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

Generalized adaptive notch filters - does gradient smoothing technique help?

Maciej Niedźwiecki and Piotr Kaczmarek
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, piokacz@proterians.net.pl

Abstract— **Generalized adaptive notch filters (GANF) 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. We analyze the enhanced GANF algorithm, proposed in the literature, which incorporates gradient smoothing. We show, both analytically and by means of computer simulation, that gradient smoothing does not improve tracking performance of generalized adaptive notch filters.**

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

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

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

We will assume that both the amplitudes $a_{li}(t), l = 1, \ldots, n$ 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 $\boldsymbol{\varphi}(t)$. One of interesting applications, which admits such problem formulation, is identification of multipath (e.g. mobile radio) channels - see e.g. [4], [5], [6].

Denote by $\boldsymbol{\alpha}_i(t) = [a_{1i}(t), \ldots, a_{ni}(t)]^{\mathrm{T}}$ the vector of system coefficients associated with a particular frequency ω_i, and let $f_i(t) = e^{j\sum_{s=1}^{t}\omega_i(s)}$. Finally,let $\boldsymbol{\beta}_i(t) = f_i(t)\boldsymbol{\alpha}_i(t)$, $i = 1, \ldots, k$ and $\boldsymbol{\beta}(t) = [\boldsymbol{\beta}_1^{\mathrm{T}}, \ldots, \boldsymbol{\beta}_k^{\mathrm{T}}]^{\mathrm{T}}$. Using the shorthand notation introduced above, (1) and (2) can be rewritten in the form

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

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

$$\begin{aligned}
\varepsilon(t) &= y(t) - \boldsymbol{\varphi}_k^{\mathrm{T}}(t)\widehat{\mathbf{A}}_n(t)\widehat{\boldsymbol{\beta}}(t-1) \\
\mathbf{P}(t) &= \frac{1}{\lambda}\,\widehat{\mathbf{A}}_n^*(t)\,\Big[\mathbf{P}(t-1) \\
&\quad - \frac{\mathbf{P}(t-1)\boldsymbol{\varphi}_k(t)\boldsymbol{\varphi}_k^{\mathrm{H}}(t)\mathbf{P}(t-1)}{\lambda + \boldsymbol{\varphi}_k^{\mathrm{H}}(t)\mathbf{P}(t-1)\boldsymbol{\varphi}_k(t)}\Big]\,\widehat{\mathbf{A}}_n(t) \\
\mathbf{l}(t) &= \mathbf{P}(t)\boldsymbol{\varphi}_k(t) \\
\widehat{\boldsymbol{\beta}}(t) &= \widehat{\mathbf{A}}_n(t)\widehat{\boldsymbol{\beta}}(t-1) + \mathbf{l}^*(t)\varepsilon(t) \\
g_i(t) &= \mathrm{Im}\{\varepsilon^*(t)e^{j\widehat{\omega}_i(t)}\boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\beta}}_i(t-1)\} \\
\widehat{\omega}_i(t+1) &= \widehat{\omega}_i(t) - \eta g_i(t) \\
i &= 1, \ldots, k \\
\widehat{\boldsymbol{\theta}}(t) &= \sum_{i=1}^{k} \widehat{\boldsymbol{\beta}}_i(t)
\end{aligned} \qquad (3)$$

where $\widehat{\mathbf{A}}_n(t) = \widehat{\mathbf{A}}(t) \otimes \mathbf{I}_n$, $\widehat{\mathbf{A}}(t) = \mathrm{diag}\{e^{j\widehat{\omega}_1(t)}, \ldots, e^{j\widehat{\omega}_k(t)}\}$ and $\boldsymbol{\varphi}_k(t) = [\underbrace{\boldsymbol{\varphi}^{\mathrm{T}}(t), \ldots, \boldsymbol{\varphi}^{\mathrm{T}}(t)}_{k}]^{\mathrm{T}}$.

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. Generally speaking, both design parameters should be chosen so as to trade-off the tracking speed of a generalized adaptive notch filter (which decreases with growing λ and increases with growing η) and its noise rejection capability (which increases with growing λ and decreases with growing η).

The initial conditions for (3) should be set to $\widehat{\boldsymbol{\beta}}(0) = \mathbf{0}$ and $\mathbf{P}(0) = c\mathbf{I}_{kn}$, where $\mathbf{I}_{kn}$ denotes the $kn \times kn$ identity

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

matrix and c is a large positive constant, which is a standard initialization procedure for all RLS-type recursive estimation algorithms [9].

Derivation of (3) is two-step. First, the problem of tracking of $\boldsymbol{\beta}(t)$ is solved for a system with fixed frequency modes (constant frequencies). Then, the frequency adaptation loop is added by performing the gradient-type minimization of the instantaneous measure of fit

$$J(t,\boldsymbol{\omega}) = |y(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\beta}}(t-1,\boldsymbol{\omega})|^2 = |\varepsilon(t,\boldsymbol{\omega})|^2$$

where $\boldsymbol{\omega} = [\omega_1, \dots, \omega_k]^{\mathrm{T}}$.

In order to smooth out the frequency estimates, and hence to improve parameter tracking, one may attempt to replace the instantaneous measure of fit with its locally averaged counterpart

$$\bar{J}(t,\boldsymbol{\omega}) = \nu\bar{J}(t-1,\boldsymbol{\omega}) + (1-\nu)J(t,\boldsymbol{\omega})$$

where $0 \le \nu < 1$ is the coefficient which decides upon the degree of smoothing. All that is needed to incorporate this change, is replace the "instantaneous" gradient terms $g_i(t)$ in (3) with the smoothed gradient terms $\bar{g}_i(t)$

$$\begin{aligned} \bar{g}_i(t) &= \nu\bar{g}_i(t-1) + (1-\nu)g_i(t) \\ \widehat{\omega}_i(t+1) &= \widehat{\omega}_i(t) - \eta\bar{g}_i(t) \end{aligned} \tag{4}$$

Note that for $\nu = 0$ the smoothing action is switched off, i.e. the modified algorithm becomes identical with the original algorithm.

The gradient smoothing technique was advocated by Tsatsanis and Giannakis in [4] and seems to be a natural way of enhancing tracking capabilities of generalized adaptive notch filters [1]. We will show that - contrary to such expectations - gradient smoothing always *worsens* tracking performance of the GANF algorithm, and hence there seems to be no good reason to use it.

II. Analysis of the signal identification algorithm

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. Actually, note that when $n = 1$ and $\varphi(t) = 1, \forall t$ the model (1) - (2) becomes a description of a noisy nonstationary multifrequency signal $s(t) = \theta(t)$

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

$$= \sum_{i=1}^{k} \beta_i(t) + v(t) \tag{5}$$

[1]Even though the form of smoothing (local arithmetic mean) proposed in [4] is slightly different from the one considered here, the qualitative effects, discussed below, are obviously the same.

In this special case the generalized adaptive notch filter (3) - (4) becomes an "ordinary" adaptive notch filter

$$\begin{aligned} \varepsilon(t) &= y(t) - \mathbf{1}_k^{\mathrm{T}}\widehat{\mathbf{A}}(t)\widehat{\boldsymbol{\beta}}(t-1) \\ \mathbf{P}(t) &= \frac{1}{\lambda}\,\widehat{\mathbf{A}}^*(t)\left[\mathbf{P}(t-1) \right. \\ &\quad - \left. \frac{\mathbf{P}(t-1)\mathbf{1}_k\mathbf{1}_k^{\mathrm{T}}\mathbf{P}(t-1)}{\lambda + \mathbf{1}_k^{\mathrm{T}}\mathbf{P}(t-1)\mathbf{1}_k}\right]\widehat{\mathbf{A}}(t) \\ \mathbf{l}(t) &= \mathbf{P}(t)\mathbf{1}_k \\ \widehat{\boldsymbol{\beta}}(t) &= \widehat{\mathbf{A}}(t)\widehat{\boldsymbol{\beta}}(t-1) + \mathbf{l}^*(t)\varepsilon(t) \\ g_i(t) &= \mathrm{Im}\{\varepsilon^*(t)e^{j\widehat{\omega}_i(t)}\widehat{\beta}_i(t-1)\} \\ \bar{g}_i(t) &= \nu\bar{g}_i(t-1) + (1-\nu)g_i(t) \\ \widehat{\omega}_i(t+1) &= \widehat{\omega}_i(t) - \eta\bar{g}_i(t) \\ i &= 1, \dots, k \end{aligned}$$

$$\widehat{s}(t) = \sum_{i=1}^{k}\widehat{\beta}_i(t) \tag{6}$$

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

The problem of elimination and extraction of complex sinusoidal signals (called cisoids) buried in noise was considered by many authors - see e.g. [7], [8] and the references therein. We will analyze (6) using the approximating linear filter (ALF) technique, introduced in [7]. 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.

A. Theoretical analysis

Similarly as in [7], we will consider the single frequency case ($k = 1$) and steady state tracking conditions. 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 (6) in a much simpler form

$$\begin{aligned} \varepsilon(t) &= y(t) - e^{j\widehat{\omega}(t)}\widehat{\beta}(t-1) \\ \widehat{\beta}(t) &= e^{j\widehat{\omega}(t)}\widehat{\beta}(t-1) + (1-\lambda)\varepsilon(t) \\ g(t) &= \mathrm{Im}\{\varepsilon^*(t)e^{j\widehat{\omega}(t)}\widehat{\beta}(t-1)\} \\ \bar{g}(t) &= \nu\bar{g}(t-1) + (1-\nu)g(t) \\ \widehat{\omega}(t+1) &= \widehat{\omega}(t) - \eta\bar{g}(t) \\ \widehat{s}(t) &= \widehat{\beta}(t) \end{aligned} \tag{7}$$

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

$$\begin{aligned} \Delta\widehat{\phi}(t) &= \beta^*(t)\Delta\widehat{\beta}(t) = \Delta\widehat{\phi}_R(t) + j\Delta\widehat{\phi}_I(t) \\ e(t) &= \beta^*(t)v(t) = e_R(t) + je_I(t) \\ w(t+1) &= \omega(t+1) - \omega(t) \end{aligned} \tag{8}$$

Using the technique proposed in [7], 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 (7) applied to signal

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

can be approximately described by the following linear filtering equations

$$\begin{aligned}
\Delta\widehat{\phi}_R(t) &\cong \lambda\Delta\widehat{\phi}_R(t-1) + (1-\lambda)e_R(t)\\
\Delta\widehat{\phi}_I(t) &\cong \lambda\Delta\widehat{\phi}_I(t-1) + \lambda b^2\Delta\widehat{\omega}(t) + (1-\lambda)e_I(t)\\
g(t) &\cong b^2\Delta\widehat{\omega}(t) + \Delta\widehat{\phi}_I(t-1) - e_I(t)\\
\bar{g}(t) &\cong \nu\bar{g}(t-1) + (1-\nu)g(t)\\
\Delta\widehat{\omega}(t+1) &\cong \Delta\widehat{\omega}(t) - w(t+1) - \eta\bar{g}(t)
\end{aligned} \tag{10}$$

where $b = |\beta(t)|$.

Proof: See Appendix.

Denote by q^{-1} the backward shift operator and let $\delta = 1 - \eta b^2$. Solving the approximating linear equations (10) with respect to $\Delta\widehat{\phi}_R(t)$, $\Delta\widehat{\phi}_I(t)$ and $\Delta\widehat{\omega}(t)$ one obtains

$$\begin{aligned}
\Delta\widehat{\phi}_R(t) &= F(q^{-1})e_R(t)\\
\Delta\widehat{\phi}_I(t) &= G_1(q^{-1})e_I(t) + G_2(q^{-1})w(t)\\
\Delta\widehat{\omega}(t) &= H_1(q^{-1})e_I(t) + H_2(q^{-1})w(t)
\end{aligned} \tag{11}$$

where

$$F(q^{-1}) = \frac{1-\lambda}{1-\lambda q^{-1}}$$

$$G_1(q^{-1}) = \frac{M_1(q^{-1})}{D(q^{-1})}\ , \quad G_2(q^{-1}) = -\frac{b^2M_2(q^{-1})}{D(q^{-1})}$$

$$H_1(q^{-1}) = \frac{N_1(q^{-1})}{b^2D(q^{-1})}\ , \quad H_2(q^{-1}) = -\frac{N_2(q^{-1})}{D(q^{-1})}$$

and

$$\begin{aligned}
M_1(q^{-1}) &= (1-\lambda) + [\lambda - \delta + \nu(2-\lambda-\delta)]q^{-1}\\
&+ \nu(1-\lambda)q^{-2}\\
M_2(q^{-1}) &= \lambda(1-\nu q^{-1})\\
N_1(q^{-1}) &= (1-\nu)(1-\delta)(1-q^{-1})q^{-1}\\
N_2(q^{-1}) &= 1 - (\lambda+\nu)q^{-1} + \nu\lambda q^{-2}\\
D(q^{-1}) &= 1 - [\lambda + \delta + \nu(2-\delta)]q^{-1}\\
&+ [\lambda + \nu(1+\lambda)]q^{-2} - \nu\lambda q^{-3}
\end{aligned}$$

It can be checked (see e.g. [12]) that for $0 < \lambda < 1$, $0 < \delta < 1$ and $0 \le \nu < 1$, all approximating filters are asymptotically stable provided that

$$\nu[1 - 3\lambda + \lambda^2 + \lambda\delta] + (1-\nu)(1-\lambda) > 0$$

The sufficient stability condition, which holds for any value of ν ($0 \le \nu < 1$) and stays in a good agreement with the optimization recommendations discussed below, takes the form

$$1 - \delta < \frac{(1-\lambda)^2}{\lambda} \tag{12}$$

Suppose that 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 $v(t)$.

For a constant-modulus signal it holds that

$$\mathrm{E}[|\Delta\widehat{\beta}(t)|^2] = \frac{\mathrm{E}[(\Delta\widehat{\phi}_R(t))^2] + \mathrm{E}[(\Delta\widehat{\phi}_I(t))^2]}{b^2}$$

One can check that $e(t)$, similarly as $v(t)$, is a complex-valued white noise obeying $\sigma_e^2 = \mathrm{E}[|e(t)|^2] = b^2\sigma_v^2$, $\mathrm{E}[e_R^2(t)] = \mathrm{E}[e_I^2(t)] = \sigma_e^2/2$, $\mathrm{E}[e_R(t)e_I(t)] = 0$. Therefore, using standard results from the linear filtering theory, one arrives at

$$\mathrm{E}[(\Delta\widehat{\phi}_R(t))^2] = I[F(z)]\ \mathrm{E}[e_R^2(t)]$$

$$\mathrm{E}[(\Delta\widehat{\phi}_I(t))^2] = I[G_1(z)]\ \mathrm{E}[e_I^2(t)] + I[G_2(z)]\ \mathrm{E}[w^2(t)]$$

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

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.

Suppose that

$$X(z^{-1}) = \frac{B(z^{-1})}{A(z^{-1})}$$

where $A(z^{-1}) = 1 + a_1z^{-1} + a_2z^{-2} + a_3z^{-3}$ and $B(z^{-1}) = b_0 + b_1z^{-1} + b_2z^{-2}$. Then it holds that [12]

$$I[X(z)] = \frac{\det(\Sigma_1)}{\det(\Sigma_2)} \tag{13}$$

where

$$\Sigma_1 = \begin{bmatrix} 1 & a_1 & a_2 & a_3\\ a_1 & 1+a_2 & a_1+a_3 & a_2\\ a_2 & a_3 & 1 & a_1\\ a_3 & 0 & 0 & 1 \end{bmatrix}$$

and

$$\Sigma_2 = \begin{bmatrix} B_0 & a_1 & a_2\\ B_1 & 1+a_2 & a_1+a_3\\ B_2 & a_3 & 1 \end{bmatrix}$$

with $B_0 = b_0^2 + b_1^2 + b_2^2$, $B_1 = 2b_1(b_0+b_2)$ and $B_2 = 2b_0b_2$. When there is no smoothing ($\nu = 0$) the problem is analytically tractable. Straightforward but tedious calculations yield [10]

$$\mathrm{E}[(\widehat{\omega}(t) - \omega(t))^2] \cong \frac{\gamma^2}{4b^2\mu}\ \sigma_v^2 + \left[\frac{\mu}{2\gamma} + \frac{1}{2\mu}\right]\sigma_w^2 \tag{14}$$

$$\mathrm{E}[|\widehat{\beta}(t)-\beta(t)|^2] \cong \left[\frac{\gamma}{4\mu}+\frac{\mu}{2}\right]\sigma_v^2+\frac{b^2}{2\mu\gamma}\,\sigma_w^2 \qquad (15)$$

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

Both mean-squared tracking errors can be minimized with respect to μ and γ. Denote by μ_ω and γ_ω the values of μ and γ which minimize the frequency tracking error (14). Similarly, denote by μ_β and γ_β the analogous settings for which the signal tracking error (15) attains its minimum. Finally, let

$$\xi = \frac{b^2\sigma_w^2}{\sigma_v^2} \qquad (16)$$

be a scalar measure of signal nonstationarity. Note that ξ is a product of the signal-to-noise ratio b^2/σ_v^2 and the variance of frequency changes σ_w^2. It can be shown that for sufficiently small values of ξ it holds that [10]

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

$$\mathrm{E}[(\widehat{\omega}(t)-\omega(t))^2|\mu_\omega,\gamma_\omega] \cong \sqrt[4]{2\xi^{-1}}\,\sigma_w^2 \qquad (17)$$

and

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

$$\mathrm{E}[|\widehat{\beta}(t)-\beta(t)|^2|\mu_\beta,\gamma_\beta] \cong \sqrt[4]{2\xi}\,\sigma_v^2 \qquad (18)$$

Remark

Note that the settings that are optimal from the frequency tracking viewpoint are not the best choice from the signal tracking viewpoint and *vice versa*. Note also that $\gamma_\omega = \mu_\omega^2/2$ and $\gamma_\beta = \mu_\beta^2$, which suggests that setting $\gamma = \mu^2/2$ (when we are primarily interested in frequency tracking) or $\gamma = \mu^2$ (when we are interested in signal tracking) may be a good way of reducing the number of design parameters of a GANF algorithm. Observe that, according to (12), in both cases discussed above the generalized adaptive notch filter is stable for any value of $0 \le \nu < 1$.

B. Numerical study

Unfortunately, when $\nu > 0$ the analytical results obtained using (13) become very complicated and cannot be easily brought down to the form revealing the effects of smoothing. For this reason we decided to analyze the problem numerically.

Figure 1 shows both theoretical and experimental results obtained for a single cisoid with a constant amplitude $b = 1$, embedded in complex-valued white Gaussian noise with variance $\sigma_v^2 = 0.2$ (SNR=7dB). The evolution of the instantaneous 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 $\omega(0) = \pi/2$.

Note that in the case considered $\xi = 5 \cdot 10^{-7}$ and hence for $\nu = 0$ (no smoothing) the optimal settings are $\mu_\omega = 0.045$, $\gamma_\omega = 0.001$ (when frequency tracking is the main objective) and $\mu_\beta = 0.032$, $\gamma_\beta = 0.001$ (when signal tracking/cancellation is our primary task). We examined the effects of smoothing in the neighborhood of the points $(\mu_\omega, \gamma_\omega)$ and $(\mu_\beta, \gamma_\beta)$. Since the amplitude adaptation gain μ does not affect in a direct way the frequency tracking loop, we fixed it at its "optimal" (as long as there is no smoothing) values μ_ω and μ_β, respectively. Hence, the only two parameters that were changed were γ and ν [2].

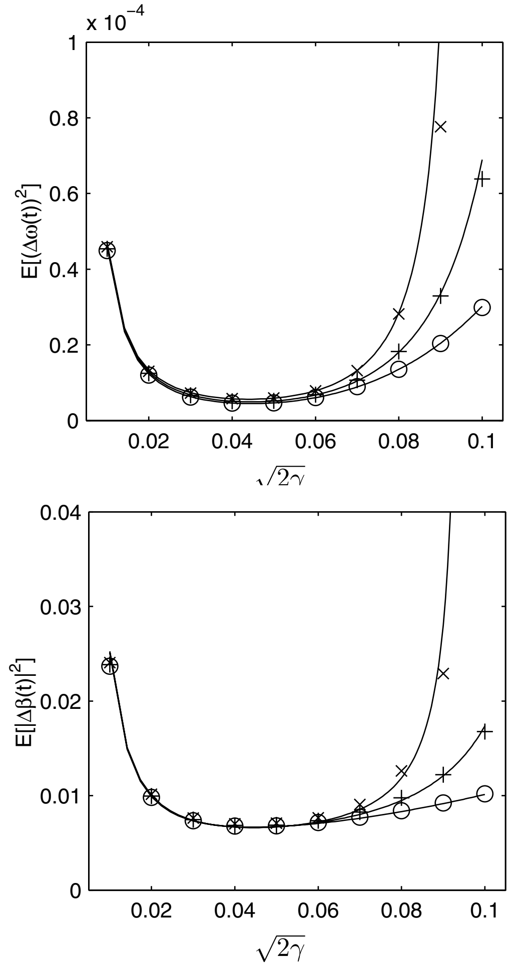

Fig. 1. Variance of the frequency estimation error $\Delta\widehat{\omega}(t)$ (upper figure) and signal estimation error $\Delta\widehat{\beta}(t)$ (lower figure) for a nonstationary noisy cisoid with randomly drifting frequency. The theoretical results (solid lines), obtained for different values of the frequency adaptation gain γ and different values of the smoothing coefficient ν, are compared with simulation results: the symbols ○ , + and × correspond to $\nu = 0$, $\nu = 0.9$ and $\nu = 0.95$, respectively. The amplitude adaptation gains μ were fixed at the values $\mu_\omega = 0.045$ (upper figure) and $\mu_\beta = 0.032$ (lower figure).

The experimental results shown in Figure 1 were obtained by double averaging. First, for a given realization of measurement noise and a given frequency trajectory, the time-averaged squared tracking errors were computed for different pairs (γ, ν) from 10000 iterations of the algorithm (after it has reached its steady state). The single-realization results obtained in this way were next averaged over 50 different realizations of $\{v(t)\}$ and $\{w(t)\}$. Note that the results of simulation experiments stay in a very good agreement with

[2]The full three-dimensional inspection of the error surface, regarded as a function of μ, γ and ν, confirmed that analysis of such two-dimensional cross-sections is representative

theoretical curves.

Both theoretical and experimental plots show clearly that smoothing can only degrade tracking capabilities of the GANF algorithm. While in the close vicinity of the points $(\mu_\omega, \gamma_\omega)$ and $(\mu_\beta, \gamma_\beta)$ smoothing has practically no effect on tracking errors, when settings are less carefully chosen the obtained results are always worse than those yielded by the unmodified GANF algorithm. This observation was confirmed by many numerical studies (not reported here), performed for different values of ξ.

III. Analysis of system identification algorithm

A. Theoretical analysis

Similarly as in Section 2, we will consider the single frequency case ($k = 1$). If the sequence of regression vectors is wide-sense stationary and persistently exciting, and λ is close to 1, one can use the following steady state approximation

$$\mathbf{P}(t) \cong (1-\lambda)(\mathbf{\Phi}^*)^{-1} \tag{19}$$

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

Using (19) the generalized adaptive notch filtering algorithm (3) can be, for a system with a single frequency mode, rewritten in a simplified form

$$\begin{aligned}
\varepsilon(t) &= y(t) - e^{j\widehat{\omega}(t)}\boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\beta}}(t-1) \\
\widehat{\boldsymbol{\beta}}(t) &= e^{j\widehat{\omega}(t)}\widehat{\boldsymbol{\beta}}(t-1) + (1-\lambda)\mathbf{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\varepsilon(t) \\
g(t) &= \mathrm{Im}\{\varepsilon^*(t)e^{j\widehat{\omega}(t)}\boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\beta}}(t-1)\} \\
\bar{g}(t) &= \nu\bar{g}(t-1) + (1-\nu)g(t) \\
\widehat{\omega}(t+1) &= \widehat{\omega}(t) - \eta\bar{g}(t) \\
\widehat{\boldsymbol{\theta}}(t) &= \widehat{\boldsymbol{\beta}}(t)
\end{aligned} \tag{20}$$

Let

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

Using the notation introduced above one can prove:

Proposition 2

Assume that all conditions of ALF analysis are fulfilled 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 (20) applied to the system governed by

$$y(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\beta}(t) + v(t), \quad \boldsymbol{\beta}(t) = e^{j\omega(t)}\boldsymbol{\beta}(t-1) \tag{21}$$

can be approximately described by equations (10) with $b^2 = \boldsymbol{\beta}_o^{\mathrm{H}}\mathbf{\Phi}\boldsymbol{\beta}_o$, where $\boldsymbol{\beta}_o = \boldsymbol{\beta}(0)$.

Proof: Because of the lack of space the proof, which combines the Taylor series approximation technique, exploited in the signal-oriented case (see Appendix), with the direct averaging approach [11], widely used for analysis of adaptive systems, is skipped.

Since the approximating linear filters associated with (20) are identical with the analogous filters derived in Section 2 for the signal identification algorithm, some of the conclusions drawn there extend to the system case. This concerns, for example, analysis of the frequency tracking error for random walk frequency variations. Unfortunately, the parameter tracking results cannot be easily generalized to the system case. This is because - unlike the signal case - the variance of the system tracking error $\varepsilon_s(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\Delta\widehat{\boldsymbol{\beta}}(t)$ (which is a natural extension, to the system case, of the signal tracking error $\Delta\widehat{\beta}(t)$) cannot be explicitly related to the quantity $\mathrm{E}[|\Delta\widehat{\phi}(t)|^2] = \mathrm{E}[|\boldsymbol{\beta}^{\mathrm{H}}(t)\mathbf{\Phi}\Delta\widehat{\boldsymbol{\beta}}(t)|^2]$, which can be evaluated using the ALF-based approach.

B. Numerical study

The results summarized below were obtained for a time-varying two-tap channel-like 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 with variance $\sigma_v^2 = 4$. The impulse response coefficients of the system were modeled as nonstationary cisoids

$$\theta_i(t) = a_i e^{j\sum_{s=1}^{t}\omega(s)}, \; i = 1, 2$$

with time-invariant complex "amplitudes" $\boldsymbol{\alpha} = [a_1, a_2]^{\mathrm{T}} = [2-j \;, \; 1+2j \;]^{\mathrm{T}}$. Note that in this case $\boldsymbol{\beta}_o = \boldsymbol{\alpha}$, $\boldsymbol{\varphi}(t) = [u(t), u(t-1)]^{\mathrm{T}}$, $\mathbf{\Phi} = \mathbf{I}_2\sigma_u^2$ and $\boldsymbol{\beta}_o^{\mathrm{H}}\mathbf{\Phi}\boldsymbol{\beta}_o/\sigma_v^2 = 5$, i.e. SNR=7dB.

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 $\omega(0) = \pi/2$.

The results, depicted in Figures 2 and 3, lead to the same conclusion as that reached in the signal case - even though gradient smoothing increases the number of design degrees of freedom ("tuning knobs") of GANF algorithms, it does not improve their tracking capabilities. Hence it cannot be recommended in practice.

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. *IEEE Trans. on Signal Processing* (accepted); 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.

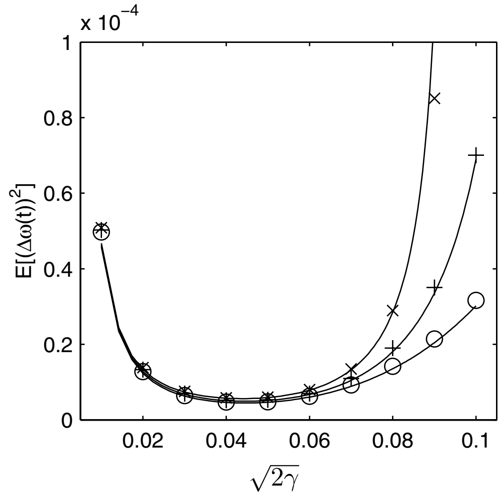

Fig. 2. Variance of the frequency estimation error $\Delta\widehat{\omega}(t)$ for a nonstationary FIR system with a single frequency mode subject to a random walk drift. The theoretical results (solid lines), obtained for different values of the frequency adaptation gain γ and different values of the smoothing coefficient ν, are compared with simulation results: the symbols ○ , + and $\times$ correspond to $\nu = 0$, $\nu = 0.9$ and $\nu = 0.95$, respectively. The amplitude adaptation gain μ was fixed at the value $\mu_\omega = 0.045$.

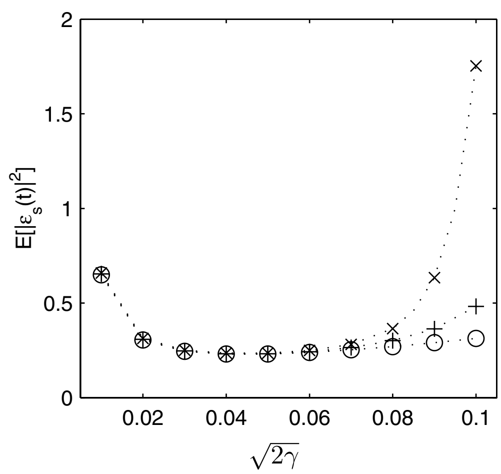

Fig. 3. Variance of the system tracking error $\varepsilon_s(t)$ for a nonstationary FIR system with a single frequency mode subject to a random walk drift. Dotted lines connect experimental points obtained for different values of the frequency adaptation gain γ and different values of the smoothing coefficient ν: the symbols ○ , + and $\times$ correspond to $\nu = 0$, $\nu = 0.9$ and $\nu = 0.95$, respectively. The amplitude adaptation gain was set to $\mu = 0.032$.

[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] Niedźwiecki, M. and P. Kaczmarek (2005). Tracking analysis of a generalized adaptive notch filter. *Technical Report ETI-1/2005*, Gdańsk University of Technology (to appear in *IEEE Trans. on Signal Processing*).

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

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

APPENDIX

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

$$\begin{aligned}\varepsilon(t) &= \beta(t) - e^{j\widehat{\omega}(t)}\widehat{\beta}(t-1) + v(t)\\ \Delta\widehat{\beta}(t) &= e^{j\widehat{\omega}(t)}\widehat{\beta}(t-1) - \beta(t) + \mu\varepsilon(t)\end{aligned}$$

Combining the last two equations one arrives at

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

Note that

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

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 $\Delta\widehat{\beta}(t-1)$ one obtains

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

Therefore

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

and

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

The first two equations of (10) follow directly from the above result.

Similar technique can be used to cope with the frequency update in (7). Note that

$$\varepsilon^*(t)e^{j\widehat{\omega}(t)}\widehat{\beta}(t-1) = e^{j\Delta\widehat{\omega}(t)}\beta^*(t-1)\widehat{\beta}(t-1)$$

$$-|\widehat{\beta}(t-1)|^2 + e^{j\widehat{\omega}(t)}\widehat{\beta}(t-1)v^*(t)$$

and

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

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

$$\cong \beta^*(t-1)\Delta\widehat{\beta}(t-1) + j|\beta(t-1)|^2\Delta\widehat{\omega}(t) + |\beta(t-1)|^2$$

Combining the last two equations, and noting that

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

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

$$\varepsilon^*(t)e^{j\widehat{\omega}(t)}\widehat{\beta}(t-1) \cong \beta^*(t-1)\Delta\widehat{\beta}(t-1) + jb^2\Delta\widehat{\omega}(t)$$

$$+|\beta(t-1)|^2 - |\widehat{\beta}(t-1)|^2 + \beta(t)v^*(t)$$

and

$$g(t) \cong b^2\Delta\widehat{\omega}(t) + \mathrm{Im}\left[\beta^*(t-1)\Delta\widehat{\beta}(t-1) + \beta(t)v^*(t)\right]$$

which is the third equation of (10).

Finally, since $\omega(t+1) = \omega(t) + w(t+1)$, the frequency estimation error can be expressed as

$$\Delta\widehat{\omega}(t+1) = \Delta\widehat{\omega}(t) - \eta\bar{g}(t) - w(t+1)$$

which is the last equation of (10).