) = \mathrm{diag}\{\widehat{f}_1(t),\ldots,\widehat{f}_k(t)\}$ and $\otimes$ denotes the Kronecker product of the corresponding matrices. Let $\widehat{\mathbf{A}}_n(t) = \widehat{\mathbf{A}}(t) \otimes \mathbf{I}_n$, $\widehat{\mathbf{A}}(t) = \mathrm{diag}\{e^{j\widehat{\omega}_1}(t),\ldots,e^{j\widehat{\omega}_k}(t)\}$ and $\boldsymbol{\varphi}_k(t) = [\underbrace{\boldsymbol{\varphi}^{\mathrm{T}}(t),\ldots,\boldsymbol{\varphi}^{\mathrm{T}}(t)}_{k}]^{\mathrm{T}}$. Using (4) one can express the algorithm (3) in the following equivalent form

$$
\begin{aligned}
\varepsilon(t) &= y(t) - \boldsymbol{\varphi}_k^{\mathrm{T}}(t)\widehat{\mathbf{A}}_n(t)\widehat{\boldsymbol{\beta}}(t-1) \\
\mathbf{P}(t) &= \frac{1}{\lambda}\,\widehat{\mathbf{A}}_n^*(t)\,[\mathbf{P}(t-1) \\
&\quad - \left.\frac{\mathbf{P}(t-1)\boldsymbol{\varphi}_k(t)\boldsymbol{\varphi}_k^{\mathrm{H}}(t)\mathbf{P}(t-1)}{\lambda + \boldsymbol{\varphi}_k^{\mathrm{H}}(t)\mathbf{P}(t-1)\boldsymbol{\varphi}_k(t)}\right]\widehat{\mathbf{A}}_n(t) \\
\mathbf{l}(t) &= \mathbf{P}(t)\boldsymbol{\varphi}_k(t) \\
\widehat{\boldsymbol{\beta}}(t) &= \widehat{\mathbf{A}}_n(t)\widehat{\boldsymbol{\beta}}(t-1) + \mathbf{l}^*(t)\varepsilon(t) \\
g_i(t) &= \mathrm{Im}\{\varepsilon^*(t)e^{j\widehat{\omega}_i(t)}\boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\beta}}_i(t-1)\} \\
\widehat{\omega}_i(t+1) &= \widehat{\omega}_i(t) - \eta g_i(t) \\
i &= 1,\ldots,k \\
\widehat{\boldsymbol{\theta}}(t) &= \sum_{i=1}^{k} \widehat{\boldsymbol{\beta}}_i(t)
\end{aligned} \tag{5}
$$

It should be stressed, that the algorithms (3) and (5) 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 $\widehat{\boldsymbol{\theta}}(t)$.

Tichavský and Händel [7] and Tichavský and Nehorai [8] have shown that the local properties of a wide range of adaptive notch filters can be analyzed using the approximating linear filtering technique. 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. We will use the same tool to analyze (5).

Similarly as in [7], we will consider a single frequency case ($k = 1$) and steady state tracking conditions. Note that in this case $\widehat{\mathbf{A}}_n(t) = e^{j\widehat{\omega}_i(t)}\mathbf{I}_n$, $\boldsymbol{\varphi}_k(t) = \boldsymbol{\varphi}(t)$ and the matrix $\mathbf{P}(t)$ can be written down in an explicit form

$$
\mathbf{P}(t) = \left[\sum_{s=1}^{t} \lambda^{t-s}\boldsymbol{\varphi}(s)\boldsymbol{\varphi}^{\mathrm{H}}(s)\right]^{-1} \tag{6}
$$

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 (6) with its expectation [11]. This results in the following steady state approximation

$$
\mathbf{P}(t) \cong \left[\sum_{s=1}^{t} \lambda^{t-s}\boldsymbol{\Phi}^*\right]^{-1} \underset{t\mapsto\infty}{\longmapsto} \quad (1-\lambda)(\boldsymbol{\Phi}^*)^{-1} \tag{7}
$$

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

Using this approximation the generalized adaptive notch filtering algorithm (5) can be, for a system with a single frequency mode, rewritten in a simplified form

$$
\begin{aligned}
\varepsilon(t) &= y(t) - e^{j\widehat{\omega}(t)}\boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\beta}}(t-1) \\
\widehat{\boldsymbol{\beta}}(t) &= e^{j\widehat{\omega}(t)}\widehat{\boldsymbol{\beta}}(t-1) + \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\varepsilon(t) \\
g(t) &= \mathrm{Im}\{\varepsilon^*(t)e^{j\widehat{\omega}(t)}\boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\beta}}(t-1)\} \\
\widehat{\omega}(t+1) &= \widehat{\omega}(t) - \eta g(t) \\
\widehat{\boldsymbol{\theta}}(t) &= \widehat{\boldsymbol{\beta}}(t)
\end{aligned} \tag{8}
$$

which will be a subject of our further analysis.

We will assume that $\boldsymbol{\theta}(t) = \boldsymbol{\alpha}_o e^{j\sum_{s=1}^{t}\omega(s)}$ is a constant-modulus quasi-periodically varying parameter vector, i.e. $\boldsymbol{\theta}(t) = \boldsymbol{\beta}(t) = \boldsymbol{\beta}(t-1)e^{i\omega(t)}$, where $\boldsymbol{\beta}(0) = \boldsymbol{\beta}_o = \boldsymbol{\alpha}_o$.

Let

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

Note that $e(t)$ is a complex-valued white noise with variance $\sigma_e^2=\mathrm{E}[|e(t)|^2]=\boldsymbol{\beta}_o^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_o\sigma_v^2$. 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 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, and that the sequence of regression vectors $\boldsymbol{\varphi}(t)$, independent of $v(t)$ and $w(t)$, is wide-sense stationary and persistently exciting. Then the generalized adaptive notch filtering algorithm (8) applied to the system governed by

$$y(t)=\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\beta}(t)+v(t),\quad \boldsymbol{\beta}(t)=e^{j\omega(t)}\boldsymbol{\beta}(t-1) \tag{9}$$

with $\boldsymbol{\beta}(0)=\boldsymbol{\beta}_o$, can be approximately described by the following linear filtering equations

$$\begin{aligned}\Delta\widehat{\phi}_I(t) &= \lambda\Delta\widehat{\phi}_I(t-1)+\lambda b^2\Delta\widehat{\omega}(t)+\mu e_I(t)\\ \Delta\widehat{\omega}(t+1) &= \delta\Delta\widehat{\omega}(t)-\eta\Delta\widehat{\phi}_I(t-1)\\ &+ \eta e_I(t)-w(t+1)\end{aligned} \tag{10}$$

where $b^2=\boldsymbol{\beta}_o^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_o$.

All approximations hold for sufficiently high signal-to-noise ratio ($\mathrm{SNR}=b^2/\sigma_v^2\gg 1$) and for sufficiently low rate of frequency changes compared with 1/SNR ($\sqrt{b\sigma_w/\sigma_v}\ll 1$).

Proof: See Appendix

Solving the approximating linear equations (10) with respect to $\Delta\widehat{\omega}(t)$ one obtains

$$\Delta\widehat{\omega}(t)=H_1(q^{-1})e_I(t)+H_2(q^{-1})w(t) \tag{11}$$

where q^{-1} denotes the backward shift operator ($q^{-1}x(t)=x(t-1)$) and

$$\begin{aligned}H_1(q^{-1}) &= \frac{(1-\delta)(1-q^{-1})q^{-1}}{b^2(1-(\lambda+\delta)q^{-1}+\lambda q^{-2})}\\ H_2(q^{-1}) &= -\frac{1-\lambda q^{-1}}{1-(\lambda+\delta)q^{-1}+\lambda q^{-2}}\end{aligned} \tag{12}$$

It is easy to check that for any λ and δ from the interval (0,1) the poles of both transfer functions in (12) lie inside the unit circle in the complex plane. Hence, under the constraint mentioned above, the approximating linear filter associated with (8) is asymptotically stable.

A. 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, independent of $v(t)$, with variance σ_w^2. Then, using standard results from the linear filtering theory, one arrives at

$$\mathrm{E}[(\Delta\widehat{\omega}(t))^2]=I[H_1(z)]\ \mathrm{E}[e_I^2(t)]+I[H_2(z)]\ \mathrm{E}[w^2(t)]$$

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. [12]) one obtains

$$I[H_1(z)]=\frac{2(1-\delta)^2}{b^4(1-\lambda)(1+2\lambda+\delta)}\cong\frac{\gamma^2}{2b^4\mu}$$

$$I[H_2(z)]=\frac{(1-\lambda)^2(1+\lambda)+2\lambda(1-\delta)}{(1-\lambda)(1-\delta)(1+2\lambda+\delta)}\cong\frac{\mu}{2\gamma}+\frac{1}{2\mu}$$

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

Finally, after combining all earlier results, one arrives at the following expression for the steady state mean-squared frequency estimation error

$$\mathrm{E}[(\widehat{\omega}(t)-\omega(t))^2]\cong\frac{\gamma^2}{4b^2\mu}\,\sigma_v^2+\left[\frac{\mu}{2\gamma}+\frac{1}{2\mu}\right]\sigma_w^2 \tag{13}$$

Observe that the derived formula contains 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) [9].

Denote by μ_ω and γ_ω the values of μ and γ that minimize the mean-squared frequency estimation error. Straightforward calculations yield

$$\mu_\omega=\sqrt[4]{8\xi},\quad \gamma_\omega=\sqrt{2\xi}$$

$$\mathrm{E}[(\widehat{\omega}(t)-\omega(t))^2|\mu_\omega,\gamma_\omega]\cong\sqrt[4]{2\xi^{-1}}\ \sigma_w^2 \tag{14}$$

where

$$\xi=\frac{b^2\sigma_w^2}{\sigma_v^2} \tag{15}$$

Note that the optimal values of design parameters and the best achievable performance are functions of a scalar coefficient ξ - a product of the signal-to-noise ratio 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 (3).

Recall that γ, equal to $b^2\eta$, is a function of $b^2=\boldsymbol{\beta}_o^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_o$. Therefore, unless $|\boldsymbol{\beta}(t)|$ is constant and the sequence of regression vectors $\boldsymbol{\varphi}(t)$ is wide-sense stationary (which we have been assuming so far), and the quantities $\boldsymbol{\beta}_o$ and $\boldsymbol{\Phi}$ are 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 (3) with the normalized correction term

$$\bar{g}(t) = \frac{g(t)}{\widehat{b}^2(t)} = \mathrm{Im}\left[\frac{\varepsilon^*(t)e^{j\widehat{\omega}(t)}\boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\beta}}(t-1)}{\widehat{b}^2(t)}\right] \qquad (16)$$

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

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

where $0 \leq \lambda_o < 1$ 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.

Our last comment will be devoted to the problem of choice of design variables μ and γ (or equivalently λ and δ). 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 (14), have some practical relevance as they suggest useful tuning rules. Observe that the optimal setting γ_ω is proportional to the square of the optimal setting μ_ω: $\gamma_\omega = \mu_\omega^2/2$. Therefore, to make tuning easier it may be worthwhile to set $\gamma = \mu^2/2$. The problem is then reduced to selection of a single design parameter μ.

B. Statistical efficiency

Consider a system governed by (9) 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. Additionally, 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 $\widehat{\boldsymbol{\omega}} = g[Y(t), U(t)] = [\widehat{\omega}(1), \dots, \widehat{\omega}(t)]^{\mathrm{T}}$ 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 [14], [13]

$$\mathrm{E}\left[(\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega})(\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega})^{\mathrm{T}}\right] \geq \mathbf{J}_t^{-1} \qquad (17)$$

where

$$\mathbf{J}_t^{-1} = -\mathrm{E}\left[\nabla^2_{\boldsymbol{\omega}\boldsymbol{\omega}} \log p(Y(t), U(t), \boldsymbol{\omega})\right] \qquad (18)$$

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 (17) implies that

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

where $\widehat{\boldsymbol{\omega}}_i = \widehat{\omega}(i)$ denotes the ith component of $\widehat{\boldsymbol{\omega}}$. Hence, the limiting steady state value of the mean-squared frequency estimation error (called posterior Cramér-Rao bound in [13]) can be obtained by examining $\lim_{t\mapsto\infty}\left[\mathbf{J}_t^{-1}\right]_{tt}$, which leads to the following result

Proposition 2

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

$$\lim_{t\mapsto\infty} \inf \mathrm{E}[(\widehat{\omega}(t) - \omega(t))^2] = f(z)\sigma_w^2 \qquad (19)$$

where $z = \xi + \sqrt{\xi^2 + 8\xi}$ and $f(z) = \sqrt{1 + 4z^{-1}}$.

Proof: The proof is a pretty straightforward extension of the proof presented in [13] for the signal case ($n = 1$, $\varphi(t) \equiv 1$). It is omitted because the lack of space.

Note that when $\sqrt{\xi} \ll 1$ it holds that $z \cong \sqrt{8\xi}$ and $f(z) \cong \sqrt{4z^{-1}} \cong \sqrt[4]{2\xi^{-1}}$. Comparison of the resulting posterior Cramér-Rao bound $\lim_{t\mapsto\infty} \inf \mathrm{E}[(\widehat{\omega}(t)-\omega(t))^2] = \sqrt[4]{2\xi^{-1}}\,\sigma_w^2$ with (14) implies that the analyzed algorithm is, in the range of applicability of the ALF approximation, a *statistically efficient* procedure for estimation/tracking of a slowly drifting system frequency.

III. Simulation results

Several simulation experiments were performed to verify results of theoretical analysis presented in Section 2. The results summarized below were obtained for a time-varying two-tap FIR system (inspired by channel equalization applications) 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 (σ_v^2=20, $2\sqrt{10}$, 2 and 0.2) to check tracking performance of the GANF algorithm under different SNR conditions (0dB, 5dB, 10dB and 20dB, respectively).

According to (14), 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 $\eta_\omega = \gamma_\omega/\boldsymbol{\beta}_o^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_o$). Figure 1 shows comparison of theoretical evaluations, based on (14), with the results of computer simulations. For each SNR the analysis was carried around the optimal 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 Figure 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 10000 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.

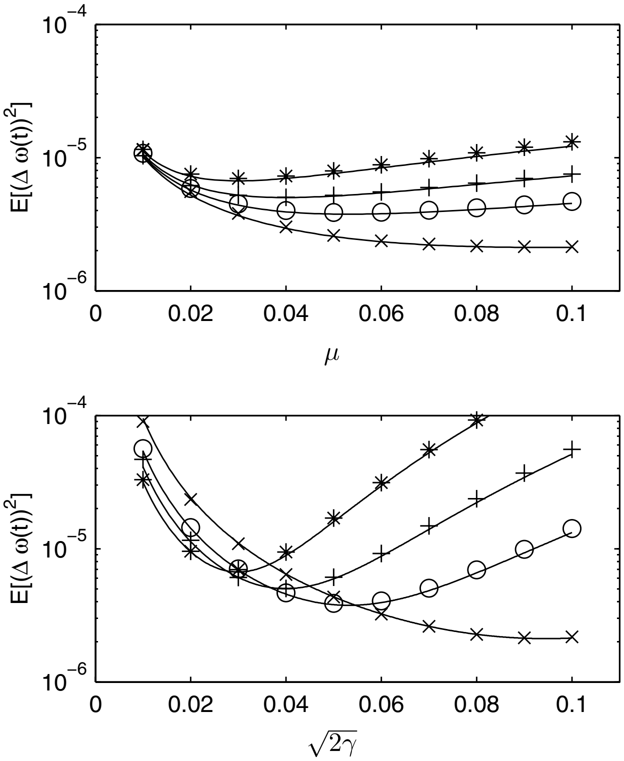

Fig. 1. Variance of the frequency estimation error $\Delta\widehat{\omega}(t)$ for an FIR system with a single frequency mode subject to a random walk drift. The theoretical results (solid lines) are compared with simulation results obtained for different values of μ given $\gamma = \gamma_\omega$ (upper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

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

Finally, Figure 3 shows the evolution of $\Delta\widehat{\omega}(t)$, $\Delta\widehat{\phi}(t)$ 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 the Appendix, the quantities $\Delta\widehat{\phi}(t)$ (a linear combination of the elements of the modified error $\Delta\widetilde{\boldsymbol{\beta}}(t)$, defined in Appendix I) and $\Delta\widehat{\omega}(t)$ vary slowly compared to $u(t)$, i.e. compared to $\boldsymbol{\varphi}(t)$. This confirms validity of the averaging approach which was used to derive the approximating linear filter in the system case.

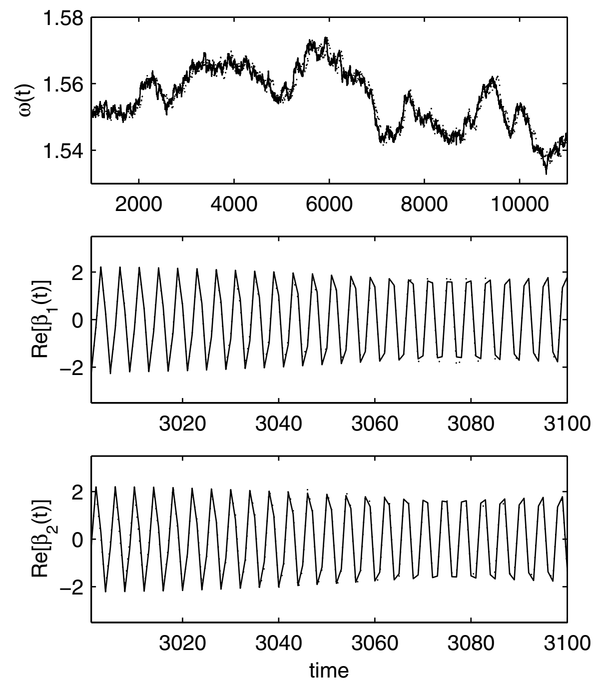

Fig. 2. Typical results of system frequency tracking (upper plot) and system parameter tracking (two lower plots). Solid lines depict true values and dotted lines show evolution of the corresponding estimates.

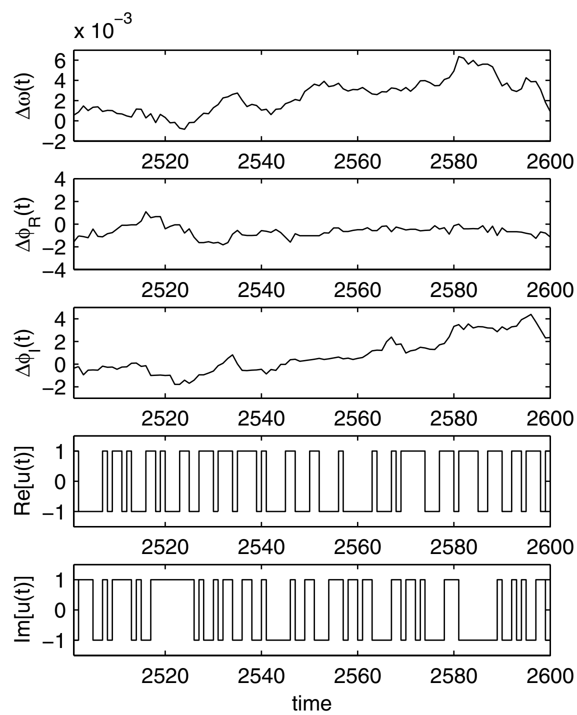

Fig. 3. Evolution of $\Delta\widehat{\omega}(t)$, $\Delta\widehat{\phi}(t)$ and $u(t)$ for a typical run of the optimally tuned GANF algorithm.

References

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

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

[3] Niedźwiecki, M. and P. Kaczmarek (2004). Identification of quasi-periodically varying systems using the combined nonparametric/parametric approach. To appear in *IEEE Trans. on Signal Processing*; see also *Proc. 10th IEEE Int. Conf. on Methods and Models in Automation and Robotics*, Międzyzdroje, Poland, 2004, 1045–1050.

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

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

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

[7] Tichavský P. and P. Händel (1995). Two algorithms for adaptive retrieval of slowly time-varying multiple cisoids in noise. *IEEE Trans. on Signal Processing*, **vol. 43**, 1116–1127.

[8] Tichavský P. and A. Nehorai (1997). Comparative study of four adaptive frequency trackers. *IEEE Trans. on Signal Processing*, **vol. 45**, 1473–1484.

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

[10] Kushner K.J. (1984). *Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic System Theory*. MIT Press. Cambridge, Mass.

[11] Niedźwiecki, M. and Guo L. (1991). Nonasymptotic results for finite-memory WLS filters. *IEEE Trans. on Automatic Control*, **vol. 36**, 198–206.

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

[13] Tichavský P. (1995). Posterior Cramér-Rao bound for adaptive harmonic retrieval. *IEEE Trans. on Signal Processing*, **vol. 43**, 1299–1302.

[14] van Trees H.L. (1968). *Detection, Estimation and Modulation Theory*. Wiley. New York.

APPENDIX

It is straightforward to check that

$$\Delta\widehat{\boldsymbol{\beta}}(t) = e^{j\widehat{\omega}(t)}\widehat{\boldsymbol{\beta}}(t-1) - \boldsymbol{\beta}(t) + \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\varepsilon(t)$$

$$= \left(\mathbf{I}_n - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\right)\left[e^{j\widehat{\omega}(t)}\widehat{\boldsymbol{\beta}}(t-1) - \boldsymbol{\beta}(t)\right]$$

$$+\mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)v(t)$$

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

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

and

$$\Delta\widehat{\boldsymbol{\beta}}(t) \cong \left(\mathbf{I}_n - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\right)e^{j\omega(t)}\Delta\widehat{\boldsymbol{\beta}}(t-1)$$

$$+j\left(\mathbf{I}_n - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\right)\boldsymbol{\beta}(t)\Delta\widehat{\omega}(t)$$

$$+\mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)v(t)$$

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

$$\Delta\widetilde{\boldsymbol{\beta}}(t) \cong \left(\mathbf{I}_n - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\right)\Delta\widetilde{\boldsymbol{\beta}}(t-1)$$

$$+j\left(\mathbf{I}_n - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\right)\boldsymbol{\beta}_o\Delta\widehat{\omega}(t)$$

$$+\mu\boldsymbol{\Phi}^{-1}f^*(t)\boldsymbol{\varphi}^*(t)v(t)$$

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

Assuming that the quantities $\Delta\widetilde{\boldsymbol{\beta}}(t)$ and $\Delta\widehat{\omega}(t)$ change slowly compared to $\boldsymbol{\varphi}(t)$, approximate analysis of the modified estimation error $\Delta\widetilde{\boldsymbol{\beta}}(t)$ can be carried out using the direct averaging technique [10]. The averaging technique was proposed and used for analysis of slowly varying adaptive systems. Since the system (9) rapidly varies with time, some additional arguments are needed to justify the slow variation assumption mentioned above. Note that $\Delta\widetilde{\boldsymbol{\beta}}(t) = f^*(t)\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_o$. It can be shown that, for small values of μ and γ, $f^*(t)\widehat{\boldsymbol{\beta}}(t)$ varies slowly compared to $\widehat{f}^*(t)\widehat{\boldsymbol{\beta}}(t) = \widehat{\boldsymbol{\alpha}}(t)$, which is itself a slowly varying quantity (in the case considered $\widehat{\boldsymbol{\alpha}}(t)$ is a long-memory estimator of a time-invariant coefficient vector $\boldsymbol{\alpha}_o = \boldsymbol{\beta}_o$). Note, in particular, that variation of $\Delta\widetilde{\boldsymbol{\beta}}(t)$ is much slower than variation of $\boldsymbol{\beta}(t)$ and $\widehat{\boldsymbol{\beta}}(t)$. Similar analysis can be carried for $\Delta\widehat{\omega}(t)$ (see Section III 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)

$$\Delta\widetilde{\boldsymbol{\beta}}(t) \cong \lambda\Delta\widetilde{\boldsymbol{\beta}}(t-1) + j\lambda\boldsymbol{\beta}_o\Delta\widehat{\omega}(t)$$

$$+\mu\boldsymbol{\Phi}^{-1}f^*(t)\boldsymbol{\varphi}^*(t)v(t)$$

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

$$\boldsymbol{\beta}^{\mathrm{H}}(t)\boldsymbol{\Phi}\Delta\widehat{\boldsymbol{\beta}}(t) \cong \lambda\boldsymbol{\beta}^{\mathrm{H}}(t-1)\boldsymbol{\Phi}\Delta\widehat{\boldsymbol{\beta}}(t-1)$$

$$+j\lambda\boldsymbol{\beta}_o^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_o\Delta\widehat{\omega}(t) + \mu\boldsymbol{\beta}^{\mathrm{H}}(t)\boldsymbol{\varphi}^*(t)v(t)$$

from which the first equation of (10) follows immediately. Applying the same technique to the frequency update recursions one arrives at

$$\Delta\widehat{\omega}(t+1) \cong \left(1 - \eta\,\boldsymbol{\beta}_o^{\mathrm{H}}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\beta}_o\right)\Delta\widehat{\omega}(t)$$

$$-\eta\,\mathrm{Im}\left[\boldsymbol{\beta}_o^{\mathrm{H}}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\Delta\widetilde{\boldsymbol{\beta}}(t-1)\right]$$

$$+\eta\,\mathrm{Im}\left[\boldsymbol{\beta}^{\mathrm{H}}(t)\boldsymbol{\varphi}^*(t)v(t)\right] - w(t+1)$$

Finally, using the averaging technique one obtains

$$\Delta\widehat{\omega}(t+1) \cong (1 - \eta\,\boldsymbol{\beta}_o^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\beta}_o)\Delta\widehat{\omega}(t)$$

$$-\eta\,\mathrm{Im}\left[\boldsymbol{\beta}^{\mathrm{H}}(t-1)\boldsymbol{\Phi}\Delta\widehat{\boldsymbol{\beta}}(t-1)\right]$$

$$+\eta\,\mathrm{Im}\left[\boldsymbol{\beta}^{\mathrm{H}}(t)\boldsymbol{\varphi}^*(t)v(t)\right] - w(t+1)$$

which constitutes the second equation of (10).