Generalized adaptive comb filters/smoothers and their application to the identification of quasi-periodically varying systems and signals

Maciej Niedźwiecki 1, Michał Meller

Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control, Gdańsk University of Technology, Narutowicza 11/12, 80-233 Gdańsk, Poland

A R T I C L E   I N F O

Article history:
Received 17 May 2012
Received in revised form 27 December 2012
Accepted 22 January 2013
Available online 26 March 2013

Keywords:
System identification
Time-varying processes

A B S T R A C T

The problem of both causal and noncausal identification of linear stochastic systems with quasi-harmonically varying parameters is considered. The quasi-harmonic description allows one to model nonsinusoidal quasi-periodic parameter changes. The proposed identification algorithms are called generalized adaptive comb filters/smoothers because in the special signal case they reduce down to adaptive comb algorithms used to enhance or suppress nonstationary harmonic signals embedded in noise. The paper presents a thorough statistical analysis of generalized adaptive comb algorithms, and demonstrates their statistical efficiency in the case where the fundamental frequency of parameter changes varies slowly with time according to the integrated random-walk model.

1. Introduction

1.1. Problem statement

We will consider the problem of identification of quasi-periodically varying complex-valued systems governed by

\[ y(t) = \sum_{i=1}^{n} \theta_i(t)u(t-i+1) + \epsilon(t) = \varphi^T(t)\theta(t) + \epsilon(t) \]

where \( t = 1, 2, \ldots \) denotes the normalized discrete time, \( y(t) \) denotes the system output, \( \varphi(t) = [u(t), \ldots, u(t-n+1)]^T \) denotes regression vector, made up of the past input samples, \( \epsilon(t) \) denotes measurement noise, and \( \theta(t) = [\theta_1(t), \ldots, \theta_n(t)]^T \) is the vector of time-varying system coefficients, modeled as weighted sums of complex exponentials

\[ \theta(t) = \sum_{k=1}^{K} \beta_k(t)e^{j\sum_{i=1}^{l} a_{ki}(t)} \]

where \( \beta_k(t) = [\beta_{k1}(t), \ldots, \beta_{kn}(t)]^T \)

\[ b_{ki}(t) = a_{ki}(t)e^{j\omega_0 t}, \quad i = 1, \ldots, n. \]

The following three types of real-valued quantities are incorporated in (2): the instantaneous angular frequencies \( \omega_k(t) \), the instantaneous amplitudes \( a_k(t) \), and the time-invariant phase shifts \( \omega_0 \). With a slight abuse of terminology, the complex-valued vectors \( \beta_k(t) \) will be further referred to as 'complex amplitudes'.

Under certain circumstances (in the presence of several strong reflectors) the model (1)–(2) can be used to describe rapidly fading mobile radio channels (Bakkoury, Roviras, Ghogho, & Castanie, 2000; Giannakis & Tepedelenlioğlu, 1998; Tsatsanis & Giannakis, 1996). In this case \( y(t) \) denotes the sampled baseband signal received by the mobile unit, \( u(t) \) denotes the sequence of transmitted symbols, and \( \epsilon(t) \) denotes channel noise.

We will assume that the frequencies \( \omega_k(t) \) are harmonically related, namely

\[ \omega_k(t) = m_k\omega_0(t), \quad k = 1, \ldots, K \]

where \( \omega(t) \) denotes the slowly varying fundamental frequency and \( m_k \) are integer numbers. Such multiple frequencies, called harmonics, appear in the Fourier series expansions of periodic signals. For example, if parameter trajectory \( \theta(t) \) is periodic with period \( L \), it admits the following Fourier representation: \( \theta(t) = \sum_{k=0}^{L-1} \beta_k e^{j2\pi k t L} \), \( \omega_0 = (2\pi)/L \).

The notion of ‘time-varying harmonics’ can be regarded as a natural extension of the Fourier analysis to quasi-periodically varying systems, such as (1)–(2). The choice of the multipliers \( m_k, k = 1, \ldots, K, \) depends on our prior knowledge of the system time variation. When all harmonics are expected to be present, one should set \( m_k = k \). In the presence of odd harmonics only, the natural choice is \( m_k = 2k - 1 \), etc.
In the special case where \( n = 1 \) and \( g(t) \equiv 1 \), Eqs. (1)-(3) describe a complex-valued harmonic signal \( s(t) = \theta(t) \) buried in noise

\[
y(t) = s(t) + v(t), \quad s(t) = \sum_{k=1}^{K} b_k(t) e^{j \sum_{i=1}^{\infty} \omega_i(t)}.
\]

(4)

The problem of either elimination or extraction of harmonic signals buried in noise can be solved using adaptive comb filters (Nehorai & Porat, 1986; Regalia, 1995). For this reason the system identification/tracking algorithm described below can be considered a generalized comb filter.

1.2. Contribution

The problem of causal identification (tracking) of single-mode quasi-periodically varying systems was studied in Niedźwiecki and Kaczmarski (2004, 2005a,b).

In the recent conference paper (Niedźwiecki & Meller, 2011a), the results presented earlier were extended to noncausal identification (smoothing). Additionally, a more sophisticated frequency estimation scheme was proposed, incorporating frequency rate tracking/smoothing and yielding better results in practice.

All papers published so far focus on the identification of single-mode systems, i.e., systems with parameters that can be modeled as complex sinusoids (cisoids) with slowly varying amplitudes and slowly varying instantaneous frequency.

This paper extends results presented in Niedźwiecki and Meller (2011a) to nonstationary systems with quasi-harmonically varying parameters, i.e., to systems with several frequency modes governed by the same slowly varying fundamental frequency. In practice such harmonic modes of variation often arise in oscillatory systems with nonlinear elements and/or loads (Neimark, 2003).

In principle, quasi-harmonically varying systems can be identified using the multiple-frequency versions of the algorithms mentioned above. Such algorithms are made up of several single-frequency sub-algorithms that work in parallel and are driven by the common prediction error. Since the estimated frequencies are in this case regarded as mutually unrelated quantities, the harmonic structure of the system/signal time variation is not exploited in any way. In this paper we present algorithms that take advantage of such a priori information, i.e., the algorithms that perform a coordinated frequency search. This allows one to improve estimation results considerably.

2. Generalized adaptive notch filter — overview of known results

Suppose that the identified nonstationary system has a single frequency mode \( (K = 1) \), i.e., it is governed by

\[
y(t) = \varphi^T(t) \Theta(t) + v(t), \quad \Theta(t) = \beta(t) e^{j \sum_{i=1}^{\infty} \omega_i(t)}
\]

(5)

where \( \beta(t) = [b_1(t), \ldots, b_N(t)]^T \) and \( \omega(t) \in (\pi - \pi, \pi) \) are slowly varying quantities. Furthermore, suppose that:

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

\( \text{(A2)} \) The sequence of regression vectors \( \{\varphi(t)\} \), independent of \( \{v(t)\} \), is nondeterministic, wide-sense stationary and ergodic with known correlation matrix \( \Phi = E[\varphi^T(t) \varphi(t)] \).

Denote by \( \alpha(t) = \omega(t + 1) - \omega(t) \) the rate of change of the instantaneous frequency \( \omega(t) \). Under assumptions \( \text{(A1)}-\text{(A2)} \), identification of the system \( (5) \) can be carried out using the following generalized adaptive notch filtering (GANF) algorithm proposed in Niedźwiecki and Meller (2011a)

\[
\hat{f}(t) = \phi \{ \hat{\Theta}(t-1) + \varepsilon(t-1) \}^T \hat{\beta}(t-1) - 1
\]

\[
\varepsilon(t) = y(t) - \varphi^T(t) \hat{f}(t) \hat{\beta}(t-1)
\]

\[
\hat{\beta}(t) = \hat{\beta}(t-1) + \mu \Phi^{-1} \varepsilon(t) \hat{f}(t)^T \varepsilon(t)
\]

\[
g(t) = \frac{\varepsilon(t) \varphi^T(t) \hat{f}(t) \hat{\beta}(t-1)}{\beta(t-1)}
\]

(6)

where \( \hat{f}(t) \) is an estimate of \( f(t) = e^{j \sum_{i=1}^{\infty} \omega_i(t)} \), and \( \mu > 0, \gamma_0 > 0, \gamma_0 > 0, \gamma_0 > 0, \) such that \( \gamma_0 \ll \gamma_0 \ll \mu \), denote small adaptation gains determining the rate of amplitude adaptation, frequency adaptation and frequency rate adaptation, respectively.

The gradient search strategy, incorporated in (6) for the purpose of tracking \( \alpha(t) \) and \( \omega(t) \), is based on minimization of the following instantaneous measure of fit \( (7) \) for the (6).

\[
f(t) = |\varepsilon(t)|^2 / 2, \text{where } \varepsilon(t) = y(t) - \varphi^T(t) \hat{f}(t) \hat{\beta}(t-1)
\]

(7)

Therefore, the term \( g(t) \) in (6) can be interpreted as a normalized estimate of the gradient (7). Normalization makes the algorithm scale-invariant. When it is not applied, the tracking properties of the GANF algorithm depend not only on the user-defined adaptation gains \( \gamma_0 \) and \( \gamma_0 \), but also on the system-related variables, which is inconvenient from the practical viewpoint.

Tracking properties of the GANF algorithm (6) were analyzed in Niedźwiecki and Meller (2011a) in the case where the vector of amplitudes \( \hat{\beta}(t) \) is unknown but constant:

\[
\text{(A3\textsuperscript{*}) } \hat{\beta}(t) \equiv \beta, \text{ i.e., } \Theta(t) = \hat{\Theta}(t-1), \forall t.
\]

Note that under (A3\textsuperscript{*}) the normalization term \( 1 / \hat{\beta}(t-1) \) can be regarded as an estimate of the power of the noiseless system output

\[
b^2 = \beta^T \Phi \beta = E[\hat{\beta}^T \varphi^T(t) \varphi(t) \hat{\beta}] = E\left[|\varphi^T(t) \Theta(t)|^2\right].
\]

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

\[
\Delta \hat{\omega}(t) \approx G_1(q^{-1}) e(t) + G_2(q^{-1}) \delta(t)
\]

\[
\Delta \hat{\alpha}(t) \approx H_1(q^{-1}) e(t) + H_2(q^{-1}) \delta(t)
\]

where \( \{e(t)\}, e(t) = -\varepsilon(t) \varphi^T(t) \hat{f}(t) \hat{\beta}(t-1) \) is a zero-mean white noise with variance \( \sigma_e^2 = \sigma_v^2 / (2B^2) \), \( \delta(t) = \alpha(t) - \alpha(t-1) \) denotes the one-step change of the frequency rate, and

\[
G_1(q^{-1}) = (1 - q^{-1})[\gamma_0 + (\gamma_0 - \gamma_0)q^{-1}] / D(q^{-1})
\]

\[
G_2(q^{-1}) = q^{-1} [1 - \gamma_0 - (1 - \mu)q^{-1}] / D(q^{-1})
\]

\[
H_1(q^{-1}) = \gamma_0 (1 - q^{-1})^2 / D(q^{-1})
\]

\[
H_2(q^{-1}) = [1 + (\mu + \gamma_0 - 2)q^{-1} + (1 - \mu)q^{-2}] / D(q^{-1})
\]

\[2\] Hereafter the symbol * will denote complex conjugation, and the symbol H — Hermitian (conjugate) transpose.
and final parameter estimate could be easily obtained by combining the \( \theta \) independent GANF algorithms of the form (6), each taking care of the signals referred to as quasi-linear.

\[
D(q^{-1}) = 1 + d_1 q^{-1} + d_2 q^{-2} + d_3 q^{-3}, d_1 = \mu + \gamma_0 + \gamma_0 - 3, d_2 = 3 - 2 \mu - \gamma_0, d_3 = \mu - 1. \]

All filters are asymptotically stable if adaptation gains fulfill the following (sufficient) stability conditions: \( 0 < \mu < 1.0 < \gamma_0 < 1.0 < \gamma_0 < 1 \) and \( \mu(\gamma_0 + \gamma_0) > \gamma_0. \)

In spite of its simplicity, the gradient frequency tracking mechanism adopted in (6) has very good statistical properties — as shown in Niedźwiecki and Meller (2011a), when the instantaneous frequency drifts according to the Gaussian integrated random-walk (IRW) model, namely

(A4) \( \delta(t) \), independent of \( [\nu(t)] \) and \( [\varphi(t)] \), is a zero-mean white sequence with variance \( \sigma^2_{\nu} \).

(A5) The sequences \( [\nu(t)] \) and \( [\delta(t)] \) are normally distributed, the optimally tuned GANF algorithm (6) is statistically efficient, i.e., it reaches the Cramér–Rao-type lower frequency and frequency rate tracking bounds.

Note that \( \alpha(t) - \alpha(t-1) = \delta(t) \) implies \( (1-q^{-1})^2 \alpha(t) = \delta(t-1) \). Since \( (1-q^{-1})^2 \alpha(t) = 0 \) entails \( \alpha(t) = \gamma_1 + \gamma_2 t \), where \( \gamma_1 \) and \( \gamma_2 \) denote arbitrary constants, the IRW model can be regarded as a perturbed linear growth/decay model — for small perturbations \( (\alpha_1 < \alpha_2) / b \) the corresponding frequency changes will be further referred to as quasi-linear.

3. Multiple-frequency GANF

Denote by \( y_k(t) = \varphi(t) \theta_k(t) + v(t) \), where \( \theta_k(t) = f_k(t) \beta_k(t) \), and \( f_k(t) \) is an estimate of \( f(t) = \varphi \sum_{i=1}^K \theta_i(t) \), the output of this subsystem of (1) which is associated with the frequency \( \omega_k \). If the signals \( y_1(t), \ldots, y_K(t) \) were measurable, one could design \( K \) independent GANF algorithms of the form (6), each taking care of a particular subsystem. Since it holds that \( \theta(t) = \sum_{k=1}^{K} \theta_k(t) \), the final parameter estimate could be easily obtained by combining the partial estimates.

Even though the outputs \( y_k(t) \) are not available, one can replace them with the surrogate (estimated) outputs obtained from

\[
\hat{y}_k(t) = y(t) - \varphi(t) \sum_{i=1}^{K} \hat{\theta}_i(t) |_{t-1}
\]

where \( \hat{\theta}_k(t-1) = \hat{f}_k(t-1) \beta_k(t-1) \) denotes the one-step-ahead prediction of \( \theta_k(t) \). Note that after replacing \( y(t) \) with \( \hat{y}_k(t) \) in (6), one obtains \( e_k(t) = \hat{y}_k(t) - \varphi(t) \hat{\theta}_k(t-1) = y(t) - \varphi(t) \sum_{i=1}^{K} \hat{f}_i(t) \beta_i(t-1) = e(t), \forall k \), which means that all sub-algorithms should be driven by the same ‘global’ prediction error. Such an approach was used, with good results, to design multiple-frequency algorithms in Niedźwiecki and Kaczmarek (2004). When applied to (6) it yields

\[
\hat{f}_k(t) = e^{\varphi \hat{\theta}_k(t-1) + c \hat{y}_k(t-1) \hat{f}_k(t-1)}
\]

\[
\varepsilon(t) = y(t) - \varphi(t) \sum_{k=1}^{K} \hat{f}_k(t) \beta_k(t-1)
\]

\[
\hat{\beta}_k(t) = \hat{\beta}_k(t-1) + \mu \Phi^{-1} \varphi(t) \hat{f}_k(t-1) e(t)
\]

\[
\varepsilon_k(t) = -e^{\varphi \hat{\theta}_k(t-1) + c \hat{y}_k(t-1) \hat{f}_k(t-1)}
\]

\[
\hat{g}_k(t) = \sum_{k=1}^{K} m_k \text{Im} \left[ e^{\varepsilon_k(t)} \varphi(t) \hat{f}_k(t) \beta_k(t-1) \right]
\]

\[
\hat{\theta}_k(t) = \sum_{k=1}^{K} \hat{\theta}_k(t), \quad \hat{\theta}(t) = \sum_{k=1}^{K} \hat{\theta}_k(t).
\]

In Niedźwiecki and Kaczmarek (2005b) it was shown that the number of frequency modes, as well as all initial conditions needed to smoothly start (start without initialization transients) the GANF algorithm, can be inferred from nonparametric DFT-based analysis of a short startup fragment of the input–output data. The tool that can be used for this purpose was termed generalized (system) periodogram, as in the signal case it reduces to the classical periodogram.

When applied to identification of the system (1)–(3), the GANF algorithm (10) has two serious drawbacks.

First, it does not take into consideration the harmonic structure (3), i.e., the estimated frequencies are regarded as mutually unrelated quantities, while the true harmonics vary in a coordinated way. Hence, even though such an unconstrained multiple-frequency generalized adaptive notch filter can be used to identify the multi-harmonic system/signal, its tracking characteristics will be generally inferior to those offered by solutions that incorporate the harmonic constraints.

Second, the algorithm (10) is not robust to incorrect frequency matching. While the strong frequency components, i.e., those characterized by large values of the signal-to-noise ratio \( \text{SNR}_k(t) = \| \hat{f}_k(t) \|_2^2 / \sigma^2 \) are usually tracked successfully, the weak ones may be difficult to follow — even if the initial frequency assignment is correct, the sub-algorithms tracking such weak components may, after some time, lock onto the neighboring, stronger components, corresponding to higher or lower frequencies. Moreover, when the system/signal is nonstationary, the ‘strength’ of different harmonic components may vary with time, which further complicates the picture.

4. Generalized adaptive comb filter

In order to arrive at the algorithm which performs a coordinated search of the instantaneous fundamental frequency \( \omega_{q0}(t) \), one should minimize \( f(t) \) for

\[
\varepsilon(t) = y(t) - \varphi(t) \sum_{k=1}^{K} \hat{f}_k(t) \beta_k(t-1)
\]

where \( \hat{f}_k(t) = e^{\varphi \hat{\theta}_k(t-1) + c \hat{y}_k(t-1) \hat{f}_k(t-1)} \). Note that

\[
\frac{\partial f(t)}{\partial \omega_{q0}(t)} = -\text{Re} \left[ \varphi(t) \left( \sum_{k=1}^{K} m_k \hat{g}_k(t) \beta_k(t-1) \right) \right]
\]

\[
= \sum_{k=1}^{K} m_k \text{Im} \left[ e^{\varepsilon_k(t)} \varphi(t) \hat{f}_k(t) \beta_k(t-1) \right].
\]

This leads to the following recursive estimation scheme which will be further referred to as a generalized adaptive comb filter (GACF)

\[
\hat{f}_k(t) = e^{\varphi \hat{\theta}_k(t-1) + c \hat{y}_k(t-1) \hat{f}_k(t-1)}
\]

\[
\varepsilon(t) = y(t) - \varphi(t) \sum_{k=1}^{K} \hat{f}_k(t) \beta_k(t-1)
\]

\[
\hat{\beta}_k(t) = \hat{\beta}_k(t-1) + \mu \Phi^{-1} \varphi(t) \hat{f}_k(t-1) e(t)
\]

\[
\hat{g}_k(t) = \sum_{k=1}^{K} m_k \text{Im} \left[ e^{\varepsilon_k(t)} \varphi(t) \hat{f}_k(t) \beta_k(t-1) \right]
\]

\[
\hat{\theta}(t) = \sum_{k=1}^{K} \hat{\theta}_k(t), \quad \hat{\theta}(t) = \sum_{k=1}^{K} \hat{\theta}_k(t).
\]
One can show, using the ALF technique, that under (A1)–(A2) and the following assumption
\[(A3) \beta_k(t) = \beta_\delta, \text{i.e., } \theta_k(t) = e^{i\omega_k(t-\delta)} \theta_k(t-1), k = 1, \ldots, K, \forall t,\]
which is a multi-frequency variant of (A3'), the frequency and frequency rate estimation errors \(\Delta \omega_0(t) = \omega_0(t) - \hat{\omega}_0(t)\), \(\Delta \dot{\omega}_0(t) = \dot{\omega}_0(t) - \hat{\dot{\omega}}_0(t)\), can be approximately expressed in the form (see Appendix A)
\[
\Delta \omega_0(t) \approx G_1(q^{-1})\varepsilon_0(t) + G_2(q^{-1})\delta(t) \tag{13}
\]
\[
\Delta \dot{\omega}_0(t) \approx H_1(q^{-1})\varepsilon_0(t) + H_2(q^{-1})\delta(t) \tag{14}
\]
where the transfer functions \(G_1(q^{-1}), G_2(q^{-1}), H_1(q^{-1})\) and \(H_2(q^{-1})\) are identical with those appearing in (8)–(9), and \(\varepsilon_0(t) = \sum_{k=1}^{K} m_k \varepsilon_k(t)\), \(\varepsilon_k(t) = -i \mbox{Im} \{ \beta_k^i \Phi(t) \dot{\phi}(t) \} / b_k^2\), \(b_k^2 = \sum_{k=1}^{K} m_k^2 \beta_k^2(t)\) and \(b_k^2 = \beta_k^i \Phi(t) \beta_k\). Note that the normalizing term \(\sum_{k=1}^{K} m_k^2 \beta_k^2(t)\) is one that appears in the expression for \(\hat{\beta}_k(t)\) in (12), can be regarded as an estimate of \(b_k^2\).

Furthermore, one can show that \(\{ \varepsilon_k(t)\}, k = 1, \ldots, K,\) are zero-mean white noise sequences with cross-correlation functions given by (see Appendix B)
\[
E[\varepsilon_k(t)\varepsilon_s(t)] = \rho_{\varepsilon_k}\delta(t) \quad \text{for } t = s, \\
E[\varepsilon_k(t)\varepsilon_s(t)] = 0 \quad \text{for } t \neq s, \tag{15}
\]
where
\[
\rho_{\varepsilon_k} = \sigma_{\varepsilon_k}^2/2b_0^2 \left\{ \Re [\beta_k^i \Phi(t)] \cos[\phi_k(t) - \dot{\phi}_0(t)] \\
- i \mbox{Im} \{ \beta_k^i \Phi(t) \} \sin[\phi_k(t) - \dot{\phi}_0(t)] \right\}, \tag{16}
\]
and \(\dot{\phi}_0(t) = \sum_{i=1}^K \alpha_i \omega_i(t) = \sum_{i=1}^K \alpha_i \omega_i(0)\).

Setting \(t = k\) in (16), one obtains
\[
\sigma_{\varepsilon_k}^2 = E[\varepsilon_k(t)^2] = \sigma_{\varepsilon_k}^2 b_k^2 / 2b_0^2 \tag{17}
\]
where \(b_k^2 = \beta_k^i \Phi(t) \beta_k\) is the k-th component of the noiseless output system.

When frequency changes are sufficiently slow, so that the functions \(\dot{\phi}_k(t)\), \(k = 1, \ldots, K\), can be regarded as locally almost periodic, the sequences \(\{ \varepsilon_k(t)\}\) are mutually orthogonal in the sense that
\[
(\rho_{\varepsilon_k}(t))_T \doteq (\rho_{\varepsilon_k}(t)|\omega_0(t) = \omega_0)_{\infty} = 0, \quad \forall k \neq l
\]
where \(x(t)_T = (1/T) \sum_{i=1}^{T} x(t - i)\) denotes the local average of \(x(t)\), and \(T \gg T_0 = 2\pi/\omega_0\). Using this result, one obtains
\[
\{ \sigma^2_{\varepsilon_k}(t) \}_T \doteq \{ E[\varepsilon_k(t)^2] \}_T \doteq \sum_{k=1}^{K} \sigma_k^2 \sigma_{\varepsilon_k}^2 = \sigma_{\varepsilon_k}^2 / 2b_0^2. \tag{18}
\]

Suppose that the assumption (A4) holds true. Then, using standard results from the linear filtering theory, one obtains
\[
E[\{ \Delta \omega_0(t) - \omega_0(t) \}^2]_T = E[\{ \Delta \hat{\omega}_0(t) \}^2]_T = g[G_1(x^{(-1)})] \sigma^2_{\omega_0} \tag{19}
\]
\[
E[\{ \Delta \hat{\omega}_0(t) - \omega_0(t) \}^2]_T = E[\{ \Delta \hat{\omega}_0(t) \}^2]_T = g[H_1(x^{(-1)})] \sigma^2_{\dot{\omega}_0}. \tag{20}
\]
where
\[
g[x(\cdot)] = \frac{1}{2\pi} \int_{-\pi/2}^{\pi/2} x(z^{-1})X(z) \frac{dz}{z}
\]
is an integral evaluated along the unit circle in the z-plane and \(X(z)\) denotes any stable proper rational transfer function.

The first term on the right-hand side of (19) constitutes the variance component of the mean-squared frequency estimation error, and the second term — its bias component. The same remark applies to (20). According to (18), the variance components in (19) and (20) are inversely proportional to the quantity which will be further referred to as effective signal-to-noise ratio (ESNR)
\[
\text{ESNR} = \frac{b_0^2}{\sigma^2_{\omega_0}} = \frac{\sum_{k=1}^{K} m_k^2 b_k^2}{\sigma^2_{\varepsilon_k} / 2b_0^2}
\]
and which differs from the signal-to-noise ratio defined as
\[
\text{SNR} = \frac{E[|\Phi(t)\theta(t)|^2]}{\sigma^2_{\varepsilon_k}} \doteq \frac{\sum_{k=1}^{K} b_k^2}{\sigma^2_{\varepsilon_k}} \leq \text{ESNR}.
\]

We note that Eqs. (13)–(14) are identical with those derived earlier for the single frequency case — the only change needed to move from (8)–(9) to (13)–(14) is replacement of the noiseless output power \(b^2\), which appears in the expression for \(\sigma^2_{\omega_0}\), with the effective output power \(b^2\), which appears in the expression for \(\sigma^2_{\omega_0}\). Since in the multi-frequency case \((K > 1)\) it holds that \(b^2 = \sum_{k=1}^{K} b_k^2\), the variance components of the mean-squared fundamental frequency and frequency rate tracking errors are smaller than the analogous errors observed, under the same SNR, in the single frequency case. This increased (compared to the unconstrained frequency estimation case) accuracy bonus is available due to incorporation in the estimation process prior information about the harmonic structure of the identified quasi-periodic phenomenon.

The same qualitative effect can be observed in time-invariant frequency estimation schemes, such as the ones described in James, Anderson, and Williamson (1994) and Nehorai and Porat (1986).

### 5. Generalized adaptive comb smoother

The important consequence of the fact that the approximate error equations (13)–(14) are identical with those derived in Niedźwiecki and Meller (2011a) for systems with a single frequency mode of parameter variation, is that the smoothing technique proposed there is directly applicable to the multiple-frequency case. Following Niedźwiecki and Meller (2011a), suppose that a pre-recorded data block \(\Omega(N) = [y(i), \phi(i), i = 1, \ldots, N] \) of length \(N\) is available, which is typical of off-line applications, i.e., those based on parameter/signal reconstruction, rather than tracking. The smoothed estimates of \(\omega_0(t), \alpha_0(t), \text{and } \theta(t), \text{based on } \Omega(N), \) will be denoted by \(\hat{\omega}_0(t), \hat{\alpha}_0(t)\) and \(\hat{\theta}(t), \) respectively.\(^3\)

To obtain smoothed estimates, one can use a cascade of postprocessing filters derived in Niedźwiecki and Meller (2011a,b). The proposed fixed-interval generalized adaptive comb smoothing (GACS) procedure, listed in Table 1, is six-step.

**Step 1:** The preliminary estimates \(\hat{\omega}_0(t)\) and \(\hat{\omega}_0(t)\) are obtained using the pilot algorithm based on (12).

**Step 2:** To obtain the smoothed frequency rate estimates \(\hat{\omega}_0(t)\), the trajectory \(\{\hat{\omega}_0(t), t = 1, \ldots, N\}\) is filtered, backward in time, using the anticausal filter \(S(q)\): \(\hat{\omega}_0(t) = S(q)\hat{\omega}_0(t)\), where
\[
S(q) = 1 - (1 - q)H_2(q) = y_\omega q / D(q),
\]

\(^3\) In the Kalman filtering/smoothing literature, the filtered and smoothed estimates of the state vector \(x(t)\) are usually denoted by \(\tilde{x}(t)\) and \(\hat{x}(N)\), respectively. In this paper a different notation is used to avoid false associations.
Table 1

<table>
<thead>
<tr>
<th>Generalized adaptive comb smoother.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pilot filter:</td>
</tr>
<tr>
<td>$e^{\theta(t)} = e^{\alpha(t)}\cos(\omega_0 t) + e^{\alpha(t)}\sin(\omega_0 t)$</td>
</tr>
</tbody>
</table>

\[ \begin{align*} 
T(q^{-1}) &= \frac{a_1 q^{-1}}{1 + a_0 q^{-1}} \quad \Rightarrow \quad a_1 = \frac{\gamma_0}{\gamma_a} \quad a_2 = \frac{\gamma_0 - \gamma_m}{\gamma_m} 
\end{align*} \]

This can be achieved by means of backward-time processing of the prefiltered trajectory: $\tilde{\omega}_0(t) = q^{-1}s(q^{-1})\tilde{\omega}_0(t)$, where $\tilde{\omega}_0(t) = q^T(q^{-1})\omega_0(t)$.

Step 4: The amplitude coefficients are re-estimated using the frequency-guided version of the pilot algorithm, obtained by replacing in (12) the causal frequency estimates $\tilde{\omega}_0(t)$ with their noncausal (smoothed) counterparts $\hat{\omega}_0(t)$, evaluated at Step 3.

Step 5: To obtain the smoothed amplitude estimates $\tilde{\beta}_0(t)$, $k = 1, \ldots, K$, the re-estimated amplitude trajectories $[\tilde{\beta}_0(t), t = 1, \ldots, N]$ are filtered, backward in time, using the anticausal filter $F(q^{-1})$: $\tilde{\beta}_0(t) = F(q^{-1})\tilde{\beta}_0(t)$, $k = 1, \ldots, K$, where $F(q) = \mu/(1 - (1 - \mu)q)$.

Step 6: To obtain the smoothed partial parameter estimates $\tilde{\theta}_0(t)$, the smoothed amplitude estimates $\tilde{\beta}_0(t)$ are combined with the smoothed phase estimates $\tilde{\alpha}_0(t) = \tilde{\alpha}_0 \tilde{\beta}_0(t)$. The smoothed parameter estimate $\hat{\theta}_0(t)$ is evaluated as a sum of its harmonic components.

Remark. We note that Step 2 above is optional — if all that is needed is an estimation of $\theta(t)$, the frequency rate smoothing part of the algorithm can be skipped.

Denote by $\Delta\tilde{\alpha}_0(t) = \omega_0(t) - \tilde{\omega}_0(t)$ and $\Delta\tilde{\beta}_0(t) = \alpha_0(t) - \tilde{\alpha}_0(t)$ the frequency and frequency rate smoothing errors, respectively. Under assumptions (A1)–(A3), the approximate error equations can be obtained in the form

\[ \begin{align*} 
\Delta\tilde{\alpha}_0(t) &\cong I_1(q^{-1})\tilde{\alpha}_0(t) + I_2(q^{-1})\delta(t) 
\Delta\tilde{\beta}_0(t) &\cong J_1(q^{-1})\tilde{\alpha}_0(t) + J_2(q^{-1})\delta(t) 
\end{align*} \]

where

\[ \begin{align*} 
I_1(q^{-1}) &= S(q)T(q^{-1})G_1(q^{-1}) 
I_2(q^{-1}) &= \frac{1 - S(q)S(q^{-1})}{1 - q^{-1}} 
J_1(q^{-1}) &= S(q)H_1(q^{-1}) 
J_2(q^{-1}) &= \frac{1 - S(q)S(q^{-1})}{1 - q^{-1}}. 
\end{align*} \]

6. Optimization and Cramér–Rao bounds

Consider a system (1)–(3) with pseudo-linear frequency changes. In order to achieve the best tracking/smoothing results, the adaptation gains of the GACF/GACS algorithms should be chosen so as to trade-off the bias and variance components in (19)–(20) and (21)–(22). Such optimal settings depend exclusively on the balance between the bias and variance error components, determined by the scalar coefficient

\[ \kappa = \frac{E[w^2(t)]}{2E[\tilde{\omega}_0^2(t)]]} \approx \frac{b_0^2\sigma^2}{\sigma^2} = \text{SNR} \cdot \sigma_i^2 \]

further referred to as the rate of nonstationarity of the analyzed system (in signal analysis a similar concept was introduced earlier in Tichavský and Händel [1995]).

Using residue calculus [ Jury, 1964], one can easily derive analytical expressions quantifying the mean-squared estimation errors in terms of $\mu$, $\gamma_0$, and $\gamma_m$. Unfortunately, these expressions (not listed here) are too complicated to enable minimization of the MSE scores in an explicit, analytical form. For this reason the optimal values of adaptation gains were searched numerically. The optimal settings that minimize the frequency and frequency rate, tracking and smoothing, errors were found to be in all four cases identical — the corresponding values, obtained for several nonstationarity rates $\kappa$, are listed in Table 2.

Our next step was to establish the Cramér–Rao-type lower tracking bounds (LBTs, LBS) and lower smoothing bounds (LSBs, LBSs), which set the upper limits for the frequency and frequency rate estimation accuracy using any causal/noncausal identification algorithm — see Table 2. Note the performance gains that can be achieved when tracking is replaced with smoothing. Even though such analysis has mainly a theoretical value, it allows one to evaluate tracking/smoothing performance of the proposed algorithms in absolute, rather than relative, terms. The LTBs and LSBs for the system governed by (1)–(3) were obtained under assumptions (A1)–(A5) and, to make the analysis easier, under
some technical constraints imposed on the initial conditions — see Appendix C. Again, rather than providing the closed-form analytical formulas, we show how LTBs and LSBs can be established numerically for a given value of $\kappa$.

It was found out that there is a perfect agreement between the lower tracking and smoothing bounds and the MSE values obtained by minimizing (19)–(20) and (21)–(22), respectively — in some cases the computed values agreed up to the sixth decimal place. This means that, at least theoretically, the optimally tuned GACF/GACS algorithms should be statistically efficient frequency and frequency rate trackers/smoothers. In the next section we will verify this statement using computer simulations.

According to the results given above, in the constant amplitude case the frequency estimates are unambiguous — in spite of the fact that the corresponding amplitude estimates are complex-valued quantities and, as such, could potentially create nonidentifiability problems. Although the same was also observed in the time-varying-amplitude case, some caution is needed in interpreting the quantities $\omega_0(t), \delta_0(t)$ and $\beta_0(t), \tilde{\beta}_0(t)$ unless identifiability is formally proved. Note, however, that this potential nonidentifiability problem does not extend to estimation of $\theta(t)$, which is our main interest here.

7. Simulation and experimental results

7.1. ALF-based analysis

To check the validity of the analytical expressions (19)–(20), based on the approximating linear filter equations (13)–(14), the following two-tap FIR system (inspired by channel equalization applications) was simulated

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

where $u(t)$ denotes a white 4-QAM [quadrature amplitude modulation — see e.g. Giannakis and Tepedelenliolu (1998)] input sequence ($u(t) = \pm 1, \sigma_u^2 = 2$) and $\nu(t)$ denotes a complex-valued Gaussian measurement noise.

Each of $n = 2$ impulse response coefficients had $K = 4$ modes of variation — system parameters varied according to

$$\theta(t) = \begin{bmatrix} \theta_1(t) \\ \theta_2(t) \end{bmatrix} = \sum_{k=1}^{4} \beta_{k0} e^{i\omega_0 \sigma_k^2 \sum_{t=1}^{r} \omega_0(t)}$$

$$\beta_{10} = [2 - j, 1 + 2j]^T, \quad \beta_{20} = [1 - j, 1]^T, \quad \beta_{30} = [1, 1]^T, \quad \beta_{40} = [0.5, 0.5]^T$$

where $m_k = k$, the phase shifts $\nu_1, \ldots, \nu_4$ were drawn independently from the uniform distribution on $[0, 2\pi)$, and the fundamental frequency $\omega_0(t)$ was governed by the integrated random-walk model [obeysing (A4) and (A5)], starting from $\omega_0(0) = \pi/4$. Note that in this case $\phi(t) = |u(t), u(t-1)|^2$. $\phi = 12\sigma_u^2$, $b_1 = 20$, $b_2 = 6$, $b_3 = 4$, $b_4 = 1$ and $b_5 = 96$ (since $\sum_{k=1}^{4} b_k^2 = 31$), there was a noticeable discrepancy between SNR and effective SNR.

The approximations (19)–(20) were checked for 3 values of the signal-to-noise ratio: $\text{SNR} = 0 \text{~dB}$ ($\sigma_v = 5.5678$), $\text{SNR} = 10 \text{~dB}$ ($\sigma_v = 1.7607$) and $\text{SNR} = 20 \text{~dB}$ ($\sigma_v = 0.5568$), for 2 values of the nonstationarity rate: $\kappa = 10^{-10}$ and $\kappa = 10^{-1}$, and for 10 values of the adaptation gain $\mu$, ranging from 0.01 to 0.1. To reduce the number of design degrees of freedom, the two other gains adopted for GACF/GACS algorithms were set to: $\gamma_0 = \mu^2/2$ and $\gamma_\alpha = \mu^2/8$ — in agreement with the general tendency observed, under different levels of effective SNR, for the optimal settings (this rule of thumb was found to work quite well in practice). The mean-squared frequency and frequency rate estimation errors were evaluated (for the optimally tuned GACS 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 1000 iterations of the GACS filter (after the algorithm had reached its steady-state). The obtained results were next averaged over 50 realizations of $\phi(t), \nu(t)$ and $\nu_1, \ldots, \nu_4$.

Figs. 1 and 2 show comparison of theoretical curves and the time-averaged values of the mean-squared frequency and frequency rate estimation errors obtained via simulation. Note the good agreement between theoretical evaluations and the actual algorithm’s performance, which can be observed for SNR $\geq 10$ dB. Generally, the degree of fit improves with decreasing $\kappa$ and increasing SNR, which is consistent with the operating range of the approximating linear filter technique. Similar results, not reported here, were obtained for the GACS algorithm.

7.2. Statistical efficiency

Fig. 3 shows a comparison of the theoretical values of the lower frequency tracking bound LTB$_{\alpha,0}$ and the lower frequency rate tracking bound LTB$_{\alpha,0}$ with experimental results obtained for an optimally tuned GACS algorithm designed for the system described in Section 7.1. In agreement with the results of theoretical analysis presented in Section 6, for small rates of system nonstationarity the proposed GACS algorithm is statistically efficient, i.e., under the conditions specified earlier, it cannot be outperformed by any other tracking algorithm. The same conclusion can be drawn after inspection of the plots shown in Fig. 4, illustrating the behavior of the optimally tuned GACS algorithm.

7.3. Performance

The aim of this simulation experiment was to compare performance of the proposed GACF/GACS algorithms with that yielded by the unconstrained multiple-frequency versions of the GANF/GANS algorithms.

The simulated two-tap linear system (25) was governed by (see Figs. 5 and 6)

$$\theta(t) = \begin{bmatrix} \theta_1(t) \\ \theta_2(t) \end{bmatrix} = \sum_{k=1}^{4} \beta_{k0} e^{i\omega_0 \sigma_k^2 \sum_{t=1}^{r} \omega_0(t)}$$

Table 2

<table>
<thead>
<tr>
<th>$\kappa$</th>
<th>$\mu^\text{opt}$</th>
<th>$\gamma_0^\text{opt}$</th>
<th>$\gamma_\alpha^\text{opt}$</th>
<th>LTB$_{\alpha,0}/\sigma_v^2$</th>
<th>LTB$_{\alpha,0}/\sigma_v^2$</th>
<th>LSB$_{\alpha,0}/\sigma_v^2$</th>
<th>LSB$_{\alpha,0}/\sigma_v^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-10}$</td>
<td>0.0472</td>
<td>0.00113</td>
<td>0.0000138</td>
<td>2.05 $10^4$</td>
<td>8.21 $10^4$</td>
<td>1.18 $10^4$</td>
<td>1.38 $10^4$</td>
</tr>
<tr>
<td>$5 \times 10^{-10}$</td>
<td>0.0613</td>
<td>0.00192</td>
<td>0.000306</td>
<td>9.09 $10^4$</td>
<td>6.28 $10^4$</td>
<td>5.28 $10^3$</td>
<td>1.05 $10^4$</td>
</tr>
<tr>
<td>$5 \times 10^{-9}$</td>
<td>0.0685</td>
<td>0.00241</td>
<td>0.000432</td>
<td>6.39 $10^4$</td>
<td>5.58 $10^4$</td>
<td>3.73 $10^3$</td>
<td>9.39</td>
</tr>
<tr>
<td>$5 \times 10^{-8}$</td>
<td>0.0886</td>
<td>0.00407</td>
<td>0.000955</td>
<td>2.82 $10^4$</td>
<td>4.26 $10^4$</td>
<td>1.67 $10^3$</td>
<td>7.18</td>
</tr>
<tr>
<td>$5 \times 10^{-7}$</td>
<td>0.0950</td>
<td>0.00509</td>
<td>0.00134</td>
<td>1.97 $10^4$</td>
<td>3.79 $10^4$</td>
<td>1.18 $10^3$</td>
<td>6.40</td>
</tr>
<tr>
<td>$5 \times 10^{-6}$</td>
<td>0.127</td>
<td>0.00852</td>
<td>0.00295</td>
<td>8.66 $10^3$</td>
<td>2.89 $10^4$</td>
<td>5.28 $10^3$</td>
<td>4.89</td>
</tr>
<tr>
<td>$5 \times 10^{-5}$</td>
<td>0.142</td>
<td>0.0106</td>
<td>0.00414</td>
<td>6.06 $10^3$</td>
<td>2.57 $10^4$</td>
<td>3.73 $10^3$</td>
<td>4.36</td>
</tr>
<tr>
<td>$5 \times 10^{-4}$</td>
<td>0.181</td>
<td>0.0177</td>
<td>0.00905</td>
<td>2.63 $10^3$</td>
<td>1.05 $10^4$</td>
<td>1.67 $10^3$</td>
<td>3.33</td>
</tr>
<tr>
<td>$5 \times 10^{-3}$</td>
<td>0.201</td>
<td>0.0219</td>
<td>0.0126</td>
<td>1.83 $10^3$</td>
<td>1.73 $10^4$</td>
<td>1.18 $10^3$</td>
<td>2.96</td>
</tr>
</tbody>
</table>
7.4. Estimation of MRI noise

Magnetic resonance imaging (MRI) equipment is used to visualize internal structures of the human body without exposing subjects to harmful radiation. It is utilized in many medical institutions for diagnostic purposes and, quite recently, as an aid during some operations — in the latter case it works in the nearly real-time mode (Kurumi et al., 2007).

MRI devices generate very loud harmonic noise (with intensity exceeding 100 dB) caused by vibration — owing to the Lorentz force — of the gradient coil. Exposure to this noise is very annoying both for the patients and for the medical staff. MRI noise can be reduced using the active noise control (ANC) techniques, the better the more accurately one can track the underlying multi-harmonic signal ($K > 30$). Since this signal is nonstationary — both the amplitudes and the fundamental frequency change over time — its estimation is a challenging task. Fig. 8 shows the time plots and periodograms of the original MRI noise (recorded in the axial mode), as well as the time plots and periodograms of prediction errors $\varepsilon(t) = y(t) - 5(t)(t - 1) = y(t) - \sum_{k=1}^{9} f_k(t) \delta_k(t - 1)$ yielded by the proposed adaptive comb filter (ACF) and by the multiple-frequency version of the adaptive notch filter (ANF). All results were obtained for $\mu = 0.01$, $\gamma_u = \mu^2/2$ and $\gamma_v = \mu^3/8$. The complex-valued version of the MRI signal was obtained using the discrete Hilbert transform.

While the time plots obtained for ANF and ACF look similar, the corresponding periodograms differ significantly. It is clear that the ACF is much more effective in suppressing signal harmonics than ANF, even though both algorithms used the same starting values. The failure of the ANF algorithm can be explained by its poor frequency matching capabilities — after some initial period...
the algorithm locks on dominant harmonics, leaving the remaining ones unattenuated.

8. Conclusion

The problem of identification of linear stochastic systems with quasi-harmonically varying parameters was considered. Both causal and noncausal identification algorithms were derived, referred to as generalized adaptive comb filters (GACFs) and generalized adaptive comb smoothers (GACSs), respectively. In both cases the frequency and frequency rate estimation properties of the proposed algorithms were analyzed using the method of approximating linear filter. It was shown, and later confirmed by means of computer simulations, that when the fundamental frequency of parameter changes varies slowly with time according to the integrated random-walk model, the optimally tuned GACF/GACS algorithms are (under Gaussian assumptions) statistically efficient frequency and frequency rate trackers/smoothers, i.e., they reach the Cramér–Rao-type lower tracking/smoothing bounds — expressions allowing one to evaluate these bounds were also derived in the paper.

Appendix A. Derivation of (13) and (14)

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

Fig. 3. 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 system with quasi-linear frequency changes for 3 different SNR values: \( \text{SNR} = 0 \, \text{dB} \) (○), \( \text{SNR} = 10 \, \text{dB} \) (×), \( \text{SNR} = 20 \, \text{dB} \) (+), and 9 different values of the rate of system nonstationarity \( \kappa \).

Fig. 4. 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 system with quasi-linear frequency changes, for 3 different SNR values: \( \text{SNR} = 0 \, \text{dB} \) (○), \( \text{SNR} = 10 \, \text{dB} \) (×), \( \text{SNR} = 20 \, \text{dB} \) (+), and 9 different values of the rate of system nonstationarity \( \kappa \).

Fig. 5. Real parts (solid lines) and imaginary parts (broken lines) of system parameters observed in a short time interval.

Fig. 6. Evolution of the instantaneous fundamental frequency.
To derive recursion for $\Delta \hat{x}(t) = \sum_{k=1}^{K} \Delta \hat{x}_k(t)$, note that
\[ \hat{\theta}_k(t) = \zeta_k(t) + \mu \Phi^{-1} \varphi^*(t) \varepsilon(t) \]
\[ \varepsilon(t) = \varphi^T(t) \sum_{k=1}^{K} \theta_k(t) + v(t) - \sum_{k=1}^{K} \xi_k(t) \]
where \[ \zeta_k(t) = e^{j \omega_k(t-1)} \hat{\omega}_k(t-1) \] and \[ \hat{\omega}_k(t) = m_k \hat{\omega}_0(t) \]
\[ \hat{\omega}_k(t) = m_k \hat{\omega}_0(t). \]
Therefore
\[ \hat{\theta}_k(t) = \left[ 1 - \mu \Phi^{-1} \varphi^*(t) \varphi(t) \right] \xi_k(t) \]
\[ + \mu \Phi^{-1} \varphi^*(t) \varphi^T(t) \theta_k(t) + \mu \Phi^{-1} \varphi^*(t) v(t) + \xi_k(t) \]
where
\[ \xi_k(t) = \mu \Phi^{-1} \varphi^*(t) \varphi(t) \sum_{i=1}^{K} \left[ \theta_i(t) - \xi_i(t) \right]. \]
Since in the case considered $\theta_k(t) = e^{j \omega_k(t)} \theta_k(t - 1)$, the last relationship leads to
\[ \Delta \hat{\theta}_k(t) = \left[ 1 - \mu \Phi^{-1} \varphi^*(t) \varphi(t) \right] \theta_k(t) \]
\[ - \left[ 1 - \mu \Phi^{-1} \varphi^*(t) \varphi^T(t) \right] \xi_k(t) \]
\[ - \mu \Phi^{-1} \varphi^*(t) v(t) - \xi_k(t). \] (27)

Note that $\xi_k(t)$ can be rewritten in the form $\xi_k(t) = e^{j \omega_k(t)} e^{-j \Delta \omega_k(t-1)} \theta_k(t - 1) - \Delta \hat{\theta}_k(t - 1)$. Using the following approximations $e^{-j \Delta \omega_k(t-1)} \approx 1 - j \Delta \omega_k(t-1)$, $e^{-j \Delta \omega_k(t-1)} \approx 1 - j \Delta \omega_k(t-1)$, that hold for small frequency and frequency rate errors, respectively, and applying ALF rules, one arrives at
\[ \xi_k(t) = e^{j \omega_k(t)} e^{-j \Delta \omega_k(t-1)} e^{-j \Delta \omega_k(t-1)} \]
\[ \times \left[ \theta_k(t - 1) - \Delta \hat{\theta}_k(t - 1) \right] \]
\[ \approx e^{j \omega_k(t)} \left[ 1 - j \Delta \omega_k(t - 1) \right] \left[ \theta_k(t - 1) - \Delta \hat{\theta}_k(t - 1) \right] \]
\[ \times \left[ \theta_k(t - 1) - \Delta \hat{\theta}_k(t - 1) \right] \]
\[ \approx \xi_k(t) - e^{j \omega_k(t)} \Delta \hat{\theta}_k(t - 1) \]
\[ - j \Delta \omega_k(t - 1) + \Delta \omega_k(t - 1) \theta_k(t). \] (28)

Combining (27) with (28), one obtains
\[ \Delta \hat{\theta}_k(t) \approx \left[ 1 - \mu \Phi^{-1} \varphi^*(t) \varphi(t) \right] e^{j \omega_k(t)} \Delta \hat{\theta}_k(t - 1) \]
\[ + j \left[ 1 - \mu \Phi^{-1} \varphi^*(t) \varphi(t) \right] \left[ \Delta \omega_k(t - 1) \right] \]
\[ + \Delta \omega_k(t - 1) \theta_k(t - 1) - \mu \Phi^{-1} \varphi^*(t) v(t) - \xi_k(t). \] (29)

Let $\Delta \hat{\theta}_k(t) = \Delta \hat{\theta}_k(t)_{\hat{x}_k}(t)$. After multiplying both sides of (29) with $f^*_k(t)$, one arrives at
\[ \Delta \hat{\theta}_k(t) \approx \left[ 1 - \mu \Phi^{-1} \varphi^*(t) \varphi(t) \right] e^{j \omega_k(t)} \Delta \hat{\theta}_k(t - 1) \]
\[ + j \left[ 1 - \mu \Phi^{-1} \varphi^*(t) \varphi(t) \right] \left[ \Delta \omega_k(t - 1) \right] \]
\[ + \Delta \omega_k(t - 1) \theta_k(t - 1) - \mu \Phi^{-1} \varphi^*(t) f^*_k(t) v(t) \]
\[ + \Delta \omega_k(t - 1) \theta_k(t - 1) - \mu \Phi^{-1} \varphi^*(t) f^*_k(t) v(t). \]
\[ -\xi_k^*(t) f_k^*(t). \]  

(30)

For small values of adaptation gains \( \mu, \gamma_0 \) and \( \gamma_0 \), the quantities \( \Delta \theta_k(t), \Delta \omega_k(t) \) and \( \Delta \omega_k(t) \) change slowly compared to \( \phi(t) \) and \( f_k(t) \), \( k = 1, \ldots, K \), i.e., (30) can be regarded as a two-scale difference equation. When solving such an equation for slowly varying quantities, one is allowed to replace some functionals of the fast varying quantities with their time averages. This is usually referred to as a deterministic averaging technique.

Denote by \( \langle x(t) \rangle_t = (1/T) \sum_{t=0}^{T} x(t - i) \) the local average of \( x(t) \), and by \( \langle x(t) \rangle_{\infty} = \lim_{t \to \infty} \langle x(t) \rangle_t \) the corresponding limiting value (provided it exists).

We will exploit the fact that for any process \( \langle \phi(t) \rangle \) obeying (A2), and for sufficiently large values of \( T \), it holds that \( \langle \phi(t) \phi(t)^\dagger \rangle_{\infty} \approx \langle \phi(t) \phi(t)^\dagger \rangle_{\infty} \approx 0 \), \( \forall \alpha, \beta \not= 0 \). According to the first relation, the data-dependent matrix \( I - \mu \Phi^{-1} \phi(t) \phi(t)^\dagger \) in (30) can be replaced with \( I - \mu \Phi^{-1} \phi(t) \phi(t)^\dagger \) \( \approx 0 \). Similarly, according to the second relation, since for sufficiently slow frequencies \( f_k(t), f_k(t), k = 1, \ldots, K \), are locally almost periodic functions of time, it holds that \( \langle \xi_k(t) \xi_k(t)^\dagger \rangle_{\infty} \approx 0 \), allowing one to neglect the last term on the right hand side of (30). Hence, using the averaging technique, one arrives at the following approximate error equation

\[ \Delta \theta_k(t) \approx (1 - \mu) \Delta \theta_k(t - 1) + j(1 - \mu) \Delta \theta_k(t - 1) \]

(31)

Multiplying both sides of (31) by \( \beta_k^I \Phi \), one obtains

\[ \beta_k^I \Phi \Delta \theta_k(t) \approx \lambda \beta_k^I \Phi \Delta \theta_k(t - 1) \]

\[ + j \mu [\Delta \theta_k(t - 1) + \Delta \theta_k(t - 1)] \beta_k^I \Phi \beta_k^I \]

\[ - \mu \beta_k^I \Phi \phi(t) \phi(t)^\dagger \nu(t) \]

(32)

where \( \lambda = 1 - \mu \). Finally, dividing both sides of Eq. (32) by \( b_k^2 \), taking imaginary parts, and noting that \( \beta_k^I \Phi \Delta \theta_k(t) = \theta_k(t)^I \Phi \Delta \theta_k(t) \) and \( \beta_k^I \Phi \phi(t)^\dagger \nu(t) = \theta_k(t)^I \phi(t)^\dagger \nu(t) \), one arrives at

\[ \Delta \theta_k(t) \approx \lambda \Delta \theta_k(t - 1) \]

\[ + \frac{\lambda b_k^2}{b_k^2} \]

\[ \times [\Delta \theta_k(t - 1) + \Delta \theta_k(t - 1)] \mu e_k(t) \]

where \( e_k(t) = -\text{Im} \left\{ \beta_k^I \Phi \phi(t) \phi(t)^\dagger \nu(t) / b_k^2 \right\} \). Note that \( \Delta \theta_k(t) = \sum_{k=1}^{K} m_k \Delta \theta_k(t), e_k(t) = \sum_{k=1}^{K} m_k e_k(t) \) and \( b_k^2 = \sum_{k=1}^{K} m_k^2 b_k^2 \). Hence, after incorporating (26), one obtains

\[ \Delta \theta_k(t) \approx \lambda \Delta \theta_k(t) + \mu e_k(t). \]

(33)

To derive recursions for \( \Delta \omega_0(t) \) and \( \Delta \omega_0(t) \), note that in the tracking mode it holds that

\[ g(t) = \sum_{k=1}^{K} m_k \text{Im} \left\{ \phi^* \phi(t) \phi(t)^\dagger \phi(t)^\dagger \right\} \]

\[ + \sum_{k=1}^{K} m_k \text{Im} \left\{ \phi^* \phi(t) \phi(t)^\dagger \phi(t)^\dagger \right\} \]

\[ \approx \sum_{k=1}^{K} m_k \text{Im} \left\{ \phi^* \phi(t) \phi(t)^\dagger \phi(t)^\dagger \right\} \]

\[ \approx \sum_{k=1}^{K} m_k \text{Im} \left\{ \phi^* \phi(t) \phi(t)^\dagger \phi(t)^\dagger \right\} \]

\[ \approx \sum_{k=1}^{K} m_k \text{Im} \left\{ \phi^* \phi(t) \phi(t)^\dagger \phi(t)^\dagger \right\} \]

\[ \approx \sum_{k=1}^{K} m_k \text{Im} \left\{ \phi^* \phi(t) \phi(t)^\dagger \phi(t)^\dagger \right\} \]

\[ \approx \sum_{k=1}^{K} m_k \text{Im} \left\{ \phi^* \phi(t) \phi(t)^\dagger \phi(t)^\dagger \right\} \]

(34)

and

\[ \varepsilon(t) = \phi^* \phi(t) \theta_k(t) - \phi^* \phi(t) \xi_k(t) + \phi^* \phi(t) \psi_k(t) + v(t) \]

\[ \psi_k(t) = \sum_{i=k}^{K} \theta_i(t) + \xi_i(t). \]

Furthermore, \( \text{Im} \left\{ e^* \phi(t) \phi(t) \xi_k(t) \right\} = \text{Im} \left\{ e^* \phi(t) \phi(t) \xi_k(t) \right\} \]

(35)

Using (28) and applying the ALF rules, one obtains the following approximation

\[ \theta_k^I(t) \phi^*(t) \phi(t) \xi_k(t) \xi_k(t) \approx \theta_k^I(t) \phi^*(t) \phi(t) \xi_k(t) \xi_k(t) \]

\[ - \theta_k^I(t - 1) \phi^*(t) \phi(t) \xi_k(t) \xi_k(t) \]

\[ - j \theta_k^I(t) \phi^*(t) \phi(t) \xi_k(t) \xi_k(t) \]

\[ = \beta_k^I \phi^*(t) \phi(t) \xi_k(t) \xi_k(t) \]

\[ - \beta_k^I \phi^*(t) \phi(t) \xi_k(t) \xi_k(t) \]

\[ \approx \beta_k^I \phi^*(t) \phi(t) \xi_k(t) \xi_k(t) \]

\[ - \beta_k^I \phi^*(t) \phi(t) \xi_k(t) \xi_k(t) \]

which, after averaging, leads to

\[ \Delta \omega_0(t) \approx \Delta \omega_0(t - 1) + \delta(t) + \nu_0(t) \]

\[ \Delta \omega_0(t) \approx \Delta \omega_0(t - 1) + \nu_0(t) \]

Note that

\[ \Delta \omega_0(t) \approx \Delta \omega_0(t - 1) + \nu_0(t) \]

Combining the last three equations, one arrives at

\[ \Delta \omega_0(t) \approx (1 - \gamma_0) \Delta \omega_0(t - 1) + \delta(t) + \nu_0(t) \]

\[ - \gamma_0 \Delta \omega_0(t - 1) + \gamma_0 \Delta \omega_0(t - 1) \]

(36)

Finally, solving the set of linear equations (33)-(35) for \( \Delta \omega_0(t) \) and \( \Delta \omega_0(t) \), one obtains (13) and (14), respectively.

Appendix B. Derivation of (16)

The relationship \( E \left\{ e_k(t) e_k(t) \right\} = 0 \), \( \forall k \not= l \) stems from the fact that \( \nu(t) \) is a sequence of zero-mean independent random variables, independent of \( \phi(t) \). To arrive at the expression for \( \rho_k(t) \) we will introduce the following notation

\[ \nu(t) = \nu(t) + j \nu(t) \]

\[ \eta_k(t) = \beta_k^I \phi(t) = \eta_k(t) + j \eta_k(t) \]

where the subscripts R/I denote the real/imaginary part of a complex variable.

Note that

\[ e_k(t) = -\text{Im} \left\{ \eta_k^I(t) e^{-j \phi_k(t)} / b_k^2 \right\} \]

\[ = \frac{1}{b_k^2} \left\{ \eta_k^R(t) \eta_k^I(t) + \eta_k^I(t) \eta_k^I(t) \right\} \sin \phi_k(t) \]

\[ - \eta_k^L(t) \eta_k^I(t) \cos \phi_k(t) \]

Since the sequence \( \nu(t) \) is circular, it holds that \( E \left\{ \nu_k^2(t) \right\} = E \left\{ \nu_k^2(t) \right\} = \sigma^2 \left( \frac{1}{2} \right) \) and \( E \left\{ \nu_k^2(t) \right\} = 0 \). Using these relationships, and the fact that the process \( \nu(t) \) is independent of \( \phi(t) \), one arrives at

\[ E \left\{ e_k(t) e_k(t) \right\} \]

\[ \sigma^2 = 2b_k \left\{ \left( \eta_k^R(t) \eta_k^I(t) + \eta_k^I(t) \eta_k^I(t) \right) \sin \phi_k(t) \right\} \]

\[ + \left( \eta_k^L(t) \eta_k^I(t) + \eta_k^I(t) \eta_k^I(t) \right) \sin \phi_k(t) \right\} \]

(37)
After elementary but tedious calculations, one can show that
\[ E\{n_{\text{tr}}(t)n_{\text{tr}}(t) + n_{\text{dr}}(t)n_{\text{dr}}(t)\} = \text{Re}\{\beta^*_k \Phi \beta_k\} \]
\[ E\{n_{\text{tr}}(t)n_{\text{dr}}(t) + n_{\text{dr}}(t)n_{\text{tr}}(t)\} = -\text{Im}\{\beta^*_k \Phi \beta_k\}. \]
Finally, combining (36) with (37), one obtains
\[ E[e_t(t)e_t(t)] = E\{E[e_t(t)e_t(t)]\} = \frac{\sigma^2}{2\rho}[\text{Re}\{\beta^*_k \Phi \beta_k\}\cos(\phi_k(t) - \phi(t)) \]
- \text{Im}\{\beta^*_k \Phi \beta_k\}\sin(\phi_k(t) - \phi(t))] = \rho_\text{v}(t).

**Appendix C. 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 (LTBs) and lower smoothing bounds (LSBs) belong to the class of posterior (or Bayesian) Cramér–Rao bounds, applicable to signals/systems with random parameters.\(^4\)

Denote by \(u\) and \(y\) the vectors of system inputs (regarded as a known deterministic sequence, e.g., a particular realization of a stochastic process), noisy outputs, and fixed initial conditions, respectively, and let \(\hat{x}(y, u, c_0)\) be an estimator of a real-valued random parameter vector \(x\) based on \((y, u, c_0)\). Then, under weak regularity conditions, one can show (van Trees, 1968)

\[ E[\tilde{\phi}(y, u, c_0) - x)(\tilde{\phi}(y, u, c_0) - x)^T[u, c_0]| \geq J^{-1}(u, c_0) \]
where
\[ J(u, c_0) = -E\left[\frac{\partial^2 \log p(y, u, c_0)}{\partial y \partial x} \right] \]
\[ = E\left[\frac{\partial \log p(y, u, c_0)}{\partial y} \cdot \frac{\partial \log p(y, u, c_0)}{\partial x} \right] \]
and \(p(y, u, c_0) = p(y|u, c_0)p(u)\) is the joint probability density function of the pair \((y, u)\) given \(u\) and \(c_0\).

When the input signal is a stochastic process, initial conditions are random, and averaging is extended to all realizations of \(u\) and \(c_0\), one obtains the following result
\[ E[\tilde{\phi}(y, u, c_0) - x)(\tilde{\phi}(y, u, c_0) - x)^T| \geq E[J^{-1}(u, c_0)] \]
\[ \geq E[J[u, c_0]]^{-1} = \tilde{J}^{-1} \]
where the second transition stems from the Jensen’s inequality for matrices — see Olkin and Pratt (1958).

In the case considered, let \(x_t = [\alpha_0(1), \ldots, \alpha_0(t)]^T, y_t = [y(1), \ldots, y(t)]\) and \(u_t = [\nu(1), \ldots, \nu(t)]\).

To simplify further analysis, we will assume, in addition to (A1)-(A5), that the complex-valued ‘amplitudes’ can be written down in the form \(\beta_k = \beta_k e^{i\phi_k}, k = 1, 2, \ldots, K\), where \(\beta_0\) is fixed (deterministic) complex-valued vectors and \(\nu_k\) are mutually independent random phase shifts, distributed uniformly over the interval \([0, 2\pi]\). Note that under the last assumption it holds that \(E[\beta_k \beta_k^*] = \mathbf{0}\). Furthermore, we will assume that \(\alpha_0(0)\) is uniformly distributed over \([\alpha_{\text{min}}, \alpha_{\text{max}}]\), and that \(\omega_0(0)\) is a known deterministic quantity. Hence, the vector of initial conditions can be specified as \(c_0 = [\beta_0, \ldots, \beta_K, \nu_1, \ldots, \nu_K, \alpha_0(0), \omega_0(0)]\),

where the quantities \(\beta_0, \ldots, \beta_K\) and \(\omega_0(0)\) are deterministic, and the quantities \(\nu_1, \ldots, \nu_K\) and \(\alpha_0(0)\) are stochastic.

First of all, note that
\[ \log p(y, x|u, c_0) = \log p(y|u, c_0) + \log p(x|c_0). \]

Since, under the assumptions listed above, the vectors \(x_t\) and \(c_0\) fully determine \(\theta(t), \tau = 1, \ldots, \tau: \theta(t) = \sum_{k=1}^{K} \beta_k e^{i\nu_k\omega(t)}\),

\[ \phi_0(t) = c_1 + \sum_{n=1}^{K} \alpha_0(n) + \sum_{n=1}^{K} \sum_{m=1}^{K} \alpha_0(n) \]

one arrives at (for normally distributed \(\{v(t)\}\))
\[ \log p(y|u, c_0) = \log p[y|\theta(1), \nu(1), \ldots, \theta(t), \nu(t)] \]
\[ = c_1 - \frac{1}{\sigma^2} \sum_{t=1}^{K} \left| v(t) \right|^2 \]
\[ = - \frac{1}{\sigma^2} \sum_{t=1}^{K} \left| y(t) - \nu(t) \omega(t) \theta(t) \right|^2 \]
where \(c_1\) is a constant independent of \(x_t\).

Differentiating (39) with respect to \(\alpha_0(m)\), one obtains
\[ \frac{\partial \log p(y|x, u, c_0)}{\partial \alpha_0(m)} = \frac{2}{\sigma^2} \sum_{t=1}^{K} \left[ \frac{\partial \theta(t)}{\partial \alpha_0(m)} \right] \]
\[ \frac{\partial \log p(y|x, u, c_0)}{\partial \alpha_0(0)} = \frac{2}{\sigma^2} \sum_{t=1}^{K} \left[ \frac{\partial \theta(t)}{\partial \alpha_0(0)} \right] \]
\[ + \nu(t) \omega(t) \left[ \frac{\partial \theta(t)}{\partial \alpha_0(0)} \right] \]
where the last transition stems from \(y(t) = \nu(t) \omega(t) + \nu(t)\).

This leads to
\[ \phi_{\text{v}}(t) = \frac{\partial \theta(t)}{\partial \alpha_0(0)} \]
\[ \phi_{\text{v}}(t) = \frac{\partial \theta(t)}{\partial \alpha_0(0)} \]
\[ \frac{\partial \theta(t)}{\partial \alpha_0(m)} \]
\[ = - \frac{2}{\sigma^2} \sum_{t=1}^{K} \left[ \frac{\partial \theta(t)}{\partial \alpha_0(m)} \right] \]
\[ = - \frac{2}{\sigma^2} \sum_{t=1}^{K} \left[ \frac{\partial \theta(t)}{\partial \alpha_0(0)} \right] \]
\[ \frac{\partial \theta(t)}{\partial \alpha_0(0)} \]

---

4 When the estimated quantities are stochastic variables, rather than unknown deterministic constants, the classical Cramér–Rao inequality does not apply.
Using (38), one arrives at

$$
\frac{\partial \theta(\tau)}{\partial \alpha_0(m)} = \sum_{k=1}^{K} m_k \beta_k e^{i \phi_k(\tau)} \sum_{l=1}^{K} \delta_{r,l} \max(\tau - m, 0)
$$

where $$\delta_{r,l} = 0$$ if $$r \neq l$$, 1 if $$r = l$$ denotes the Kronecker delta. This allows one to reach

$$
E \left[ \frac{\partial^2 \log p(\theta)}{\partial \theta^2} \right] = \frac{2b_0^2}{\sigma^2} A_t
$$

where $$[A_t]_{mn} = \sum_{r=1}^{K} \max(\tau - m, 0) \cdot \max(\tau - n, 0)$$.

In an analogous way, one can derive the second component of the generalized Fisher matrix. First, note that

$$
\log p(x_t) = \log p(\alpha_0(1), \alpha_0(2) - \alpha_0(1), \ldots, \alpha_0(t) - \alpha_0(t - 1))
$$

where $$c_2 = \log(1/(\alpha_{\text{max}} - \alpha_{\text{min}}))$$ and $$c_3$$ are constants independent of $$x_t$$.

Differentiation of (41) results in

$$
\frac{\partial \log p(x_t)}{\partial x_t} = \frac{1}{\alpha_0(m)} \frac{\partial}{\partial \alpha_0(m)} [\alpha_0(t) - \alpha_0(t - 1)]^2
$$

which leads to

$$
E \left[ \frac{\partial \log p(x_t)}{\partial x_t} \frac{\partial \log p(x_t)}{\partial \theta} \right] = \frac{1}{\sigma^2} B_t
$$

where

$$
B_t = \begin{bmatrix}
1 & -1 & 0 & \cdots & 0 \\
-1 & 2 & -1 & \cdots & 0 \\
0 & -1 & 2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & -1 & 2 & -1 \\
0 & \cdots & 0 & 0 & -1 & 1
\end{bmatrix}
$$

Combining (40) with (42), one obtains

$$
\frac{\partial}{\partial \alpha}(41) = \frac{2b_0^2}{\sigma^2} A_t + \frac{1}{\sigma^2} B_t = \frac{1}{\sigma^2} [2x_t A_t + B_t].
$$

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

$$
\text{LTB}_{\alpha_0} = \lim_{n \to \infty} \inf_{\{\alpha_0(t) - \bar{\alpha}_0(t)\}^2} = \lim_{n \to \infty} \inf_{\{\alpha_0(t) - \bar{\alpha}_0(t)\}^2} = \lim_{n \to \infty} [E^{1/2}]_{\bar{\alpha}_0(t)}
$$

where $$b_1^T = \{1, \ldots, 0\}$$, and 1, denotes the vector of ones of length t. The analogous expressions for lower smoothing bounds read

$$
\text{LSB}_{\alpha_0} = \lim_{n \to \infty} \inf_{\{\alpha_0(t) - \bar{\alpha}_0(t)\}^2} = \lim_{n \to \infty} [c_1^{T} E_{\bar{\alpha}_0(t)}^{-1} c_1]
$$

where $$c_1^T = \{1, \ldots, 0\}$$, and 0 denotes the vector of zeros of length t. The values of LTBS and LSB, shown in Table 1, were computed numerically for t ranging from 100 to 600 (the convergence is slower for smaller values of k).

References


Maciej Niedźwiecki was born in Poznań, Poland in 1953. He received the M.Sc. and Ph.D. degrees from the Gdańsk University of Technology, Gdańsk, Poland, and the Dr. Hab. (D.Sc.) degree from the Technical University of Warsaw, Warsaw, Poland, in 1977, 1981 and 1991, respectively. He spent three years as a Research Fellow with the Department of Systems Engineering, Australian National University from 1986 to 1989. From 1990 to 1993 he served as a Vice-Chairman of the Technical Committee on Theory of the International Federation of Automatic Control (IFAC). He is currently a Professor and Head of the Department of Automatic Control, Faculty of Electronics, Telecommunications and Computer Science, Gdańsk University of Technology. His main areas of research interest include system identification, statistical signal processing and adaptive systems. He is the author of the book *Identification of Time-varying Processes* (Wiley, 2000).

Dr. Niedźwiecki is currently Associate Editor for IEEE Transactions on Signal Processing, a member of the IFAC committees on Modeling, Identification and Signal Processing and on Large Scale Complex Systems, and a member of the Automatic Control and Robotics Committee of the Polish Academy of Sciences (PAN).

Michał Meller received the M.Sc. and Ph.D. degrees in Automatic Control from the Gdańsk University of Technology, Gdańsk, Poland, in 2007 and 2010, respectively. Since 2007 he has been working in the Department of Signal and Information Processing, Bumar Elektronika, Gdańsk Division. In 2010 he also joined the Department of Automatic Control at the Gdańsk University of Technology, Faculty of Electronics, Telecommunications and Computer Science. His professional interests include signal processing and adaptive systems.