**Preprints of the**
**15th IFAC Symposium on System Identification**
**Saint-Malo, France, July 6-8, 2009**

# SELF-OPTIMIZING REAL-VALUED GENERALIZED ADAPTIVE NOTCH FILTER

**Maciej Niedźwiecki** * **Adam Sobociński** *

* *Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control, Gdańsk University of Technology ul. Narutowicza 11/12, Gdańsk , Poland e-mail: maciekn@eti.pg.gda.pl, adsob@eti.pg.gda.pl*

Abstract: Generalized adaptive notch filters (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 a GANF algorithm depend on two adaptation gains, which should be chosen so as to match the degree of nonstationarity of the identified system. First, an analytical study of a tracking performance of a real-valued GANF algorithm is presented. Then, based on the obtained theoretical results, a self-optimizing algorithm is proposed, capable of automatic tuning of its adaptation gains. The paper is an extension of the previous work devoted to complex-valued GANF algorithms.

Keywords: system identification, time-varying processes, frequency estimation

## 1. INTRODUCTION

Consider the following real-valued quasi-periodically varying system

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

where $t = 1, 2, ...$ denotes the normalized discrete time, $y(t)$ denotes the system output, $\boldsymbol{\varphi}(t) = [\varphi_1(t), ..., \varphi_n(t)]^{\mathrm{T}}$ is the regression verctor, $v(t)$ is an additive white noise, uncorrelated with $\boldsymbol{\varphi}(t)$, and $\boldsymbol{\theta}(t) = [\theta_1(t), ..., \theta_2(t)]^{\mathrm{T}}$ denotes the vector of time varying system coefficients, modeled as weighted sums of sinusoidal and cosinusoidal functions:

$$\theta_l(t) = c_{l,0}(t) + \sum_{l=1}^{k}[c_{l,2i-1}(t)\sin\phi_i(t)$$
$$+ c_{l,2i}(t)\cos\phi_i(t)] \ , \quad l = 1, ..., n \qquad (2)$$

where

$$\phi_i(t) = \sum_{\tau=1}^{t} \omega_i(\tau) \qquad (3)$$

It is assumed that, for every i, the quantaties $c_{l,i}$, $l = 1, ..., n$ and $\omega_i(t)$ are slowly time-varying. It is also assumed that the sequence $\{\boldsymbol{\varphi}(t)\}$ is zero-mean, wide-sense stationary and ergodic, with covariance matrix $\boldsymbol{\Phi}$.

The models similar to (1) - (3) arise in many practical applications such as equalization of rapidly fading communication channels and array processing (Tsatsanis & Giannakis, 1996), (Giannakis & Tepedelenlioğlu, 1998), (Bakkoury et al., 2000).

The parameters of the quasi-periodically varying system (1)–(3) can be tracked using generalized adaptive notch filters (GANFs). The name "generalized adaptive notch filters" stems from the fact that in the special signal case ($n = 1$, $\varphi(t) \equiv 1$) GANF algorithms turn into classical adaptive notch filters (ANFs) – devices used to either extract or eliminate nonstationary sinusoidal signals buried in noise.

The first GANF algorithms, in their complex-valued version, were presented in (Niedźwiecki & Kaczmarek, 2005a) and (Niedźwiecki & Kaczmarek, 2005b). The first real-valued GANF algorithm was presented in (Niedźwiecki & Kaczmarek, 2005c) and its frequency tracking properties were analyzed in (Niedźwiecki & Sobociński, 2007). The main purpose of the current study is to extend results presented in (Niedźwiecki &

Kaczmarek, 2006a) for complex-valued GANF algorithms. We will show that, despite some obvious qualitative similarities between the complex case and the real case, analysis of the real-valued algorithm leads to slightly different quantitative results than those obtained earlier for the complex-valued algorithm. It should be stressed that the investigation presented below is not a special case of the analysis carried out in (Niedźwiecki & Kaczmarek, 2006a) – as a matter of fact real-valued algorithms are considerably more difficult to analyze than complex-valued algorithms.

## 2. REAL-VALUED GANF ALGORITHM

As in (Niedźwiecki & Sobociński, 2007), the analytical study will be focused on systems with a single nonzero frequency mode. In such a case

$$\theta_l(t) = c_{l,1}(t)\sin\phi(t) + c_{l,2}(t)\cos\phi(t)$$

$$\phi(t) = \sum_{\tau=1}^{t}\omega(\tau)\ , \quad l = 1,\dots,n \tag{4}$$

The corresponding GANF algorithm can be written down in the form

$$\begin{aligned}
\mathbf{z}(t) &= [\mathbf{I}_n \otimes \widehat{\mathbf{G}}(t)]^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(t-1)\\
\varepsilon(t) &= y(t) - [\boldsymbol{\varphi}(t)\otimes\mathbf{f}(0)]^{\mathrm{T}}\,\mathbf{z}(t)\\
\widehat{\boldsymbol{\beta}}(t) &= \mathbf{z}(t) + \mu\left[(\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}(t))\otimes\mathbf{f}(0)\right]\varepsilon(t)\\
g(t) &= [\boldsymbol{\varphi}(t)\otimes\mathbf{h}(0)]^{\mathrm{T}}\,\mathbf{z}(t)\varepsilon(t)\\
\widehat{\omega}(t+1) &= \widehat{\omega}(t) + \eta g(t)\\
\widehat{\boldsymbol{\theta}}(t) &= [\mathbf{I}_n\otimes\mathbf{f}(0)]^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(t)
\end{aligned} \tag{5}$$

where $\otimes$ is the symbol of a Kronecker product, $\mathbf{I}_n$ denotes the $n\times n$ identity matrix,

$$\begin{aligned}
\mathbf{f}(t) &= [\sin\phi(t), \cos\phi(t)]^{\mathrm{T}}\\
\mathbf{h}(t) &= [\cos\phi(t), -\sin\phi(t)]^{\mathrm{T}}\\
\mathbf{f}(0) &= [0,1]^{\mathrm{T}},\ \mathbf{h}(0) = [1,0]^{\mathrm{T}}
\end{aligned}$$

and

$$\widehat{\mathbf{G}}(t) = \begin{bmatrix} \cos\widehat{\omega}(t) & \sin\widehat{\omega}(t)\\ -\sin\widehat{\omega}(t) & \cos\widehat{\omega}(t)\end{bmatrix}$$

is an orthogonal rotation matrix.

Tracking performance of this algorithm can be controlled by two user-dependent scalar coefficients: the small adaptation gain $\mu > 0$, which decides upon the speed of amplitude tracking, and another gain $\eta > 0$, which determines the speed of frequency tracking.

## 3. TRACKING ANALYSIS

Denote by $w(t)$ the one-step frequency change: $w(t) = \omega(t) - \omega(t-1)$. We will restrict tracking analysis to the case where

$$\boldsymbol{\beta}(t) = (\mathbf{I}_n\otimes\mathbf{G}(t))^{\mathrm{T}}\boldsymbol{\beta}(t-1), \forall t$$

$$\mathbf{G}(t) = \begin{bmatrix} \cos\omega(t) & \sin\omega(t)\\ -\sin\omega(t) & \cos\omega(t)\end{bmatrix}$$

In such a case the changes of $\boldsymbol{\theta}(t)$ can be attributed exclusively to the changes in $\omega(t)$, and the Euclidean norm of $\boldsymbol{\beta}(t)$ is constant: $||\boldsymbol{\beta}(t)||^2 = ||\boldsymbol{\beta}(0)||^2 = ||\boldsymbol{\beta}_0||^2$.

Denote by $\Delta\widehat{\omega}(t) = \widehat{\omega}(t) - \omega(t)$ and $\Delta\widehat{\boldsymbol{\beta}}(t) = \widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)$ the corresponding tracking errors and let

$$\begin{aligned}
e(t) &= \boldsymbol{\beta}_0^{\mathrm{T}}[\boldsymbol{\varphi}(t)\otimes\mathbf{h}(t)]v(t)\\
\Delta\widetilde{\boldsymbol{\beta}}(t) &= [\mathbf{I}_n\otimes\mathbf{B}(t)]\,\Delta\widehat{\boldsymbol{\beta}}(t)
\end{aligned}$$

$$\mathbf{B}(t) = \prod_{i=1}^{t}\mathbf{G}(i)\ , \quad \mathbf{J} = \begin{bmatrix} 0 & 1\\ -1 & 0\end{bmatrix}$$

By combining the approximating linear filtering (ALF) technique, proposed in (Tichavský & Händel, 1995) for analysis of adaptive notch filters, with the method of deterministic averaging, the following approximation formulas were derived in (Niedźwiecki & Sobociński, 2007)

$$\Delta\widehat{\omega}(t) \cong H_1(q^{-1})e(t) + H_2(q^{-1})w(t) \tag{6}$$

$$\begin{aligned}
\Delta\widetilde{\boldsymbol{\beta}}(t) \cong\ & \lambda\Delta\widetilde{\boldsymbol{\beta}}(t-1) - \lambda\left(\mathbf{I}_n\otimes\mathbf{J}\right)\boldsymbol{\beta}_o\Delta\widehat{\omega}(t)\\
& + \mu\left(\boldsymbol{\Phi}^{-1}\otimes\mathbf{I}_2\right)[\boldsymbol{\varphi}(t)\otimes\mathbf{f}(t)]v(t)
\end{aligned} \tag{7}$$

where $q^{-1}$ denotes the backward shift operator, the approximation filters have the form

$$H_1(q^{-1}) = \frac{2(1-\delta)(1-q^{-1})q^{-1}}{b^2(1-(\lambda+\delta)q^{-1}+\lambda q^{-2})}$$

$$H_2(q^{-1}) = -\frac{1-\lambda q^{-1}}{1-(\lambda+\delta)q^{-1}+\lambda q^{-2}}$$

and $\lambda = 1-\mu/2$, $\delta = 1-\eta b^2/2$, $b^2 = \boldsymbol{\beta}_o^{\mathrm{T}}(\boldsymbol{\Phi}\otimes\mathbf{I}_2)\boldsymbol{\beta}_o$. The filters $H_1(q^{-1})$ and $H_2(q^{-1})$ are asymptotically stable for any $\lambda$ and $\delta$ from the interval $(0,1)$.

After combining (6) with (7), and noting that $\mathbf{h}(t) = \mathbf{J}\mathbf{f}(t)$, one arrives at

$$\Delta\widetilde{\boldsymbol{\beta}}(t) \cong \mathbf{K}_{\mathrm{n}}\left(q^{-1}\right)\mathbf{n}(t) + \mathbf{K}_w\left(q^{-1}\right)w(t)$$

where $\mathbf{n}(t) = [\boldsymbol{\varphi}(t)\otimes\mathbf{f}(t)]v(t)$ and

$$\begin{aligned}
\mathbf{K}_{\mathrm{n}}\left(q^{-1}\right) =\ & \frac{\mu\left(\boldsymbol{\Phi}^{-1}\otimes\mathbf{I}_2\right)}{1-\lambda q^{-1}}\\
& - \frac{\lambda\left(\mathbf{I}_n\otimes\mathbf{J}\right)\boldsymbol{\beta}_o\boldsymbol{\beta}_o^{\mathrm{T}}\left(\mathbf{I}_n\otimes\mathbf{J}\right)H_1\left(q^{-1}\right)}{1-\lambda q^{-1}}
\end{aligned}$$

$$\mathbf{K}_w\left(q^{-1}\right) = -\frac{\lambda\left(\mathbf{I}_n\otimes\mathbf{J}\right)\boldsymbol{\beta}_o H_2\left(q^{-1}\right)}{1-\lambda q^{-1}}$$

### *3.1 Evaluation of the excess prediction error*

To arrive at analytical results we will assume that the instantaneous frequency $\omega(t)$ evolves according to the random walk model, i.e. that the frequency increments $w(t)$ form a white noise sequence with variance $\sigma_w^2$, independent of v(t). Denote by $\widehat{s}(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}(t)$ the estimate of the noiseless system output $s(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t)$. Define the excess prediction error $\varepsilon_s(t)$ as

$$\begin{aligned}
\varepsilon_s(t) &= \widehat{s}(t) - s(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\Delta\widehat{\boldsymbol{\theta}}(t)\\
&= \boldsymbol{\varphi}^{\mathrm{T}}(t)\left(\mathbf{I}_n\otimes\mathbf{f}^{\mathrm{T}}(0)\right)\Delta\widehat{\boldsymbol{\beta}}(t)
\end{aligned}$$

where $\Delta\widehat{\boldsymbol{\theta}}(t) = \widehat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t)$.

Since $\mathbf{B}(t)\mathbf{f}(0) = \mathbf{f}(t)$ and $\mathbf{B}^{-1}(t) = \mathbf{B}^{\mathrm{T}}(t)$, it holds that

$$\varepsilon_s(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\left(\mathbf{I}_n \otimes \mathbf{f}^{\mathrm{T}}(t)\right)\Delta\widetilde{\boldsymbol{\beta}}(t)$$

As it was shown in (Niedźwiecki & Sobociński, 2007), when the adaptation gains $\mu$ and $\gamma$ are small, the quantity $\Delta\widetilde{\boldsymbol{\beta}}(t)$ varies slowly compared to $\boldsymbol{\varphi}(t)$ and $\mathbf{f}(t)$. Denote by $\langle\cdot\rangle_T$ a local time average, where $T$ is the length of the analysis window, centered at $t$. Due to ergodicity of $\{\boldsymbol{\varphi}(t)\}$

$$\langle\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\rangle_T \cong \langle\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\rangle_\infty = \boldsymbol{\Phi}$$

Furthermore, for sufficiently slow frequency variations, $\sin\phi(t)$ and $\cos\phi(t)$ are locally almost periodic functions of time: $\sin\phi(t+i) \cong \sin[\omega_0\cdot(t+i)]$, $\cos\phi(t+i) \cong \cos[\omega_0\cdot(t+i)]$, $i \in [-T/2, T/2-1]$, where $\omega_0 = \phi(t)/t$. Therefore, when $T \gg 2\pi/\omega_0$, one arrives at

$$\langle\mathbf{f}(t)\mathbf{f}^{\mathrm{T}}(t)\rangle_T \cong \langle\mathbf{f}(t)\mathbf{f}^{\mathrm{T}}(t)|\omega(t)\equiv\omega_0\rangle_\infty = \frac{1}{2}\,\mathbf{I}_2$$

Combining both results one obtains

$$\langle\mathrm{E}\left[\varepsilon_s^2(t)\right]\rangle_T \cong \frac{1}{2}\,\mathrm{E}\left[\Delta\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}(t)\left(\boldsymbol{\Phi}\otimes\mathbf{I}_2\right)\Delta\widetilde{\boldsymbol{\beta}}(t)\right]$$

As shown in the Appendix, the time-averaged mean-squared excess prediction error for the algorithm (5), applied to the system (4), with frequency changing according to the random walk model, can be approximated by the following expression

$$\langle\mathrm{E}\left[\varepsilon_s^2(t)\right]\rangle_T \cong \left[\frac{n\mu}{2} + \frac{\gamma}{\mu}\right]\sigma_v^2 + \frac{b^2}{2\mu\gamma}\sigma_w^2 \qquad (8)$$

where where $\gamma = 1-\delta = \eta b^2/2$ and $n = \dim\boldsymbol{\theta}(t)$ denotes the number of estimated system coefficients.

Denote by $\mu_s$ and $\gamma_s$ the values of $\mu$ and $\gamma$ that minimize (8). It is straightforward to show that

$$\mu_s = 2\sqrt[4]{\frac{\xi}{n^2}}, \quad \gamma_s = \sqrt{\xi} \qquad (9)$$

where

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

### *3.2 Experimental verification*

We will check the approximation (8) using computer simulation. Consider the following two-tap FIR system:

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

where $u(t)$ denotes a white PRBS (pseudo-random binary) input sequence ($u(t) = \pm1, \sigma_u^2 = 1$) and $v(t)$ denotes Gaussian measurement noise. The impulse response coefficients of this system were modeled as nonstationary harmonic signals

$$\theta_1(t) = \sqrt{15}\sin\phi(t) + \sqrt{5}\cos\phi(t)$$
$$\theta_2(t) = \sqrt{5}\sin\phi(t) + \sqrt{15}\cos\phi(t)$$
$$\phi(t) = \sum_{\tau=1}^{t}\omega(\tau)$$

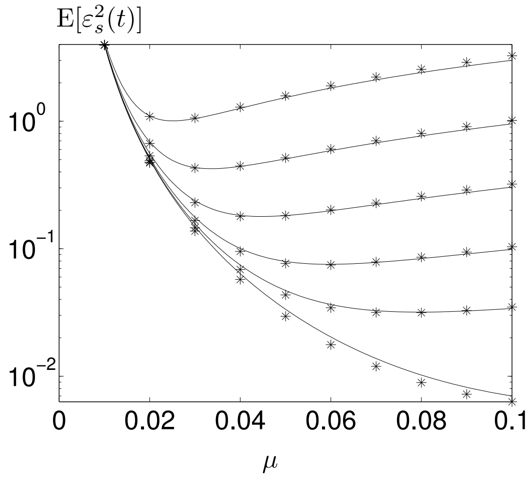


Fig. 1. Variance of the excess prediction error $\varepsilon_s(t)$ for a two-tap system with a single frequency mode, subject to a random walk drift. The theoretical results (solid lines) are compared with simulation results (500 runs for each point) obtained for different values of $\mu$; to reduce the number of degrees of freedom, the second adaptation gain $\gamma$ was set to $n\mu^2/4$. The corresponding signal-to-noise ratios were equal to (from top to bottom): 0 dB, 5 dB, 10 dB, 15 dB, 20 dB and 30 dB.

Note that in this case $\boldsymbol{\varphi}(t) = [u(t), u(t-1)]^{\mathrm{T}}$, $\boldsymbol{\Phi} = \mathbf{I}_2$ and $b^2 = 40$.

Fig. 1 shows comparison of theoretical evaluations, based on (8), with the results of computer simulations. The plot illustrates dependence of the average mean-squared system tracking error $\langle\mathrm{E}[\widehat{s}(t) - s(t))^2]\rangle_T$ on $\mu$ given $\gamma = n\mu^2/4$. All experimental points were obtained by means of double averaging: over time (2000 iterations, after the algorithm has reached its steady state behavior) and over different realizations of $\{u(t), v(t), w(t)\}$ (500 simulation runs). Note the good agreement between theoretical curves and the results of computer simulations.

## 4. SELF-OPTIMIZING GANF ALGORITHM

In order to guarantee satisfactory performance of the GANF algorithm, the values of adaptation gains should match the degree of system nonstationarity, so as to trade-off the filter's tracking speed (which increases with growing gains), and its tracking accuracy (which decreases with growing gains). A self-optimizing complex-valued GANF algorithm, capable of automatic tuning of its adaptation gains, was recently proposed in (Niedźwiecki & Kaczmarek, 2006b). Below we will extend this approach to the real-valued algorithm (5).

Similarly as it was done in (Niedźwiecki & Kaczmarek, 2006b), we will start from "pre-optimizing" the GANF algorithm. Assume, for the time being, that the scalar coefficient $b^2$ is constant and known (we will relax this assumption later on). When tuning the original GANF algorithm (5) one needs to simultaneously adjust two adaptation gains: $\mu$ and $\eta$ (or $\gamma$). According to (9), the optimal value of $\gamma$ is proportional to the square of the optimal value of $\mu$: $\gamma_s = n\mu_s^2/4$. Therefore, to make tuning easier, it may be worthwhile to reduce the number of design degrees of freedom from two ($\mu$, $\eta$) to one ($\mu$) by setting $\gamma = n\mu^2/4$, or equivalently $\eta = n\mu^2/(2b^2)$. In such a case the frequency update recursion in (5) should be replaced with

$$\widehat{\omega}(t+1) = \widehat{\omega}(t) + \kappa\mu^2 g(t) \tag{10}$$

where $\kappa = n/(2b^2)$. Simulation experiments confirm that – even though derived for a specific, random walk model of frequency changes – the update rule (10) can be very useful in practice: it works pretty satisfactorily under a wide range of more "realistic" frequency variation scenarios.

We will adjust the adaptation gain $\mu$ recursively, by minimizing the following local measure of fit, made up of exponentially weighted prediction errors

$$V(t,\mu) = \frac{1}{2}\sum_{\tau=1}^{t} \rho^{t-\tau}\varepsilon^2(\tau,\mu)$$

The forgetting constant $\rho$ ($0 < \rho < 1$) decides upon the effective averaging range.

To evaluate the estimate $\widehat{\mu}(t) = \arg\min_\mu V(t,\mu)$ we will use the standard RPE approach. According to (Söderström & Stoica, 1996), the RPE algorithm can be expressed in the form

$$\begin{aligned}\widehat{\mu}(t) &= \widehat{\mu}(t-1) \\ &\quad - \left[V''(t,\widehat{\mu}(t-1))\right]^{-1} V'(t,\widehat{\mu}(t-1))\end{aligned}$$

where

$$V'(t,\widehat{\mu}(t-1)) \cong \varepsilon(t,\widehat{\mu}(t-1))\frac{\partial\varepsilon(t,\widehat{\mu}(t-1))}{\partial\mu}$$

$$\begin{aligned}V''(t,\widehat{\mu}(t-1)) &\cong \rho V''(t-1,\widehat{\mu}(t-2)) \\ &\quad + \left[\frac{\partial\varepsilon(t,\widehat{\mu}(t-1))}{\partial\mu}\right]^2\end{aligned}$$

and all derivatives are taken with respect to $\mu$.

Let

$$\boldsymbol{\vartheta}(t) = \frac{\partial\mathbf{z}(t,\widehat{\mu}(t-1))}{\partial\mu}, \quad \zeta(t) = \frac{\partial\varepsilon(t,\widehat{\mu}(t-1))}{\partial\mu}$$

$$\boldsymbol{\psi}(t) = \frac{\partial\widehat{\boldsymbol{\beta}}(t,\widehat{\mu}(t-1))}{\partial\mu}, \quad \varrho(t) = \frac{\partial g(t,\widehat{\mu}(t-1))}{\partial\mu}$$

$$\chi(t) = \frac{\partial\widehat{\omega}(t,\widehat{\mu}(t-1))}{\partial\mu}, \quad r(t) = V''(t,\widehat{\mu}(t-1))$$

Straightforward calculations lead to

$$\begin{aligned}
\boldsymbol{\vartheta}(t) =& [\mathbf{I}_n \otimes (\mathbf{J}\widehat{\mathbf{G}}(t))]^{\mathrm{T}}\chi(t)\widehat{\boldsymbol{\beta}}(t-1) \\
&+ [\mathbf{I}_n \otimes \widehat{\mathbf{G}}(t)]^{\mathrm{T}}\boldsymbol{\psi}(t-1) \\
\zeta(t) =& -\ [\boldsymbol{\varphi}(t)\otimes\mathbf{f}(0)]^{\mathrm{T}}\boldsymbol{\vartheta}(t) \\
\boldsymbol{\psi}(t) =& \boldsymbol{\vartheta}(t) + [\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}(t)\otimes\mathbf{f}(0)] \\
&\times[\varepsilon(t) + \widehat{\mu}(t-1)\zeta(t)] \\
\varrho(t) =& [\boldsymbol{\varphi}(t)\otimes\mathbf{h}(0)]^{\mathrm{T}}[\boldsymbol{\vartheta}(t)\varepsilon(t) + \mathbf{z}(t)\zeta(t)] \\
r(t) =& \rho\, r(t-1) + \zeta^2(t) \\
\widehat{\mu}(t) =& \widehat{\mu}(t-1) - \frac{\varepsilon(t)\zeta(t)}{r(t)} \\
\chi(t+1) =& \chi(t) - \kappa\widehat{\mu}(t)[2g(t) + \widehat{\mu}(t)\varrho(t)]
\end{aligned} \tag{11}$$

When the true value of $b^2$ is unknown and possibly time-varying, one can replace it with the estimate

$$\widehat{b}^2(t) = \widehat{\boldsymbol{\beta}}^{\mathrm{T}}(t)(\boldsymbol{\Phi}\otimes\mathbf{I}_2)\widehat{\boldsymbol{\beta}}(t) \tag{12}$$

When the covariance matrix $\boldsymbol{\Phi}$ is also unknown, one can replace it in (5), (11) and (12) with the exponentially weighted estimate

$$\widehat{\boldsymbol{\Phi}}(t) = \lambda_o\widehat{\boldsymbol{\Phi}}(t-1) + (1-\lambda_o)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) \tag{13}$$

where $\lambda_o$, $0 < \lambda_o < 1$, denotes forgetting constant. Combining (5), (10), (11), (12) and (13) one finally arrives at the following self-optimizing version of the real-valued GANF algorithm

$$\begin{aligned}
\mathbf{z}(t) &= [\mathbf{I}_n\otimes\widehat{\mathbf{G}}(t)]^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(t-1) \\
\varepsilon(t) &= y(t) - [\boldsymbol{\varphi}(t)\otimes\mathbf{f}(0)]^{\mathrm{T}}\,\mathbf{z}(t) \\
\widehat{\boldsymbol{\Phi}}(t) &= \lambda_o\widehat{\boldsymbol{\Phi}}(t-1) + (1-\lambda_o)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) \\
\boldsymbol{\vartheta}(t) &= [\mathbf{I}_n\otimes(\mathbf{J}\widehat{\mathbf{G}}(t))]^{\mathrm{T}}\chi(t)\widehat{\boldsymbol{\beta}}(t-1) \\
&\quad + [\mathbf{I}_n\otimes\widehat{\mathbf{G}}(t)]^{\mathrm{T}}\boldsymbol{\psi}(t-1) \\
\zeta(t) &= -\,[\boldsymbol{\varphi}(t)\otimes\mathbf{f}(0)]^{\mathrm{T}}\boldsymbol{\vartheta}(t) \\
\boldsymbol{\psi}(t) &= \boldsymbol{\vartheta}(t) + [\widehat{\boldsymbol{\Phi}}^{-1}(t)\boldsymbol{\varphi}(t)\otimes\mathbf{f}(0)] \\
&\quad \times[\varepsilon(t) + \widehat{\mu}(t-1)\zeta(t)] \\
\varrho(t) &= [\boldsymbol{\varphi}(t)\otimes\mathbf{h}(0)]^{\mathrm{T}}[\boldsymbol{\vartheta}(t)\varepsilon(t) + \mathbf{z}(t)\zeta(t)] \\
r(t) &= \rho\, r(t-1) + \zeta^2(t) \\
\widehat{\mu}(t) &= \widehat{\mu}(t-1) - \frac{\varepsilon(t)\zeta(t)}{r(t)} \\
\widehat{\boldsymbol{\beta}}(t) &= \mathbf{z}(t) + \widehat{\mu}(t)\left[(\widehat{\boldsymbol{\Phi}}^{-1}(t)\boldsymbol{\varphi}(t))\otimes\mathbf{f}(0)\right]\varepsilon(t) \\
g(t) &= [\boldsymbol{\varphi}(t)\otimes\mathbf{h}(0)]^{\mathrm{T}}\,\mathbf{z}(t)\varepsilon(t) \\
\widehat{b}^2(t) &= \widehat{\boldsymbol{\beta}}^{\mathrm{T}}(t)[\widehat{\boldsymbol{\Phi}}(t)\otimes\mathbf{I}_2]\widehat{\boldsymbol{\beta}}(t) \\
\kappa(t) &= n/[2\widehat{b}^2(t)] \\
\widehat{\omega}(t+1) &= \widehat{\omega}(t) + \kappa(t)\widehat{\mu}^2(t)g(t) \\
\chi(t+1) &= \chi(t) - \kappa(t)\widehat{\mu}(t)[2g(t) + \widehat{\mu}(t)\varrho(t)] \\
\widehat{\boldsymbol{\theta}}(t) &= [\mathbf{I}_n\otimes\mathbf{f}(0)]^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(t)
\end{aligned} \tag{14}$$

Using the parallel estimation approach, described in (Niedźwiecki & Kaczmarek, 2006b), the algorithm (14) can be easily extended to the multiple frequencies case, i.e. to systems governed by (1) - (3).

The algorithm suitable for the signal case

$$\begin{aligned}
y(t) &= s(t) + v(t) \\
s(t) &= c_1(t)\sin\phi(t) + c_2(t)\cos\phi(t) \\
\phi(t) &= \sum_{\tau=1}^{t}\omega(\tau)
\end{aligned} \tag{15}$$

can be obtained from (14) using the following substitutions: $n = 1$, $\varphi(t) \equiv 1$, $\Phi(t) \equiv 1$ and

$\widehat{\beta}(t) = \widehat{s}(t)$. The resulting simplified recursions are

$$\begin{aligned}
\mathbf{z}(t) &= \widehat{\mathbf{G}}^{\mathrm{T}}(t)\widehat{\boldsymbol{\beta}}(t-1)\\
\varepsilon(t) &= y(t) - \mathbf{f}^{\mathrm{T}}(0)\mathbf{z}(t)\\
\boldsymbol{\vartheta}(t) &= \widehat{\mathbf{G}}^{\mathrm{T}}(t)[\mathbf{J}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(t-1)\ \chi(t) + \boldsymbol{\psi}(t-1)]\\
\zeta(t) &= -\ \mathbf{f}^{\mathrm{T}}(0)\boldsymbol{\vartheta}(t)\\
\boldsymbol{\psi}(t) &= \boldsymbol{\vartheta}(t) + \mathbf{f}(0)[\varepsilon(t) + \widehat{\mu}(t-1)\zeta(t)]\\
\varrho(t) &= \mathbf{h}^{\mathrm{T}}(0)[\boldsymbol{\vartheta}(t)\varepsilon(t) + \mathbf{z}(t)\zeta(t)]\\
r(t) &= \rho\, r(t-1) + \zeta^2(t)\\
\widehat{\mu}(t) &= \widehat{\mu}(t-1) - \frac{\varepsilon(t)\zeta(t)}{r(t)}\\
\widehat{\boldsymbol{\beta}}(t) &= \mathbf{z}(t) + \widehat{\mu}(t)\mathbf{f}(0)\varepsilon(t)\\
g(t) &= \mathbf{h}^{\mathrm{T}}(0)\mathbf{z}(t)\varepsilon(t)\\
\widehat{b}^2(t) &= ||\widehat{\boldsymbol{\beta}}(t)||^2\\
\kappa(t) &= 1/[2\widehat{b}^2(t)]\\
\widehat{\omega}(t+1) &= \widehat{\omega}(t) + \kappa(t)\widehat{\mu}^2(t)g(t)\\
\chi(t+1) &= \chi(t) - \kappa(t)\widehat{\mu}(t)[2g(t) + \widehat{\mu}(t)\varrho(t)]\\
\widehat{s}(t) &= \mathbf{f}^{\mathrm{T}}(0)\widehat{\boldsymbol{\beta}}(t)
\end{aligned} \qquad (16)$$

## 5. SIMULATION RESULTS

Our simulation experiment aimed at checking performance of the self-optimizing adaptive notch filter (16). The analyzed signal, governed by (15), was a noisy cisoid ($c_1(t) = c_2(t) \equiv 1$, $\sigma_v = 0.1$) with frequency drifting according to the random-walk model, starting from $\omega(0) = \pi/8$ – a typical frequency trajectory is shown in Fig. 2. The rate of frequency variation was time-varying:

$$\sigma_w(t) = \begin{cases} 0.0005 \text{ for } & t \in [1, 5000] \\ 0.001 \text{ for } & t \in [5001, 10000] \\ 0.0001 \text{ for } & t \in [10001, 15000] \end{cases}$$

The self-optimizing algorithm (16) was implemented with $\rho = 0.995$. Fig. 3 shows the ensemble average of $\widehat{\mu}(t)$, evaluated for 100 simulation runs, corresponding to different realizations of the processes $\{v(t)\}$ and $\{w(t)\}$. The obtained results confirm that the proposed gain scheduling rule works correctly.

## 6. CONCLUSION

We have shown that a real-valued generalized adaptive notch filter (GANF), used for identification/tracking of quasi-periodically varying systems, can be equipped with an auto-tuning loop, capable of performing on-line optimization of the algorithm's tracking performance. The proposed adaptation mechanism combines analytical results, derived for the real-valued GANF algorithm, with the recursive prediction error approach to optimization of adaptive filters.

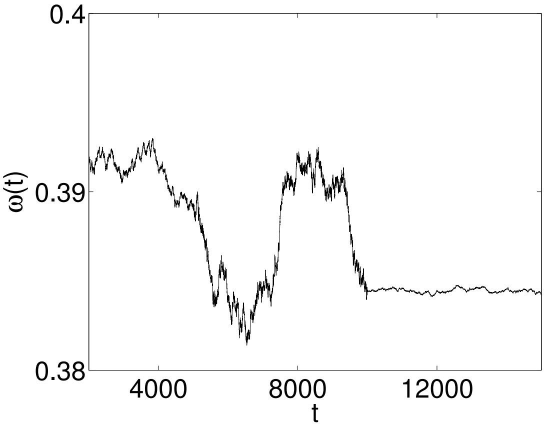


Fig. 2. Evolution of the instantaneous frequency of the analyzed signal during a typical simulation run.

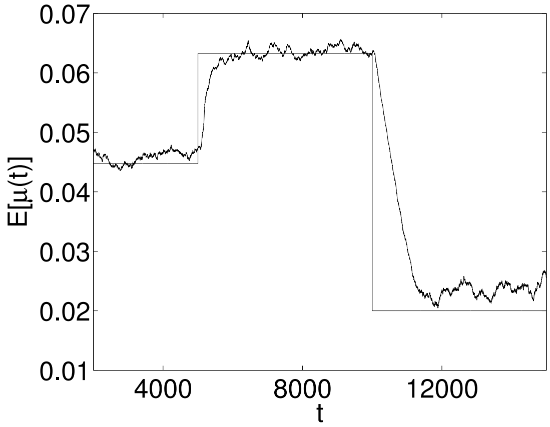


Fig. 3. Ensemble average (obtained for 100 simulation runs) of the gain estimate $\widehat{\mu}(t)$; thin line shows the optimal value of $\mu$.

## REFERENCES


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

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.

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

Niedźwiecki, M. and P Kaczmarek (2005b). Identification of quasi-periodically varying systems using the combined nonparametric/parametric approach. *IEEE Trans. on Signal Processing*, **vol. 53**, 4588-4588.

Niedźwiecki, M. and P Kaczmarek (2005c). Generalized adaptive notch and comb filters for identificatjion of quasi-periodically varying systems. *IEEE Trans. on Signal Processing*, **vol. 53**, 4599-4609.

Niedźwiecki, M. and P Kaczmarek (2006a). Tracking analysis of a generalized adaptive notch filter. *IEEE Trans. on Signal Processing*, **vol. 54**, 304-314.

Niedźwiecki, M. and P Kaczmarek (2006b). Generalized adaptive notch filter with a self-optimization capability. *IEEE Trans. on Signal Processing*, **vol. 54**, 4185-4193.

Niedźwiecki, M. and A. Sobociński (2007). On tracking properties of real-valued generalized adaptive notch filters. *IEEE Transactions on Signal Processing*, **vol. 55**, No. 5, 2007, 1688-1695.

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.

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

Söderström, T. and P. Stoica (1988). *System Identification.* Prentice Hall. Englewood Cliffs NJ.

## APPENDIX [derivation of (8)]

Note that

$$\langle \mathrm{E}\left[\varepsilon_s^2(t)\right]\rangle_T \cong \frac{1}{2}\,\mathrm{E}\left[\Delta\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}(t)\,(\boldsymbol{\Phi}\otimes\mathbf{I}_2)\,\Delta\widetilde{\boldsymbol{\beta}}(t)\right] = \frac{1}{2}\,\mathrm{tr}\{(\boldsymbol{\Phi}\otimes\mathbf{I}_2)\,\boldsymbol{\Sigma}\}$$

where $\boldsymbol{\Sigma} = \mathrm{E}\left[\Delta\widetilde{\boldsymbol{\beta}}(t)\Delta\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}(t)\right]$. Furthermore

$$\boldsymbol{\Sigma} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathbf{S}_{\Delta\widetilde{\boldsymbol{\beta}}}(\omega)d\omega$$

where $\mathbf{S}_{\Delta\widetilde{\boldsymbol{\beta}}}(\omega)$ denotes the spectral density matrix of the sequence $\{\Delta\widetilde{\boldsymbol{\beta}}(t)\}$.
Since $\mathbf{n}(t)$ and $w(t)$ are mutually uncorrelated, using standard linear filtering results one arrives at

$$\mathbf{S}_{\Delta\widetilde{\boldsymbol{\beta}}}(\omega) = \mathbf{K}_{\mathrm{n}}\left(e^{-j\omega}\right)\mathbf{D}_{\mathrm{n}}\mathbf{K}_{\mathrm{n}}^{\mathrm{T}}\left(e^{j\omega}\right) + \left|\mathbf{K}_w\left(e^{-j\omega}\right)\right|^2\sigma_w^2$$

where

$$\mathbf{D}_{\mathrm{n}} = \frac{1}{2}\,(\Phi\otimes\mathbf{I}_2)\,\sigma_v^2$$

Using the identity $\mathbf{J}^{\mathrm{T}} = -\mathbf{J}$ one arrives at

$$\begin{aligned}\mathbf{K}_{\mathrm{n}}\left(e^{-j\omega}\right)\mathbf{D}_{\mathrm{n}}\mathbf{K}_{\mathrm{n}}^{\mathrm{T}}\left(e^{j\omega}\right) = \\ = \frac{\sigma_v^2}{2|1-\lambda e^{-j\omega}|^2}\,\{\mu^2(\boldsymbol{\Phi}^{-1}\otimes\mathbf{I}_2) + \\ +\lambda\mu(\mathbf{I}_n\otimes\mathbf{J})\boldsymbol{\beta}_o[(\mathbf{I}_n\otimes\mathbf{J})\boldsymbol{\beta}_o]^{\mathsf{T}}H_1\left(e^{-j\omega}\right) \\ +\lambda\mu(\mathbf{I}_n\otimes\mathbf{J})\boldsymbol{\beta}_o[(\mathbf{I}_n\otimes\mathbf{J})\boldsymbol{\beta}_o]^{\mathsf{T}}H_1\left(e^{j\omega}\right) \\ +\lambda^2(\mathbf{I}_n\otimes\mathbf{J})\boldsymbol{\beta}_o[(\mathbf{I}_n\otimes\mathbf{J})\boldsymbol{\beta}_o]^{\mathsf{T}}(\Phi\otimes\mathbf{I}_2) \\ \times(\mathbf{I}_n\otimes\mathbf{J})\boldsymbol{\beta}_o[(\mathbf{I}_n\otimes\mathbf{J})\boldsymbol{\beta}_o]^{\mathsf{T}}|H_1\left(e^{-j\omega}\right)|^2\}\end{aligned}$$

$$\left|\mathbf{K}_w\left(e^{-j\omega}\right)\right|^2\sigma_w^2 = \frac{\lambda^2\sigma_w^2}{|1-\lambda e^{-j\omega}|^2} \times(\mathbf{I}_n\otimes\mathbf{J})\boldsymbol{\beta}_o[(\mathbf{I}_n\otimes\mathbf{J})\boldsymbol{\beta}_o]^{\mathsf{T}}\left|H_2\left(e^{-j\omega}\right)\right|^2$$

Note that

$$\begin{aligned}&\mathrm{tr}\{(\Phi\otimes\mathbf{I}_2)\,[(\mathbf{I}_n\otimes\mathbf{J})\,\boldsymbol{\beta}_o]\,[(\mathbf{I}_n\otimes\mathbf{J})\,\boldsymbol{\beta}_o]^{\mathsf{T}}\} \\ &= [(\mathbf{I}_n\otimes\mathbf{J})\,\boldsymbol{\beta}_o]^{\mathsf{T}}\,(\Phi\otimes\mathbf{I}_2)\,[(\mathbf{I}_n\otimes\mathbf{J})\,\boldsymbol{\beta}_o] \\ &= \boldsymbol{\beta}_o^{\mathsf{T}}\,(\Phi\otimes\mathbf{I}_2)\,\boldsymbol{\beta}_o = b^2\end{aligned}$$

Hence

$$\begin{aligned}\mathrm{tr}\{(\Phi\otimes\mathbf{I}_2)\,\mathbf{S}_{\Delta\widetilde{\boldsymbol{\beta}}}(\omega)\} &= \frac{\sigma_v^2}{2|1-\lambda e^{-j\omega}|^2} \\ &\times\{2n\mu^2 + \lambda\mu b^2H_1\left(e^{-j\omega}\right) + \lambda\mu b^2 H_1\left(e^{j\omega}\right) \\ &+ |\lambda b^2 H_1\left(e^{-j\omega}\right)|^2\} + \sigma_w^2\left|\frac{\lambda b\,H_2\left(e^{-j\omega}\right)}{1-\lambda e^{-j\omega}}\right|^2 \\ &= \frac{(2n-1)\sigma_v^2}{2}|F_R\left(e^{-j\omega}\right)|^2 + \frac{\sigma_v^2}{2}|G_{R1}\left(e^{-j\omega}\right)|^2 \\ &+ b^2\sigma_w^2|G_{R2}\left(e^{-j\omega}\right)|^2\end{aligned}$$

where

$$F_R\left(q^{-1}\right) = \frac{\mu}{1-\lambda q^{-1}}$$

$$G_{R1}\left(q^{-1}\right) = \frac{\mu+\lambda b^2 H_1\left(q^{-1}\right)}{1-\lambda q^{-1}}$$

$$G_{R2}\left(q^{-1}\right) = \frac{\lambda H_2\left(q^{-1}\right)}{1-\lambda q^{-1}}$$

Note that

$$\begin{aligned}&\mathrm{tr}\{(\Phi\otimes\mathbf{I}_2)\,\boldsymbol{\Sigma}\} \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{tr}\{(\Phi\otimes\mathbf{I}_2)\,\mathbf{S}_{\Delta\widetilde{\boldsymbol{\beta}}}(\omega)\}d\omega \\ &= \frac{(2n-1)\sigma_v^2}{2}\,I\left[F_R\left(z^{-1}\right)\right] \\ &+ \frac{\sigma_v^2}{2}\,I\left[G_{R1}\left(z^{-1}\right)\right] + b^2\sigma_w^2\,I\left[G_{R2}\left(z^{-1}\right)\right]\end{aligned}$$

where

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

is an integral evaluated along the unit circle in the $\mathcal{Z}$-plane, and $X(z^{-1})$ denotes any stable proper rational transfer function.
Using residue calculus one obtains

$$I\left[F_R\left(z^{-1}\right)\right] = \frac{4(1-\lambda)}{1+\lambda} \cong \mu$$

$$I\left[G_{R1}\left(z^{-1}\right)\right] = \frac{4(1+\delta-\lambda-3\lambda\delta+2\lambda^2)}{(1-\lambda)(1+2\lambda+\delta)} \cong \frac{4\gamma}{\mu} + \mu$$

$$I\left[G_{R2}\left(z^{-1}\right)\right] = \frac{\lambda^2(1+\lambda)}{(1-\lambda)(1-\delta)(1+2\lambda+\delta)} \cong \frac{1}{\mu\gamma}$$

which, after straightforward calculations, leads to (8).