



Gdańsk University of Technology

Technical Report #ETI 2/2010

Identification of quasi-periodically varying systems with quasi-linear frequency changes

by

Maciej Niedźwiecki and Michał Meller

Abstract

The problem of identification of linear quasi-periodically varying systems is considered. This problem can be solved using generalized adaptive notch filtering (GANF) algorithms. It is shown that accuracy of system parameter estimation can be increased if the results obtained from GANF are further processed using a cascade of appropriately designed filters. The resulting generalized adaptive notch smoothing (GANS) algorithms can be employed in off-line applications where causality constraints do not apply. When the instantaneous frequency of parameter changes varies in a sufficiently smooth manner, the proposed GANS algorithm, based on a new, quasi-linear model of frequency drift, outperforms the existing solutions.

I. INTRODUCTION

Consider the problem of identification of quasi-periodically varying complex-valued systems, i.e., systems governed by

$$y(t) = \boldsymbol{\varphi}^T(t)\boldsymbol{\theta}(t) + v(t) \quad (1)$$

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

$$\boldsymbol{\theta}(t) = \sum_{i=1}^k \boldsymbol{\beta}_i(t) e^{j \sum_{l=1}^t \omega_i(l)}. \quad (2)$$

All quantities in (1)–(2), except angular frequencies $\omega_1(t), \dots, \omega_k(t)$, are complex-valued. Since the complex “amplitudes” $\boldsymbol{\beta}_i(t) = [b_{i1}(t), \dots, b_{in}(t)]^T$ incorporate both magnitude and phase information, there is no explicit phase component in (2).

An interesting application, which under certain conditions admits the formulation presented above, is adaptive equalization of rapidly fading multipath telecommunication channels - see e.g. (Tsatsanis & Giannakis, 1996), (Giannakis & Tepedelenlioğlu, 1998), (Bakkoury et al., 2000). In this particular case, $y(t)$ is the sampled baseband signal, received by the mobile radio system, the regression vector $\boldsymbol{\varphi}(t)$ is made up of past input (transmitted) symbols, $\boldsymbol{\theta}(t)$ is the vector of time-varying impulse response coefficients of the channel, and the angular frequencies $\omega_1(t), \dots, \omega_k(t)$ 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).

Identification of the system (1)–(2) can be carried out using the algorithms known as generalized adaptive notch filters (GANFs) (Niedźwiecki & Kaczmarek, 2004). In the special case where $n = 1$ and $\varphi(t) \equiv 1$, equations (1)–(2) describe a mixture of complex-valued sinusoidal signals buried in noise [$s(t) = \theta(t)$]

$$y(t) = s(t) + v(t)$$

$$s(t) = \sum_{i=1}^k b_i(t) e^{j \sum_{l=1}^t \omega_i(l)} \quad (3)$$

and GANF filters become “ordinary” adaptive notch filters (ANFs) – devices used for a variety of purposes, such as line enhancement, mitigation of narrowband interferences in communication channels, biomedical signal processing, or elimination of narrowband disturbances (generated by power lines and/or electronic circuitry) from audio signals – see e.g. (Widrow & Stearns, 1985), (Regalia, 1995), (Haykin, 1996).

GANF filters are causal estimation algorithms, i.e., at each time instant they yield parameter estimates that are functions of the current and past measurements only. Even though the causality constraint is inevitable when tasks must be performed in real time, quite a number of applications exist where causality restrictions do not apply since the entire data record is available (incorporating both “past” and “future” observations). Reconstruction of parameter trajectories of a time-varying communication channel, based on prerecorded input/output sequences (e.g. for simulation purposes), is a good example of such a problem. Such tasks can be accomplished using generalized adaptive notch smoothing (GANS) algorithms.

The problem of generalized adaptive notch smoothing was studied in (Niedźwiecki & Sobociński, 2007), (Niedźwiecki & Sobociński, 2008) and (Niedźwiecki, 2010). For real-valued systems, simple versions of GANS algorithms, obtained by means of compensating estimation delays that arise in the frequency tracking and amplitude tracking loops of GANF algorithms, was proposed in (Niedźwiecki & Sobociński, 2008) [for complex-valued systems a simplified version of this solution was presented earlier in (Niedźwiecki & Sobociński, 2007)]. The approach described in (Niedźwiecki, 2010) is more sophisticated. Based on the analysis of the tracking properties of a complex-valued GANF algorithm, a cascade of postprocessing filters that increase accuracy of frequency and amplitude estimation was elaborated. It was shown that in the important benchmark case, where the instantaneous frequency of the single-mode system drifts according to the random-walk model, the proposed GANS algorithm is, under Gaussian assumptions, the statistically efficient frequency smoother, i.e., it reaches the corresponding Cramér-Rao-type lower smoothing bound. It is robust to frequency/amplitude model misspecification, and it yields better estimation results than the GANF algorithm, irrespective of the choice of adaptation gains.

In many practical applications the frequencies of system modes change in a linear or quasi-linear manner – see e.g. (Bakkoury et al., 2000). Whenever known in advance, this fact can be taken advantage of, as it allows one to design algorithms with improved tracking capability. In this paper we will propose such an improved generalized adaptive notch tracker, prove its statistical efficiency (under Gaussian assumptions), and show how it can be turned into a statistically efficient generalized adaptive notch smoother. The new scheme yields better results than those described in (Niedźwiecki & Sobociński, 2007)–(Niedźwiecki, 2010).

II. PROBLEM STATEMENT

To simplify further considerations, we will assume that the analyzed system has a single frequency mode ($k = 1$), i.e., that it is governed by

$$\begin{aligned} y(t) &= \boldsymbol{\varphi}^T(t)\boldsymbol{\theta}(t) + v(t) \\ \boldsymbol{\theta}(t) &= \boldsymbol{\beta}(t)e^{j\sum_{l=1}^t\omega(l)}. \end{aligned} \quad (4)$$

Based on the results presented in (Niedźwiecki & Sobociński, 2007), the algorithm derived below can be easily extended to systems with multiple frequency modes ($k > 1$).

We will assume that the complex-valued vector of “amplitudes” $\boldsymbol{\beta}(t) = [b_1(t), \dots, b_n(t)]^T$ and the real-valued instantaneous frequency $\omega(t) \in (-\pi, \pi]$ are slowly varying quantities. Additionally, we will assume that:

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

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

A. Quasi-linear Frequency Changes

Consider frequency variations governed by the following model

$$\begin{aligned} \omega(t) &= \omega(t-1) + \alpha(t-1) \\ \alpha(t) &= \alpha(t-1) + w(t) \end{aligned} \quad (5)$$

where $\alpha(t)$ denotes frequency rate and $w(t)$ – the one-step frequency rate change. According to (5) it holds that $(1 - q^{-1})^2 \omega(t) = q^{-1} w(t)$, where q^{-1} is the backward shift operator. Since $(1 - q^{-1})^2 \omega(t) = 0$ implies $\omega(t) = \omega_0 + \delta_\omega t$, (5) can be regarded as a perturbed linear growth/decay model. We will assume that

(A3) $\{w(t)\}$, independent of $\{v(t)\}$ and $\{\varphi(t)\}$, is a zero-mean white sequence with variance σ_w^2 .

Under this assumption, (5) defines the so-called second-order random-walk model. The corresponding frequency changes will be further referred to as *quasi-linear*.

B. Pilot GANF and Its Tracking Properties

Consider the following GANF algorithm, which combines frequency tracking with frequency rate tracking

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

where $*$ denotes complex conjugation, H denotes conjugate transpose, and $\mu > 0$, $\gamma_\omega > 0$, $\gamma_\alpha > 0$, $\gamma_\alpha \ll \gamma_\omega \ll \mu$, are small adaptation gains determining the rate of amplitude adaptation, frequency adaptation and frequency rate adaptation, respectively. We will show that this algorithm, further referred to as pilot GANF, can favorably cope with quasi-linear frequency changes. Note that for $\gamma_\alpha = 0$ and under zero initial conditions $\hat{\alpha}(0) = 0$, (6) reduces down to the algorithm studied in (Niedźwiecki, 2010) – the equivalence holds for $\gamma = \gamma_\omega$, where γ is the adaptation gain used in (Niedźwiecki, 2010).

Tracking properties of the pilot GANF will be analyzed in the case where the vector of “amplitudes” $\beta(t)$ is unknown but constant [in such a case, parameter changes of

the analyzed system can be attributed exclusively to the changes of the instantaneous frequency $\omega(t)$]:

$$(A4) \quad \boldsymbol{\beta}(t) \equiv \boldsymbol{\beta}_0, \text{ i.e., } \boldsymbol{\theta}(t) = e^{j\omega(t)}\boldsymbol{\theta}(t-1), \quad \forall t.$$

Using the approximating linear filter (ALF) technique – the stochastic linearization approach proposed in (Tichavský & P.Händel, 1995) – one can show that the frequency and frequency rate estimation errors $\Delta\hat{\omega}(t) = \omega(t) - \hat{\omega}(t)$, $\Delta\hat{\alpha}(t) = \alpha(t) - \hat{\alpha}(t)$, can be approximately expressed in the form (see Appendix 1)

$$\Delta\hat{\omega}(t) = H_1(q^{-1})e(t) + H_2(q^{-1})w(t) \quad (7)$$

$$\Delta\hat{\alpha}(t) = I_1(q^{-1})e(t) + I_2(q^{-1})w(t) \quad (8)$$

where $\{e(t)\}$, $e(t) = -\text{Im}[\boldsymbol{\beta}_0^H \boldsymbol{\varphi}^*(t) f^*(t) v(t) / b_0^2]$, $b_0^2 = \boldsymbol{\beta}_0^H \boldsymbol{\Phi} \boldsymbol{\beta}_0$, is a zero-mean white noise, independent of $\{w(t)\}$, with variance $\sigma_e^2 = \sigma_v^2 / (2b_0^2)$,

$$H_1(q^{-1}) = (1 - q^{-1})[\gamma_\omega + (\gamma_\alpha - \gamma_\omega)q^{-1}] / D(q^{-1})$$

$$H_2(q^{-1}) = q^{-1}[1 - \gamma_\omega - (1 - \mu)q^{-1}] / D(q^{-1})$$

$$I_1(q^{-1}) = \gamma_\alpha(1 - q^{-1})^2 / D(q^{-1})$$

$$I_2(q^{-1}) = [1 + (\mu + \gamma_\omega - 2)q^{-1} + (1 - \mu)q^{-2}] / D(q^{-1})$$

and

$$D(q^{-1}) = 1 + d_1q^{-1} + d_2q^{-2} + d_3q^{-3}$$

$$d_1 = \mu + \gamma_\omega + \gamma_\alpha - 3$$

$$d_2 = 3 - 2\mu - \gamma_\omega$$

$$d_3 = \mu - 1.$$

All filters are asymptotically stable if adaptation gains fulfill the following (sufficient) stability conditions: $0 < \mu < 1$, $0 < \gamma_\omega < 1$, $0 < \gamma_\alpha < 1$ and $\mu(\gamma_\omega + \gamma_\alpha) > \gamma_\alpha$.

It is worth noticing that ALF approximations remain valid for any uniformly small sequences $\{v(t)\}$ and $\{w(t)\}$, i.e., they are not restricted to sequences obeying assumptions (A1) and (A3). Additionally, the functional form of ALF equations does not change when the amplitude is also slowly varying with time, i.e., when assumption (A4) does not hold

true. These facts have important implications when it comes to robustness analysis of GANF/GANS algorithms.

We will show that under Gaussian assumptions:

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

the optimally tuned algorithm (6) is a statistically efficient frequency and frequency rate tracker, i.e., it reaches the corresponding lower tracking bounds. First, note that – due to orthogonality of $e(t)$ and $w(t)$ – the mean-squared tracking errors can be expressed in the form

$$\begin{aligned} \mathbb{E}\{[\Delta\hat{\omega}(t)]^2\} &= J[H_1(z^{-1})]\mathbb{E}[e^2(t)] + J[H_2(z^{-1})]\mathbb{E}[w^2(t)] \\ &= \sigma_w^2 F_\omega(\mu, \gamma_\omega, \gamma_\alpha; \kappa) \end{aligned} \quad (9)$$

$$\begin{aligned} \mathbb{E}\{[\Delta\hat{\alpha}(t)]^2\} &= J[I_1(z^{-1})]\mathbb{E}[e^2(t)] + J[I_2(z^{-1})]\mathbb{E}[w^2(t)] \\ &= \sigma_w^2 F_\alpha(\mu, \gamma_\omega, \gamma_\alpha; \kappa) \end{aligned} \quad (10)$$

where

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

is the integral evaluated along the unit circle [$X(z^{-1})$ is assumed to be a stable proper rational transfer function] and

$$\kappa = \frac{\mathbb{E}[w^2(t)]}{2\mathbb{E}[e^2(t)]} = \frac{b_0^2\sigma_w^2}{\sigma_v^2} = \text{SNR} \cdot \sigma_w^2 \quad (11)$$

denotes the rate of nonstationarity of the analyzed system (Tichavský & P.Händel, 1995) (SNR= b_0^2/σ_v^2 stands for the signal-to-noise ratio).

Optimal settings μ^{opt} , $\gamma_\omega^{\text{opt}}$ and $\gamma_\alpha^{\text{opt}}$ should be chosen so as to minimize (9) – to achieve the best frequency tracking, or to minimize (10) – to achieve the best frequency rate tracking. Note that in both cases the optimal values of μ , γ_ω and γ_α depend exclusively on the rate of signal nonstationarity κ . Even though the analytical expressions for $F_\omega(\cdot)$ and $F_\alpha(\cdot)$ can be easily derived using residue calculus [the corresponding formulas can be found e.g. in (Jury, 1964)], they are too lengthy and too complicated to enable optimization in a “symbolic” form. For this reason they are not presented here. For a given value of κ the optimal parameter values can be found using numerical search – the results are shown in

Table 1. Optimal GANF settings and the corresponding normalized lower tracking bounds.

κ	μ^{opt}	$\gamma_{\omega}^{\text{opt}}$	$\gamma_{\alpha}^{\text{opt}}$	$\text{LTB}_{\omega}/\sigma_w^2$	$\text{LTB}_{\alpha}/\sigma_w^2$
10^{-10}	0.0472	0.00113	0.0000138	$2.05 \cdot 10^5$	$8.21 \cdot 10^1$
$5 \cdot 10^{-10}$	0.0613	0.00192	0.0000306	$9.09 \cdot 10^4$	$6.28 \cdot 10^1$
10^{-9}	0.0685	0.00241	0.0000432	$6.39 \cdot 10^4$	$5.58 \cdot 10^1$
$5 \cdot 10^{-9}$	0.0886	0.00407	0.0000955	$2.82 \cdot 10^4$	$4.26 \cdot 10^1$
10^{-8}	0.0990	0.00509	0.000134	$1.97 \cdot 10^4$	$3.79 \cdot 10^1$
$5 \cdot 10^{-8}$	0.127	0.00852	0.000295	$8.66 \cdot 10^3$	$2.89 \cdot 10^1$
10^{-7}	0.142	0.0106	0.000414	$6.06 \cdot 10^3$	$2.57 \cdot 10^1$
$5 \cdot 10^{-7}$	0.181	0.0177	0.000905	$2.63 \cdot 10^3$	$1.95 \cdot 10^1$
10^{-6}	0.201	0.0219	0.00126	$1.83 \cdot 10^3$	$1.73 \cdot 10^1$
$5 \cdot 10^{-6}$	0.254	0.0359	0.00273	$7.81 \cdot 10^2$	$1.32 \cdot 10^1$
10^{-5}	0.281	0.0443	0.00379	$5.39 \cdot 10^2$	$1.17 \cdot 10^1$
$5 \cdot 10^{-5}$	0.350	0.0712	0.00806	$2.25 \cdot 10^2$	8.84
10^{-4}	0.384	0.0869	0.0111	$1.54 \cdot 10^2$	7.83

Table 1, along with the corresponding values of lower tracking bounds LTB_{ω} and LTB_{α} , derived in Appendix 2. Since it was found that the functions $F_{\omega}(\cdot)$ and $F_{\alpha}(\cdot)$ attain their minima for the same values of μ , γ_{ω} and γ_{α} , only one set of optimal gains was listed for each value of κ .

There is a perfect agreement between the lower tracking bounds and the values obtained by minimizing (9) and (10), respectively – in some cases the computed values agreed up to the six decimal place. (however, due to space constraints only two decimal places are shown). This means that, at least theoretically, the optimally tuned pilot GANF (6) should be a statistically efficient frequency and frequency rate tracker.

To check validity of theoretical results, the following two-tap FIR system (inspired by

channel equalization applications) was simulated

$$y(t) = \theta_1(t)u(t) + \theta_2(t)u(t-1) + v(t) \quad (12)$$

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-valued Gaussian measurement noise. The impulse response coefficients of this system were modeled as nonstationary cisoids $\theta_i(t) = b_i e^{j \sum_{\tau=1}^t \omega(\tau)}$, $i = 1, 2$, with time-invariant complex “amplitudes” $\beta = [b_1, b_2]^T = [2 - j, 1 + 2j]^T$. Note that in this case $\varphi(t) = [u(t), u(t-1)]^T$ and $\Phi = \mathbf{I}_2 \sigma_u^2$.

Fig. 1 shows comparison of theoretical values of the lower tracking bounds with experimental results obtained for the system (12) obeying assumptions (A1)–(A5), 3 different SNR values (0 dB, 10 dB and 20 dB) and 13 different values of the rate of nonstationarity κ , ranging from 10^{-10} to 10^{-4} . The mean-squared frequency estimation errors were evaluated (for the optimally tuned GANF algorithm) by means of joint time and ensemble averaging. First, for each realization of the measurement noise sequence and each realization of the frequency trajectory, the mean-squared errors were computed from 40000 iterations of the GANF filter (after the algorithm has reached its steady-state). The obtained results were next averaged over 20 realizations of $\{w(t), v(t)\}$. There are no results for SNR=0 dB and $\kappa > 10^{-6}$ since in the majority of runs the algorithm was unable to track under such extremely difficult conditions.

Note the very good agreement between the theoretical curves and the results of computer simulations.

C. From GANF to GANS

The pilot algorithm (6) and its approximate error model (7)–(8) will serve as a starting point for derivation of an adaptive notch smoother. Basically, we will follow the steps of (Niedźwiecki, 2010), i.e., we will design a cascade of postprocessing filters increasing accuracy of amplitude, frequency, and frequency rate estimation. We will show that, using such a multistage scheme, one can significantly improve estimation results. Moreover, for quasi-linear frequency changes and under Gaussian assumptions, the GANS algorithm obtained in this way is a statistically efficient frequency and frequency rate smoother, i.e., when optimally tuned it reaches the corresponding lower smoothing bounds.

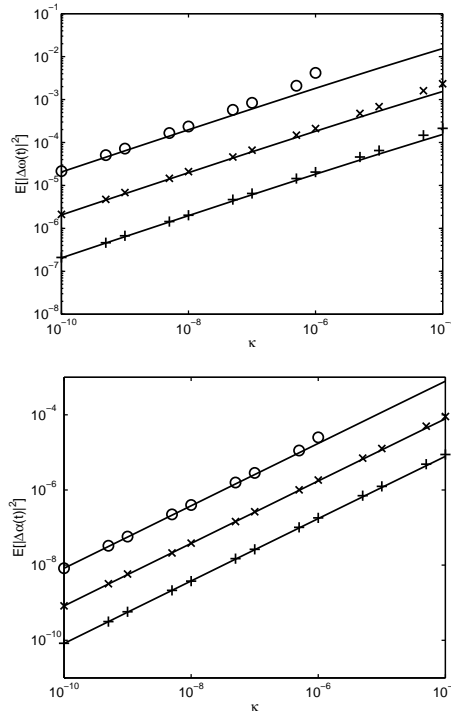


Fig. 1. Comparison of the theoretical values of the lower frequency (upper figure) and frequency rate (lower figure) tracking bounds (solid lines) with experimental results obtained for the signal with quasi-linear frequency changes for 3 different SNR values: SNR=0 dB (\circ), SNR=10 dB (\times), SNR=20 dB ($+$), and 13 different values of the rate of nonstationarity κ .

III. FREQUENCY RATE SMOOTHING

Smoothed frequency rate estimates can be obtained by means of lowpass filtering of the results yielded by the pilot GANF. Such an approach will be called postfiltering. We will start from considering a general postfiltering scheme, incorporating any linear noncausal filter. Then we will show that the best results can be obtained when the smoothing filter is anticausal and “matched” to the frequency characteristics of optimally tuned GANF. Finally, we will explain why the proposed scheme should work satisfactorily for any slow frequency rate variations, not necessarily of the random-walk type, and for any adaptation gains, not necessarily optimally tuned.

A. Optimization

Suppose that the entire measurement record is available up to the instant N : $\Omega(N) = \{\mathcal{Y}(N), \Phi(N)\}$ where $\mathcal{Y}(N) = \{y(1), \dots, y(N)\}$, $\Phi(N) = \{\varphi(1), \dots, \varphi(N)\}$. To obtain a

fixed-interval smoothed estimate of $\alpha(t)$, further denoted by $\tilde{\alpha}(t)$, we will pass the estimates $\hat{\alpha}(t)$ through a noncausal filter $R(q^{-1}) = \dots + r_{-1}q^{-1} + r_0 + r_1q^1 + \dots$

$$\tilde{\alpha}(t) = R(q^{-1})\hat{\alpha}(t). \quad (13)$$

The filter $R(q^{-1})$ will be designed so as to minimize the mean-squared frequency estimation error $E\{[\Delta\tilde{\alpha}(t)]^2\}$, where $\Delta\tilde{\alpha}(t) = \alpha(t) - \tilde{\alpha}(t)$. After elementary but tedious calculations, one can show that

$$\Delta\tilde{\alpha}(t) = (1 - q^{-1})^2 X(q^{-1})e(t+1) + \frac{1 - X(q^{-1})}{1 - q^{-1}} w(t) \quad (14)$$

where

$$\begin{aligned} X(q^{-1}) &= R(q^{-1})S(q^{-1}) \\ S(q^{-1}) &= 1 - (1 - q^{-1})I_2(q^{-1}) = \gamma_\alpha q^{-1}/D(q^{-1}). \end{aligned} \quad (15)$$

Let $\Delta(q^{-1}) = 1/(1 - q^{-1})$. Due to orthogonality of $\{e(t)\}$ and $\{w(t)\}$, one obtains

$$E\{[\Delta\tilde{\alpha}(t)]^2\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} h[X(e^{-j\xi})] d\xi \quad (16)$$

where

$$h[X] = \frac{XX^*}{|\Delta|^4} \sigma_e^2 + |\Delta|^2(1 - X)(1 - X^*) \sigma_w^2. \quad (17)$$

Minimization of (17) can be achieved by minimizing $h[X(e^{-j\xi})]$ for every value of $\xi \in (-\pi, \pi]$.

Denote by

$$\begin{aligned} \frac{\partial}{\partial z} &= \frac{1}{2} \left[\frac{\partial}{\partial \text{Re}[z]} - j \frac{\partial}{\partial \text{Im}[z]} \right] \\ \frac{\partial}{\partial z^*} &= \frac{1}{2} \left[\frac{\partial}{\partial \text{Re}[z]} + j \frac{\partial}{\partial \text{Im}[z]} \right] \end{aligned}$$

the so-called Wirtinger derivatives – symbolic derivatives with respect to a complex variable z , applicable to non-analytic functions, such as $h(\cdot)$ (Haykin, 1996).

To find the optimal transfer function $X_0(q^{-1})$, i.e., the one that minimizes (16), we will request that

$$\left. \frac{\partial h}{\partial X^*} \right|_{X=X_0} = \frac{X_0}{|\Delta|^4} \sigma_e^2 - |\Delta|^2(1 - X_0) \sigma_w^2 = 0. \quad (18)$$

Solving (18), one obtains

$$X_0(q^{-1}) = \frac{2\kappa}{2\kappa + (1 - q^{-1})^3(1 - q)^3}. \quad (19)$$

One of the key observations of this paper is that the transfer function $X_0(q^{-1})$ can be factorized as follows

$$X_0(q^{-1}) \cong S_0(q^{-1})S_0(q) \quad (20)$$

where $S_0(q^{-1}) = S(q^{-1})|_{(\mu, \gamma_\omega, \gamma_\alpha)^{\text{opt}}}$ and $\mu^{\text{opt}}, \gamma_\omega^{\text{opt}}, \gamma_\alpha^{\text{opt}}$ denote optimal settings for the adaptive notch tracker (6).

The relationship (20) can be conjectured from an analogous result proved analytically (with equality sign instead of approximate equality) in (Niedźwiecki, 2010) for a simpler case of a random-walk frequency drift [see equation (15) in (Niedźwiecki, 2010)]. Since the optimal gain settings for quasi-linear frequency changes were established numerically, the same approach must be applied to verification of (20). Our check was based on comparison of frequency characteristics $X_0(e^{-j\xi})$ and $S_0(e^{-j\xi})S_0(e^{j\xi})$. Since an ideal match was observed for all values of $\kappa \in [10^{-10}, 10^{-1}]$, and all values of $\xi \in (-\pi, \pi]$, we have the right to claim that the possible factorization errors, if any, are negligible.

Combining (15) with (20), one arrives at the following transfer function of the optimal postfiltering scheme

$$R_0(q^{-1}) = \frac{X_0(q^{-1})}{S_0(q^{-1})} \cong S_0(q). \quad (21)$$

Since the filter $S_0(q)$ is anticausal, postfiltering can be realized by means of backward-time filtering of the frequency rate estimates $\hat{\alpha}(t)$ yielded by the pilot algorithm. Later on, in Section 7, we will show that under assumptions (A1)–(A5) the optimized scheme based on postfiltering is statistically efficient, i.e., it reaches the lower frequency rate lower smoothing bound that can be derived for quasi-linear frequency changes.

B. Robust Smoothing Scheme

Although interesting from the theoretical viewpoint, the facts established so far are of little practical value – the optimality results are restricted to a specific model of frequency rate changes and they were derived under assumption that the optimal settings for the pilot filter are known. We will show that when postfiltering has the form

$$\tilde{\alpha}(t) = S(q)\hat{\alpha}(t) \quad (22)$$

improvement in estimation accuracy can be expected for *any* smooth frequency rate trajectory and for *any* choice of adaptation gains. Our robustness analysis will be based on the following relationship

$$\hat{\alpha}(t) = S(q^{-1})\alpha(t) - I_1(q^{-1})e(t) \quad (23)$$

which is an equivalent of (8). Importantly, as already stressed in Section 2.2, ALF approximations (7) and (8) [and hence also the relationship (23)] remain valid for any sequence of small one-step frequency rate changes $\{w(t)\}$, i.e., for any trajectory $\{\alpha(t)\}$ that is sufficiently smooth and hence can be regarded as a lowpass signal.

Note that for zero-mean measurement noise it holds that $E[e(t)] = 0$, leading to

$$E[\hat{\alpha}(t)|\alpha(s), s \leq t] \cong S(q^{-1})\alpha(t) . \quad (24)$$

Note also that $S(q^{-1})$ is a lowpass filter with unity static gain $S(1) = 1$. Hence, when the instantaneous frequency rate varies slowly with time, the mean path of frequency rate estimates is approximately the time-delayed version of the true trajectory

$$E[\hat{\alpha}(t)|\alpha(s), s \leq t] \cong \alpha(t - \tau_\alpha) \quad (25)$$

where $\tau_\alpha = \text{int}[t_\alpha]$ denotes the so-called estimation delay, equal to the integer part of the nominal (low-frequency) delay of the filter $S(q^{-1})$. The nominal delay of $S(q^{-1})$ is defined as $t_\alpha = -\lim_{\xi \rightarrow 0} d\phi(\xi)/d\xi = -\lim_{\xi \rightarrow 0} \phi(\xi)/\xi$ where $\phi(\xi) = \arg[S(e^{-i\xi})]$ is the phase characteristic of $S(q^{-1})$. One can show that in the case considered

$$t_\alpha = I_2(1) = \frac{\gamma_\omega}{\gamma_\alpha} . \quad (26)$$

Estimation delay is a source of the bias error (sometimes called lag error) which, especially for large values of τ_α , may seriously degrade tracking performance of the pilot GANF.

For the smoothed estimate (22) the situation is different. Combining (22) with (23), one arrives at

$$\tilde{\alpha}(t) = S(q)S(q^{-1})\alpha(t) - S(q)I_1(q^{-1})e(t) . \quad (27)$$

Since the nominal delay of the filter $S(q)S(q^{-1})$ is equal to zero, and its static gain is equal to 1, the steady-state ($1 \ll t \ll N$) smoothed estimate is approximately unbiased

$$E[\tilde{\alpha}(t)|\alpha(s), 1 \leq s \leq N] \cong \alpha(t) \quad (28)$$

which considerably increases estimation accuracy. Importantly, this conclusion holds true irrespective of the shape of the estimated frequency rate trajectory (as long as it is sufficiently smooth), and for any choice of adaptation gains μ , γ_ω and γ_α that guarantee stable operation of the pilot GANF.

IV. FREQUENCY SMOOTHING

Frequency smoothing can be achieved in an analogous way as frequency rate smoothing, i.e., by means of postfiltering of the estimated frequency trajectory yielded by the pilot GANF. Similar to frequency rate smoothing, the proposed approach to frequency smoothing is robust to modeling errors.

A. Optimization

Denote by $\tilde{\omega}(t)$ the smoothed frequency estimate

$$\tilde{\omega}(t) = P(q^{-1})\hat{\omega}(t) \quad (29)$$

where $P(q^{-1}) = \dots + p_{-1}q^{-1} + p_0 + p_1q^1 + \dots$ is any noncausal filter. We will proceed similarly as in Section 3.1, i.e., we will find the filter $P(q^{-1})$ that minimizes the mean-squared frequency estimation error $E\{[\Delta\tilde{\omega}(t)]^2\}$, where $\Delta\tilde{\omega}(t) = \omega(t) - \tilde{\omega}(t)$. One can show that

$$\Delta\tilde{\omega}(t) = (1 - q^{-1})Y(q^{-1})e(t) + \frac{1 - Y(q^{-1})}{(1 - q^{-1})^2} w(t - 1) \quad (30)$$

where

$$\begin{aligned} Y(q^{-1}) &= P(q^{-1})Q(q^{-1}) \\ Q(q^{-1}) &= 1 - q(1 - q^{-1})^2 H_2(q^{-1}) = [\gamma_\omega + (\gamma_\alpha - \gamma_\omega)q^{-1}]/D(q^{-1}). \end{aligned} \quad (31)$$

Due to orthogonality of $\{e(t)\}$ and $\{w(t)\}$, (30) leads to

$$E\{[\Delta\tilde{\omega}(t)]^2\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} g[Y(e^{-j\xi})] d\xi \quad (32)$$

where

$$g[Y] = \frac{YY^*}{|\Delta|^2} \sigma_e^2 + |\Delta|^4(1 - Y)(1 - Y^*) \sigma_w^2. \quad (33)$$

The optimal transfer function $Y_0(q^{-1})$ can be obtained by solving

$$\left. \frac{\partial g}{\partial Y^*} \right|_{Y=Y_0} = \frac{Y_0}{|\Delta|^2} \sigma_e^2 - |\Delta|^4 (1 - Y_0) \sigma_w^2 = 0 \quad (34)$$

which leads to

$$Y_0(q^{-1}) = X_0(q^{-1}). \quad (35)$$

Finally, after combining (35) with (31) and (20), one arrives at

$$P_0(q^{-1}) \cong \frac{S_0(q^{-1})S_0(q)}{Q_0(q^{-1})} \quad (36)$$

where $Q_0(q^{-1}) = Q(q^{-1})|_{(\mu, \gamma_\omega, \gamma_\alpha)^{\text{opt}}}$. In Section 7 we will show that under assumptions (A1)–(A5) the resulting optimized smoothing scheme attains the lower frequency smoothing bound, which means that for quasi-linear frequency changes it is statistically efficient.

B. Robust Smoothing Scheme

Based on (36) we propose the following postfiltering scheme

$$\tilde{\omega}(t) = S(q)T(q^{-1})\hat{\omega}(t) \quad (37)$$

where

$$T(q^{-1}) = \frac{S(q^{-1})}{Q(q^{-1})} = \frac{\gamma_\alpha q^{-1}}{\gamma_\omega + (\gamma_\alpha - \gamma_\omega)q^{-1}} = \frac{b_1 q^{-1}}{1 + c_1 q^{-1}}$$

$$b_1 = \frac{\gamma_\alpha}{\gamma_\omega}, \quad c_1 = \frac{\gamma_\alpha - \gamma_\omega}{\gamma_\omega}.$$

Note that the filter $T(q^{-1})$ is causal and the filter $S(q)$ is anticausal. Therefore, postfiltering can be realized by means of forward-time filtering of the frequency trajectory $\{\hat{\omega}(t)\}$ using the filter $T(q^{-1})$, followed by backward-time filtering of the results using the filter $S(q)$.

Robustness analysis of (37) is similar to that carried for (22). Using (7), one can show that

$$\hat{\omega}(t) = Q(q^{-1})\omega(t) - H_1(q^{-1})e(t) \quad (38)$$

leading to (in steady-state)

$$\begin{aligned} \mathbb{E}[\tilde{\omega}(t)|\omega(s), 1 \leq s \leq N] &= S(q)T(q^{-1})Q(q^{-1})\omega(t) \\ &= S(q)S(q^{-1})\omega(t) \cong \omega(t) \end{aligned} \quad (39)$$

for any frequency trajectory that can be modeled as a lowpass signal and for any selection of stabilizing adaptation gains.

Suppose that postfiltering is restricted to the forward pass

$$\bar{\omega}(t) = T(q^{-1})\hat{\omega}(t).$$

Then it holds that (in steady-state)

$$\begin{aligned} \mathbb{E}[\bar{\omega}(t)|\omega(s), s \leq t] &= T(q^{-1})Q(q^{-1})\omega(t) \\ &= S(q^{-1})\omega(t) \cong \omega(t - \tau_\omega) \end{aligned}$$

where $\tau_\omega = \tau_\alpha$ is the delay introduced by the filter $S(q^{-1})$ – identical with that established in Section 3 for frequency rate tracking.

V. AMPLITUDE SMOOTHING

Having smoothed frequency estimates, one can re-estimate the amplitude coefficients. This can be achieved using the frequency-guided version of the algorithm (6), obtained by replacing the filtering-type (causal) frequency estimates $\hat{\omega}(t)$ with their smoothed (non-causal) counterparts $\tilde{\omega}(t)$, derived in Section 4

$$\begin{aligned} \tilde{f}(t) &= e^{j\tilde{\omega}(t)}\tilde{f}(t-1) \\ \bar{\varepsilon}(t) &= y(t) - \boldsymbol{\varphi}^T(t)\tilde{f}(t)\bar{\boldsymbol{\beta}}(t-1) \\ \bar{\boldsymbol{\beta}}(t) &= \bar{\boldsymbol{\beta}}(t-1) + \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\tilde{f}^*(t)\bar{\varepsilon}(t) \\ \bar{\boldsymbol{\theta}}(t) &= \bar{\boldsymbol{\beta}}(t)\tilde{f}(t). \end{aligned} \tag{40}$$

Since the frequency-guided GANF (40) is identical with the analogous algorithm studied in (Niedźwiecki, 2010), we will simply repeat conclusions reached there: the smoothed estimates of amplitude coefficients can be obtained by means of backward-time filtering of the estimates yielded by the algorithm (40)

$$\tilde{\boldsymbol{\beta}}(t) = F(q)\bar{\boldsymbol{\beta}}(t) \tag{41}$$

where

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

is the unity-gain anticausal lowpass filter “matched” to tracking characteristics of (40). When the amplitude trajectory $\{\boldsymbol{\beta}(t)\}$ can be modeled as a lowpass process and $1 \ll t \ll N$ (steady-state conditions), it holds that

$$\mathbb{E}[\tilde{\boldsymbol{\beta}}(t)|\boldsymbol{\beta}(s), 1 \leq s \leq N] = F(q)F(q^{-1})\boldsymbol{\beta}(t) \cong a(t) \quad (42)$$

which means that the smoothed amplitude estimates are approximately unbiased - see (Niedźwiecki, 2010) for more details.

When smoothing is not incorporated, the corresponding relationship reads

$$\mathbb{E}[\bar{\boldsymbol{\beta}}(t)|\boldsymbol{\beta}(s), s \leq t] = F(q^{-1})\boldsymbol{\beta}(t) \cong \boldsymbol{\beta}(t - \tau_\beta) \quad (43)$$

where $\tau_\beta = \text{int}[t_\beta]$ denotes estimation delay arising in the amplitude tracking loop of the frequency-guided GANF, equal to the integer part of the nominal delay of the filter $F(q^{-1})$: $t_\beta = (1 - \mu)/\mu$.

VI. GENERALIZED ADAPTIVE NOTCH SMOOTHER

The fixed-interval GANS algorithm can be obtained by combining all steps described in Sections 2-5 and adding the final parameter reconstruction step

$$\tilde{\boldsymbol{\theta}}(t) = \tilde{\boldsymbol{\beta}}(t)\tilde{f}(t). \quad (44)$$

The proposed adaptive smoothing algorithm is a six-step procedure, combining results yielded by the: pilot GANF algorithm (6), frequency rate smoother (22) [optionally], frequency smoother (37), frequency-guided GANF algorithm (40), amplitude smoother (41) and output filter (44) - all steps are summarized in Table 2.

When the covariance matrix $\boldsymbol{\Phi}$ is not known and/or it is time-varying, one can replace it with the following exponentially weighted estimate

$$\hat{\boldsymbol{\Phi}}(t) = \lambda_o \hat{\boldsymbol{\Phi}}(t-1) + (1 - \lambda_o) \boldsymbol{\varphi}^*(t) \boldsymbol{\varphi}^T(t)$$

where $0 < \lambda_o < 1$ denotes the forgetting constant. We note that the inverse of $\hat{\boldsymbol{\Phi}}(t)$ can also be evaluated recursively by exploiting the well-known matrix inversion lemma (Haykin, 1996).

Table 2. Generalized adaptive notch smoother

pilot filter :

$$\begin{aligned}
\hat{f}(t) &= e^{j[\hat{\omega}(t-1)+\hat{\alpha}(t-1)]}\hat{f}(t-1) \\
\varepsilon(t) &= y(t) - \boldsymbol{\varphi}^T(t)\hat{f}(t)\hat{\boldsymbol{\beta}}(t-1) \\
\hat{\boldsymbol{\beta}}(t) &= \hat{\boldsymbol{\beta}}(t-1) + \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\hat{f}^*(t)\varepsilon(t) \\
\delta(t) &= \text{Im} \left[\frac{\varepsilon^*(t)\boldsymbol{\varphi}^T(t)\hat{f}(t)\hat{\boldsymbol{\beta}}(t-1)}{\hat{\boldsymbol{\beta}}^H(t-1)\boldsymbol{\Phi}\hat{\boldsymbol{\beta}}(t-1)} \right] \\
\hat{\alpha}(t) &= \hat{\alpha}(t-1) - \gamma_\alpha\delta(t) \\
\hat{\omega}(t) &= \hat{\omega}(t-1) + \hat{\alpha}(t-1) - \gamma_\omega\delta(t) \\
t &= 1, \dots, N
\end{aligned}$$

frequency rate smoother [optional] :

$$\begin{aligned}
\tilde{\alpha}(N-i) &= \hat{\alpha}(N-i), \quad i = 0, 1, 2 \\
\tilde{\alpha}(t) &= -d_1\tilde{\alpha}(t+1) - d_2\tilde{\alpha}(t+2) \\
&\quad -d_3\tilde{\alpha}(t+3) + \gamma_\alpha\hat{\alpha}(t+1) \\
t &= N-3, \dots, 1
\end{aligned}$$

frequency smoother :

$$\begin{aligned}
\bar{\omega}(1) &= \hat{\omega}(1) \\
\bar{\omega}(t) &= -c_1\bar{\omega}(t-1) + b_1\hat{\omega}(t-1) \\
t &= 1, \dots, N
\end{aligned}$$

$$\begin{aligned}
\tilde{\omega}(N-i) &= \bar{\omega}(N-i), \quad i = 0, 1, 2 \\
\tilde{\omega}(t) &= -d_1\tilde{\omega}(t+1) - d_2\tilde{\omega}(t+2) \\
&\quad -d_3\tilde{\omega}(t+3) + \gamma_\alpha\bar{\omega}(t+1) \\
t &= N-3, \dots, 1
\end{aligned}$$

frequency – guided filter :

$$\begin{aligned}
\tilde{f}(t) &= e^{j\tilde{\omega}(t)}\tilde{f}(t-1) \\
\bar{\varepsilon}(t) &= y(t) - \boldsymbol{\varphi}^T(t)\tilde{f}(t)\bar{\boldsymbol{\beta}}(t-1) \\
\bar{\boldsymbol{\beta}}(t) &= \bar{\boldsymbol{\beta}}(t-1) + \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\tilde{f}^*(t)\bar{\varepsilon}(t) \\
t &= 1, \dots, N
\end{aligned}$$

amplitude smoother :

$$\begin{aligned}
\tilde{\boldsymbol{\beta}}(N) &= \bar{\boldsymbol{\beta}}(N) \\
\tilde{\boldsymbol{\beta}}(t) &= (1-\mu)\tilde{\boldsymbol{\beta}}(t+1) + \mu\bar{\boldsymbol{\beta}}(t) \\
t &= N-1, \dots, 1
\end{aligned}$$

output filter :

$$\begin{aligned}
\tilde{\theta}(t) &= \tilde{\boldsymbol{\beta}}(t)\tilde{f}(t) \\
t &= 1, \dots, N
\end{aligned}$$

Unless a specific prior knowledge about the analyzed signal is available, we recommend to set $\hat{f}(0) = 1$, $\hat{\alpha}(0) = 0$, and to adopt for $\hat{\omega}(0)$ the value that stems from spectral analysis of a short initial fragment of the available data record – a suitable algorithm, based on the so-called generalized (system) periodogram, can be found in (Niedźwiecki & Kaczmarek, 2005). Less careful choice of initial conditions may result in long initialization transients (especially for small adaptation gains) but does not prevent the pilot algorithm from reaching the small-error tracking mode after a sufficiently long initial convergence period.

VII. SIMULATION RESULTS

Our simulation study will focus on two aspects of the proposed generalized adaptive notch smoothing scheme: optimality and robustness.

Although demonstration of the algorithm's optimality, i.e., its ability to reach the Cramér-Rao-type lower smoothing bounds is mainly of theoretical value, it is important as it allows one to specify conditions under which some absolute performance limits can be reached. From the practical viewpoint, the most important property of the estimation algorithm is its robustness, i.e., insensitivity to modeling errors. We will show that, exactly as predicted, the proposed GANS algorithm outperforms its GANF counterpart for a wide range of operating conditions, including different (nonstandard) frequency/amplitude trajectories and different (non-optimal) values of adaptation gains.

A. Optimality

Fig. 2 shows comparison of theoretical values of the lower smoothing bounds (see Appendix 2) with experimental results obtained for the system (12) obeying assumptions (A1)–(A5), 3 different SNR values (0 dB, 10 dB and 20 dB) and 13 different values of the rate of nonstationarity κ , ranging from 10^{-10} to 10^{-4} . Similarly as in Section 2.2, all MSE values were computed for the optimally tuned GANS algorithm by means of joint time and ensemble averaging. Note the good agreement between the theoretical curves and the results of computer simulations.

The possible margins of improvement LTB_ω/LSB_ω and LTB_α/LSB_α , which are functions of κ , are depicted in Fig. 3.

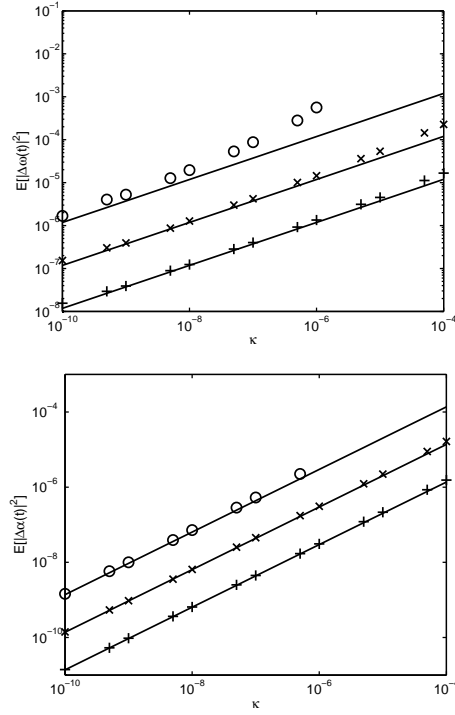


Fig. 2. Comparison of the theoretical values of the lower frequency (upper figure) and frequency rate (lower figure) smoothing bounds (solid lines) with experimental results obtained for the signal with quasi-linear frequency changes, for 3 different SNR values: SNR=0 dB (\circ), SNR=10 dB (\times), SNR=20 dB ($+$), and 13 different values of the rate of nonstationarity κ .

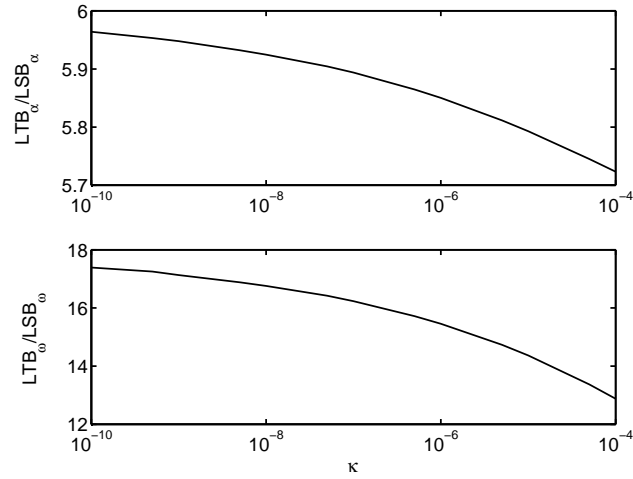


Fig. 3. Dependence of the limiting tracking versus smoothing variance ratios on the degree of signal nonstationarity κ .

B. Robustness

To check performance of the GANS algorithm under nonstandard conditions, i.e., for systems with parameters that do not change according to the second-order random-walk model, a system (12) with fast sinusoidal amplitude and frequency changes was simulated:

$$b_1(t) = 1 + 0.5 \cos(2\pi t/2000), \quad b_2(t) = 1 + 0.5 \sin(2\pi t/2000), \quad \omega(t) = \sin(2\pi t/2000).$$

Figs. 4 and 5 show the comparison of the steady-state mean-squared frequency and parameter estimation errors, yielded by: the GANF algorithm without frequency rate estimation, the GANS algorithm without frequency rate estimation, the GANF algorithm with frequency rate estimation, and the GANS algorithm with frequency rate estimation. The comparison was made for 40 different values of the adaptation gain μ and for two noise intensities: $\sigma_v = 0.56$ (SNR=5 dB) and $\sigma_v = 0.1$ (SNR=20 dB). To reduce the number of design degrees of freedom, the two other gains adopted for GANF/GANS algorithms with frequency rate estimation were set to: $\gamma_\omega = \mu^2/2$ and $\gamma_\alpha = \mu\gamma_\omega/4$ – in agreement with the general tendency revealed in Table 1. The GANF/GANS algorithms without frequency rate estimation were obtained by setting $\gamma_\alpha = 0$. All MSE values were obtained by means of joint time averaging (the evaluation interval [2001,8000] was placed inside a wider analysis interval [1,10000]), and ensemble averaging (100 realizations of measurement noise were used). As expected, the GANS algorithms yielded uniformly better results than their GANF counterparts. The achievable variance reduction is approximately equal to 20 dB for frequency estimation and 10 dB for parameter estimation, respectively. Rather surprisingly, for the particular system analyzed, a relatively small parameter estimation gain can be achieved by switching from the GANS algorithm without frequency rate estimation, to the GANS algorithm with frequency rate estimation (Fig. 4), even though the frequency estimation benefits are evident (Fig. 4).

VIII. CONCLUSION

Identification of quasi-periodically varying systems can be carried out using generalized adaptive notch filters (GANF). We have designed a new GANF algorithm based on quasi-linear model of frequency changes, and we have shown that, under Gaussian assumptions, it is a statistically efficient frequency and frequency rate tracker. Then, based on analysis

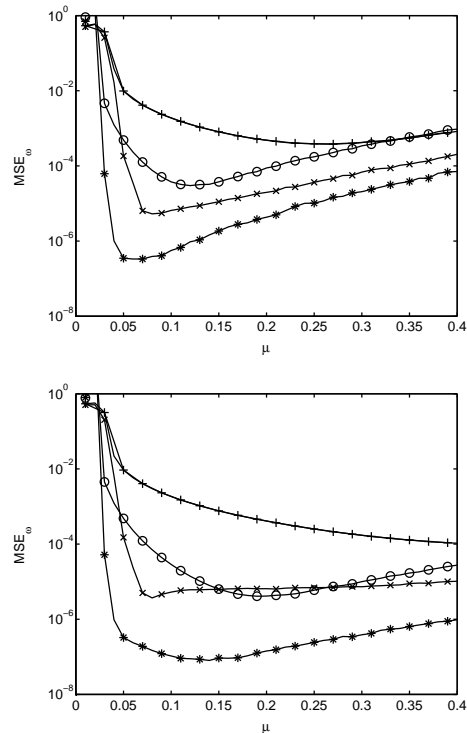


Fig. 4. Dependence of the mean-squared frequency estimation error on adaptation gain μ for: GANF algorithm without frequency rate estimation (+), GANS algorithm without frequency rate estimation (\times), GANF algorithm with frequency rate estimation (O), and GANS algorithm with frequency rate estimation (*). Upper figure: SNR=5 dB, lower figure: SNR= 20 dB.

of tracking properties of the proposed algorithm, we have designed a cascade of post-processing filters increasing accuracy of parameter estimation. We have shown that the resulting generalized adaptive notch smoother (GANS) is statistically efficient (under ideal conditions) and robust to system model misspecification. It yields improved estimation results compared to its causal (GANF) counterpart. It also works better than the currently available, simpler GANS algorithm.

REFERENCES

- Bakkoury, J., Roviras, D., Ghogho, M., & Castanie, F. (2000). Adaptive MLSE receiver over rapidly fading channels. *Signal Processing*, 80, 1347–1360.
- Giannakis, G.B. ,& Tepedelenlioğlu, C. (1998). Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels. *Proc. IEEE*, 86, 1969–1986.
- Haykin, S. (1996). *Adaptive Filter Theory*. Englewood Cliffs NJ: Prentice Hall.
- Jury, M. (1964). *Theory and Application of the Z-transform Method*. New York: Wiley.

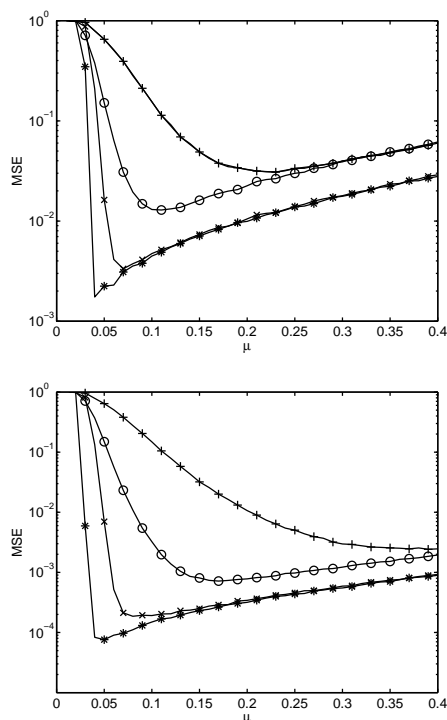


Fig. 5. Dependence of the mean-squared parameter estimation error on adaptation gain μ for: GANF algorithm without frequency rate estimation (+), GANS algorithm without frequency rate estimation (\times), GANF algorithm with frequency rate estimation (\circ), and GANS algorithm with frequency rate estimation (*). Upper figure: SNR=5 dB, lower figure: SNR= 20 dB.

- Niedźwiecki, M., & Kaczmarek, P. (2004). Generalized adaptive notch filters. *in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Montreal, Canada, II-657–II-660.
- Niedźwiecki, M., & Kaczmarek, P. (2005). Identification of quasi-periodically varying systems using the combined nonparametric/parametric approach. *IEEE Trans. Signal Process.*, 53, 4588–4598.
- Niedźwiecki, M., & Sobociński, A. (2007). A simple way of increasing estimation accuracy of generalized adaptive notch filters. *IEEE Signal Process. Lett.*, 14, 217–220.
- Niedźwiecki, M., & Sobociński, A. (2008). Generalized adaptive notch smoothers for real-valued signals and systems. *IEEE Trans. Signal Process.*, 56, 125–133.
- Niedźwiecki, M. (2010). Generalized adaptive notch smoothing revisited. *IEEE Trans. Signal Process.*, 58, pp. 1565–1576.
- Olkin, I., & Pratt, J. (1958). A multivariate Tchebycheff inequality. *Annals of Mathematical Statistics*, 29, 226–234.
- Regalia, P.A. (1995). *Adaptive IIR Filtering in Signal Processing and Control*. New York: Marcel Dekker.
- Tichavský, P., & Händel, P. (1995). Two algorithms for adaptive retrieval of slowly time-varying multiple cisoids in noise. *IEEE Trans. on Signal Process.*, 43, 1116–1127.
- Tsatsanis, M.K., & Giannakis, G.B. (1996). Modeling and equalization of rapidly fading channels. *Int. J. Adaptive Contr. Signal Processing*, 10, 159–176.
- Widrow, B., & Stearns, S.D. (1985). *Adaptive Signal Processing*. Prentice-Hall.
- van Trees, H. *Detection, Estimation and Modulation Theory*. New York: Wiley.

APPENDIX 1

Derivation of (7) and (8)

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

To derive recursion for $\Delta\hat{x}(t)$, note that

$$\begin{aligned}\hat{\boldsymbol{\theta}}(t) &= \boldsymbol{\zeta}(t) + \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\varepsilon(t) \\ \varepsilon(t) &= \boldsymbol{\varphi}^T(t)\boldsymbol{\theta}(t) + v(t) - \boldsymbol{\varphi}^T(t)\boldsymbol{\zeta}(t)\end{aligned}$$

where

$$\boldsymbol{\zeta}(t) = e^{j[\hat{\omega}(t-1)+\hat{\alpha}(t-1)]}\hat{\boldsymbol{\theta}}(t-1).$$

Therefore

$$\begin{aligned}\hat{\boldsymbol{\theta}}(t) &= [\mathbf{I} - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^T(t)]\boldsymbol{\zeta}(t) \\ &\quad + \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}(t)\boldsymbol{\theta}(t) + \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)v(t).\end{aligned}$$

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

$$\begin{aligned}\Delta\hat{\boldsymbol{\theta}}(t) &= [\mathbf{I} - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^T(t)]\boldsymbol{\theta}(t) \\ &\quad - [\mathbf{I} - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^T(t)]\boldsymbol{\zeta}(t) - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)v(t).\end{aligned}\tag{45}$$

Note that $\boldsymbol{\zeta}(t)$ can be rewritten in the form

$$\boldsymbol{\zeta}(t) = e^{j\omega(t)}e^{-j\Delta\hat{\omega}(t-1)}e^{-j\Delta\hat{\alpha}(t-1)}[\boldsymbol{\theta}(t-1) - \Delta\hat{\boldsymbol{\theta}}(t-1)].$$

Using the following approximations $e^{-j\Delta\hat{\omega}(t-1)} \cong 1 - j\Delta\hat{\omega}(t-1)$, $e^{-j\Delta\hat{\alpha}(t-1)} \cong 1 - j\Delta\hat{\alpha}(t-1)$, that hold for small frequency and frequency rate errors, respectively, and applying ALF rules, one arrives at

$$\begin{aligned}\boldsymbol{\zeta}(t) &= e^{j\omega(t)}e^{-j\Delta\hat{\omega}(t-1)}e^{-j\Delta\hat{\alpha}(t-1)}[\boldsymbol{\theta}(t-1) - \Delta\hat{\boldsymbol{\theta}}(t-1)] \\ &\cong e^{j\omega(t)}[1 - j\Delta\hat{\omega}(t-1)][1 - j\Delta\hat{\alpha}(t-1)][\boldsymbol{\theta}(t-1) - \Delta\hat{\boldsymbol{\theta}}(t-1)] \\ &\cong e^{j\omega(t)}[1 - j\Delta\hat{\omega}(t-1) - j\Delta\hat{\alpha}(t-1)][\boldsymbol{\theta}(t-1) - \Delta\hat{\boldsymbol{\theta}}(t-1)] \\ &\cong \boldsymbol{\theta}(t) - e^{j\omega(t)}\Delta\hat{\boldsymbol{\theta}}(t-1) - j[\Delta\hat{\omega}(t-1) + \Delta\hat{\alpha}(t-1)]\boldsymbol{\theta}(t).\end{aligned}\tag{46}$$

Combining (45) with (46), one obtains

$$\begin{aligned}\Delta\widehat{\boldsymbol{\theta}}(t) &\cong [\mathbf{I} - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^T(t)] e^{j\omega(t)}\Delta\widehat{\boldsymbol{\theta}}(t-1) \\ &+ j [\mathbf{I} - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^T(t)] [\Delta\widehat{\omega}(t-1) + \Delta\widehat{\alpha}(t-1)]\boldsymbol{\theta}(t) \\ &- \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)v(t).\end{aligned}\quad (47)$$

Let $\Delta\widetilde{\boldsymbol{\theta}}(t) = \Delta\widehat{\boldsymbol{\theta}}(t)f^*(t)$. After multiplying both sides of (47) with $f^*(t)$, one arrives at

$$\begin{aligned}\Delta\widetilde{\boldsymbol{\theta}}(t) &\cong [\mathbf{I} - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^T(t)] \Delta\widetilde{\boldsymbol{\theta}}(t-1) \\ &+ j [\mathbf{I} - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^T(t)] [\Delta\widehat{\omega}(t-1) + \Delta\widehat{\alpha}(t-1)]\boldsymbol{\beta}_0 \\ &- \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)f^*(t)v(t).\end{aligned}\quad (48)$$

For small values of adaptation gains μ , γ_ω and γ_α , the quantities $\Delta\widetilde{\boldsymbol{\theta}}(t)$, $\Delta\widehat{\omega}(t)$ and $\Delta\widehat{\alpha}(t)$ change slowly compared to $\boldsymbol{\varphi}(t)$. Under such circumstances the error analysis can be carried out using the stochastic averaging technique, i.e., by studying the following equation

$$\begin{aligned}\Delta\widetilde{\boldsymbol{\theta}}(t) &\cong (1 - \mu)\Delta\widetilde{\boldsymbol{\theta}}(t-1) + j(1 - \mu)[\Delta\widehat{\omega}(t-1) + \Delta\widehat{\alpha}(t-1)]\boldsymbol{\beta}_0 \\ &- \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)f^*(t)v(t)\end{aligned}\quad (49)$$

obtained by replacing in (48) the data-dependent matrix $\mathbf{I} - \mu\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi}^*(t)\boldsymbol{\varphi}^T(t)$ with its expectation: $(1 - \mu)\mathbf{I}$.

Multiplying both sides of (49) with $\boldsymbol{\beta}_0^H\boldsymbol{\Phi}$, one obtains

$$\begin{aligned}\boldsymbol{\beta}_0^H\boldsymbol{\Phi}\Delta\widetilde{\boldsymbol{\theta}}(t) &\cong \lambda\boldsymbol{\beta}_0^H\boldsymbol{\Phi}\Delta\widetilde{\boldsymbol{\theta}}(t-1) + j\lambda[\Delta\widehat{\omega}(t-1) + \Delta\widehat{\alpha}(t-1)]\boldsymbol{\beta}_0^H\boldsymbol{\Phi}\boldsymbol{\beta}_0 \\ &- \mu\boldsymbol{\beta}_0^H\boldsymbol{\varphi}^*(t)f^*(t)v(t).\end{aligned}\quad (50)$$

Finally, dividing both sides of (50) by b_0^2 , taking imaginary parts, and noting that $\boldsymbol{\beta}_0^H\boldsymbol{\Phi}\Delta\widetilde{\boldsymbol{\theta}}(t) = \boldsymbol{\theta}(t)^H\boldsymbol{\Phi}\Delta\widehat{\boldsymbol{\theta}}(t)$, $\boldsymbol{\beta}_0^H\boldsymbol{\Phi}\boldsymbol{\beta}_0 = b_0^2$ and $\boldsymbol{\beta}_0^H f^*(t) = \boldsymbol{\theta}^H(t)$, one arrives at

$$\begin{aligned}\Delta\widehat{x}(t) &\cong \lambda\Delta\widehat{x}(t-1) + \lambda[\Delta\widehat{\omega}(t-1) + \Delta\widehat{\alpha}(t-1)] \\ &+ \mu e(t).\end{aligned}\quad (51)$$

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

$$\delta(t) = \frac{\text{Im}[\varepsilon^*(t)\boldsymbol{\varphi}^T(t)\boldsymbol{\zeta}(t)]}{\widehat{\boldsymbol{\beta}}^H(t-1)\boldsymbol{\Phi}\widehat{\boldsymbol{\beta}}(t-1)} \cong \frac{\text{Im}[\varepsilon^*(t)\boldsymbol{\varphi}^T(t)\boldsymbol{\zeta}(t)]}{b_0^2}.$$

Furthermore

$$\text{Im}[\varepsilon^*(t)\varphi^T(t)\zeta(t)] = \text{Im}[z(t) - |\varphi^T(t)\zeta|^2] = \text{Im}[z(t)]$$

where

$$z(t) = \boldsymbol{\theta}^H(t)\varphi^*(t)\varphi^T(t)\zeta(t) + v^*(t)\varphi^T(t)\zeta(t)$$

Using (46) and applying the ALF rules, one obtains the following approximation

$$\begin{aligned} z(t) &\cong \boldsymbol{\theta}^H(t)\varphi^*(t)\varphi^T(t)\boldsymbol{\theta}(t) - j\boldsymbol{\theta}^H(t)\varphi^*(t)\varphi^T(t)\boldsymbol{\theta}(t)[\Delta\widehat{\omega}(t-1) + \Delta\widehat{\alpha}(t-1)] \\ &\quad - \boldsymbol{\theta}^H(t-1)\varphi^*(t)\varphi^T(t)\Delta\widehat{\boldsymbol{\theta}}(t-1) + v^*(t)\varphi^T(t)\boldsymbol{\theta}(t) \\ &= \boldsymbol{\beta}_0^H\varphi^*(t)\varphi^T(t)\boldsymbol{\beta}_0 - j\boldsymbol{\beta}_0^H\varphi^*(t)\varphi^T(t)\boldsymbol{\beta}_0[\Delta\widehat{\omega}(t-1) + \Delta\widehat{\alpha}(t-1)] \\ &\quad - \boldsymbol{\beta}_0^H\varphi^*(t)\varphi^T(t)\Delta\widetilde{\boldsymbol{\theta}}(t-1) + v^*(t)\varphi^T(t)\boldsymbol{\theta}(t) \end{aligned}$$

which, after stochastic averaging, leads to

$$\begin{aligned} z(t) &\cong \boldsymbol{\beta}_0^H\boldsymbol{\Phi}\boldsymbol{\beta}_0 - j\boldsymbol{\beta}_0^H\boldsymbol{\Phi}\boldsymbol{\beta}_0[\Delta\widehat{\omega}(t-1) + \Delta\widehat{\alpha}(t-1)] \\ &\quad - \boldsymbol{\beta}_0^H\boldsymbol{\Phi}\Delta\widetilde{\boldsymbol{\theta}}(t-1) + v^*(t)\varphi^T(t)\boldsymbol{\theta}(t). \end{aligned}$$

Since $\boldsymbol{\beta}_0^H\boldsymbol{\Phi}\boldsymbol{\beta}_0 = b_0^2$, one arrives at

$$\delta(t) = \text{Im}[z(t)/b_0^2] \cong -\Delta\widehat{x}(t-1) - \Delta\widehat{\omega}(t-1) - \Delta\widehat{\alpha}(t-1) + e(t).$$

Note that

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

Combining the last three equations, one arrives at

$$\Delta\widehat{\alpha}(t) \cong (1 - \gamma_\alpha)\Delta\widehat{\alpha}(t-1) + w(t) + \gamma_\alpha e(t) - \gamma_\alpha\Delta\widehat{\omega}(t-1) - \gamma_\alpha\Delta\widehat{x}(t-1) \quad (52)$$

$$\Delta\widehat{\omega}(t) \cong (1 - \gamma_\omega)\Delta\widehat{\omega}(t-1) + (1 - \gamma_\omega)\Delta\widehat{\alpha}(t-1) + \gamma_\omega e(t) - \gamma_\omega\Delta\widehat{x}(t-1). \quad (53)$$

Finally, solving the set of linear equations (51), (52) and (53) for $\Delta\widehat{\omega}(t)$ and $\Delta\widehat{\alpha}(t)$, one obtains (7) and (8), respectively.

APPENDIX 2

Computation of Lower Tracking/Smoothing Bounds

In this appendix, we will derive expressions for theoretical upper bounds that limit tracking/smoothing capabilities of *any* causal/noncausal frequency and frequency rate estimation algorithms applied to quasi-periodically varying systems with quasi-linear frequency changes. The corresponding lower tracking bounds (LTB) and lower smoothing bounds (LSB) belong to the class of posterior (or Bayesian) Cramér-Rao bounds, applicable to signals/systems with random parameters¹.

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

$$\mathbb{E}[(\widehat{\mathbf{x}}(\mathbf{y}, \boldsymbol{\phi}) - \mathbf{x})(\widehat{\mathbf{x}}(\mathbf{y}, \boldsymbol{\phi}) - \mathbf{x})^T | \boldsymbol{\phi}] \geq \mathbf{J}^{-1}(\boldsymbol{\phi})$$

where

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

and $p(\mathbf{y}, \mathbf{x} | \boldsymbol{\phi}) = \mathbf{p}(\mathbf{y} | \mathbf{x}, \boldsymbol{\phi}) \mathbf{p}(\mathbf{x})$ is the joint probability density function of the pair (\mathbf{y}, \mathbf{x}) given $\boldsymbol{\phi}$.

When the input signal is a stochastic process, and when averaging is extended to all realizations of $\boldsymbol{\phi}$, one obtains the following result

$$\mathbb{E}[(\widehat{\mathbf{x}}(\mathbf{y}, \boldsymbol{\phi}) - \mathbf{x})(\widehat{\mathbf{x}}(\mathbf{y}, \boldsymbol{\phi}) - \mathbf{x})^T] \geq \mathbb{E}[\mathbf{J}^{-1}(\boldsymbol{\phi})] \geq \{\mathbb{E}[\mathbf{J}(\boldsymbol{\phi})]\}^{-1} = \bar{\mathbf{J}}^{-1}$$

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

In the case considered, let $\mathbf{x}_t = [\alpha(\mathbf{1}), \alpha(\mathbf{2}), \dots, \alpha(\mathbf{t})]^T$, $\mathbf{y}_t = [\mathbf{y}(\mathbf{1}), \mathbf{y}(\mathbf{2}), \dots, \mathbf{y}(\mathbf{t})]^T$ and $\boldsymbol{\phi}_t = [\boldsymbol{\varphi}^T(1), \boldsymbol{\varphi}^T(2), \dots, \boldsymbol{\varphi}^T(t)]^T$. To simplify our analysis (without restricting its gener-

¹When the estimated quantities are stochastic variables, rather than unknown deterministic constants, the classical Cramér-Rao inequality does not apply.

ality), we will assume, in addition to (A1)–(A4), that $\boldsymbol{\theta}(0) = \boldsymbol{\beta}_0$ is a known deterministic quantity, that $\alpha(0)$ is uniformly distributed over $[\alpha_{\min}, \alpha_{\max}]$, and that $\omega(0) = 0$.

First of all, note that

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

Since, under the assumptions listed above, the vector \mathbf{x}_t fully determines $\boldsymbol{\theta}(k)$, $k = 1, \dots, t$

$$\begin{aligned} \boldsymbol{\theta}(k) &= \boldsymbol{\beta}_0 e^{j\psi(k)} \\ \psi(k) &= \sum_{n=1}^k \omega(n) = \sum_{n=1}^k \sum_{m=1}^{n-1} \alpha(m) \end{aligned} \quad (54)$$

one can write

$$\begin{aligned} \log p(\mathbf{y}_t | \mathbf{x}_t, \boldsymbol{\phi}_t) &= \log \mathbf{p}[\mathbf{y}_t | \boldsymbol{\theta}(1), \boldsymbol{\varphi}(1), \dots, \boldsymbol{\theta}(t), \boldsymbol{\varphi}(t)] \\ &= c_1 - \frac{1}{\sigma_v^2} \sum_{k=1}^t |v(k)|^2 = c_1 - \frac{1}{\sigma_v^2} \sum_{k=1}^t |y(k) - \boldsymbol{\varphi}^T(k) \boldsymbol{\theta}(k)|^2 \end{aligned} \quad (55)$$

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

Differentiating (55) with respect to $\alpha(i)$, one obtains

$$\begin{aligned} &\frac{\partial \log p(\mathbf{y}_t | \mathbf{x}_t, \boldsymbol{\phi}_t)}{\partial \alpha(m)} \\ &= -\frac{1}{\sigma_v^2} \frac{\partial}{\partial \alpha(m)} \sum_{k=1}^t [|y(k)|^2 - y^*(k) \boldsymbol{\varphi}^T(k) \boldsymbol{\theta}(k) y(k) \boldsymbol{\varphi}^H(k) \boldsymbol{\theta}^*(k) + \boldsymbol{\theta}^H(k) \boldsymbol{\varphi}^*(k) \boldsymbol{\varphi}^T(k) \boldsymbol{\theta}(k)] \\ &= \frac{2}{\sigma_v^2} \sum_{k=1}^t \operatorname{Re} \left\{ y^*(k) \boldsymbol{\varphi}^T(k) \frac{\partial \boldsymbol{\theta}(k)}{\partial \alpha(m)} \right\} \end{aligned}$$

where the second transition follows from the fact that the terms $|y(k)|^2$ and $\boldsymbol{\theta}^H(k) \boldsymbol{\varphi}^*(k) \boldsymbol{\varphi}^T(k) \boldsymbol{\theta}(k) = \boldsymbol{\beta}_0^H \boldsymbol{\varphi}^*(k) \boldsymbol{\varphi}^T(k) \boldsymbol{\beta}_0$ do not depend on $\alpha(m)$.

Using (54), one arrives at

$$\frac{\partial \boldsymbol{\theta}(k)}{\partial \alpha(m)} = j \boldsymbol{\theta}(k) \sum_{s=1}^k \sum_{l=1}^{s-1} \delta_{m,l} = j \boldsymbol{\theta}(k) \max(k-m, 0)$$

where $\delta_{m,l} = \{0 \text{ if } m \neq l, 1 \text{ if } m = l\}$ denotes the Kronecker delta. This leads to

$$\frac{\partial \log p(\mathbf{y}_t | \mathbf{x}_t, \boldsymbol{\phi}_t)}{\partial \alpha(m)} = \frac{2}{\sigma_v^2} \sum_{k=1}^t \operatorname{Re} \left\{ j y^*(k) \boldsymbol{\varphi}^T(k) \boldsymbol{\theta}(k) \max(k-m, 0) \right\}$$

and

$$\frac{\partial^2 \log p(\mathbf{y}_t | \mathbf{x}_t, \boldsymbol{\phi}_t)}{\partial \alpha(m) \partial \alpha(n)} = -\frac{2}{\sigma_v^2} \sum_{k=1}^t \operatorname{Re} \{ y^*(k) \boldsymbol{\varphi}^T(k) \boldsymbol{\theta}(k) \max(k-m, 0) \max(k-n, 0) \}$$

Since $y^*(k) = \boldsymbol{\varphi}^H(k) \boldsymbol{\theta}^*(k) + v^*(k)$, one obtains

$$\begin{aligned} \mathbb{E}[y^*(k) \boldsymbol{\varphi}^T(k) \boldsymbol{\theta}(k)] &= \mathbb{E}[\boldsymbol{\theta}^H(k) \boldsymbol{\varphi}^*(k) \boldsymbol{\varphi}^T(k) \boldsymbol{\theta}(k)] + \mathbb{E}[v^*(k) \boldsymbol{\varphi}^T(k) \boldsymbol{\theta}(k)] \\ &= \mathbb{E}[\boldsymbol{\beta}_0^H \boldsymbol{\varphi}^*(k) \boldsymbol{\varphi}^T(k) \boldsymbol{\beta}_0] = \boldsymbol{\beta}_0^H \boldsymbol{\Phi} \boldsymbol{\beta}_0 = b_0^2 \end{aligned}$$

which leads to

$$-\mathbb{E} \left[\frac{\partial^2 \log p(\mathbf{y}_t | \mathbf{x}_t, \boldsymbol{\phi}_t)}{\partial \mathbf{x}_t \partial \mathbf{x}_t^T} \right] = \frac{2b_0^2}{\sigma_v^2} \mathbf{A}_t \quad (56)$$

where

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

In an analogous way one can derive the second component of the generalized Fisher matrix.

First, note that

$$\begin{aligned} \log p(\mathbf{x}_t) &= \log p[\alpha(1), \alpha(2) - \alpha(1), \dots, \alpha(t) - \alpha(t-1)] \\ &= c_\alpha + c_2 - \frac{1}{2\sigma_w^2} \sum_{k=2}^t w^2(k) = c_\alpha + c_2 - \frac{1}{2\sigma_w^2} \sum_{k=2}^t [\alpha(k) - \alpha(k-1)]^2 \end{aligned} \quad (57)$$

where $c_\alpha = \log[1/(\alpha_{\max} - \alpha_{\min})]$ and c_2 are constants independent of \mathbf{x}_t .

Differentiation of (57) results in

$$\begin{aligned} \frac{\log p(\mathbf{x}_t)}{\alpha(m)} &= -\frac{1}{2\sigma_w^2} \sum_{k=2}^t \frac{\partial}{\partial \alpha(m)} [\alpha(k) - \alpha(k-1)]^2 \\ &= \frac{1}{\sigma_w^2} \begin{cases} \alpha(2) - \alpha(1) & \text{for } m = 1 \\ \alpha(m+1) - 2\alpha(m) + \alpha(t-1) & \text{for } 1 < m < t \\ \alpha(t-1) - \alpha(t) & \text{for } m = t \end{cases} \\ &= \frac{1}{\sigma_w^2} \begin{cases} w(2) & \text{for } m = 1 \\ w(m+1) - w(m) & \text{for } 1 < m < t \\ -w(t) & \text{for } m = t \end{cases} \end{aligned}$$

which leads to

$$\mathbb{E} \left[\frac{\partial \log p(\mathbf{x}_t)}{\partial \mathbf{x}_t} \frac{\partial \log p(\mathbf{x}_t)}{\partial \mathbf{x}_t^T} \right] = \frac{1}{\sigma_w^2} \mathbf{B}_t \quad (58)$$

where

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

Combining (56) with (58), one obtains

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

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

$$\text{LTB}_\omega = \lim_{t \rightarrow \infty} \inf_{\hat{\omega}(\cdot)} \mathbb{E} \{ [\omega(t) - \hat{\omega}(t)]^2 \} = \lim_{t \rightarrow \infty} \mathbf{b}_t^T \mathbf{J}_t^{-1} \mathbf{b}_t$$

$$\text{LTB}_\alpha = \lim_{t \rightarrow \infty} \inf_{\hat{\alpha}(\cdot)} \mathbb{E} \{ [\alpha(t) - \hat{\alpha}(t)]^2 \} = \lim_{t \rightarrow \infty} [\mathbf{J}_t^{-1}]_{tt}$$

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

$$\text{LSB}_\omega = \lim_{t \rightarrow \infty} \inf_{\tilde{\omega}(\cdot)} \mathbb{E} \{ [\omega(t) - \tilde{\omega}(t)]^2 \} = \lim_{t \rightarrow \infty} \mathbf{c}_t^T \mathbf{J}_{2t}^{-1} \mathbf{c}_t$$

$$\text{LSB}_\alpha = \lim_{t \rightarrow \infty} \inf_{\tilde{\alpha}(\cdot)} \mathbb{E} \{ [\alpha(t) - \tilde{\alpha}(t)]^2 \} = \lim_{t \rightarrow \infty} [\mathbf{J}_{2t}^{-1}]_{tt}$$

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