

IEEE Copyright Notice

- ©20xx IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.
- This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder.

Statistically Efficient Smoothing Algorithm for Time-Varying Frequency Estimation

Maciej Niedźwiecki

Abstract—The problem of extraction/elimination of a nonstationary sinusoidal signal from noisy measurements is considered. This problem is usually solved using adaptive notch filtering (ANF) algorithms. It is shown that the accuracy of frequency estimates can be significantly increased if the results obtained from ANF are backward-time filtered by an appropriately designed lowpass filter. The resulting adaptive notch smoothing (ANS) algorithm can be employed to perform many offline signal processing tasks, such as elimination of sinusoidal interference from a prerecorded signal, for example. We show that when the unknown signal frequency drifts according to the random-walk model, the optimally tuned ANS algorithm is, under Gaussian assumptions, statistically efficient, i.e., it attains the Cramér–Rao-type lower smoothing bound, which limits accuracy of any (whether causal or not) frequency estimation scheme.

Index Terms—Adaptive notch filtering, adaptive notch smoothing, frequency estimation.

I. INTRODUCTION

CONSIDER the problem of extraction or elimination of a nonstationary complex sinusoidal signal (cisoid) $s(t)$ from noisy measurements $y(t)$

$$\begin{aligned} y(t) &= s(t) + v(t) \\ s(t) &= ae^{j\sum_{l=1}^t \omega(l)}. \end{aligned} \quad (1)$$

We will assume that the complex-valued amplitude $a = be^{j\phi_0}$ (which incorporates the initial phase ϕ_0 of the signal) is constant, and that the real-valued instantaneous frequency $\omega(t) \in (-\pi, \pi]$ is a slowly varying quantity. In addition, we will assume the following:

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

Note that, according to (1), $s(t)$ is a constant-modulus signal ($|s(t)| = b$), i.e., its nonstationarity is caused by the instantaneous frequency changes only (the amplitude is assumed to be unknown but constant).

Due to its large practical relevance, the problem of retrieving noisy sinusoidal signals has attracted a great deal of interest in the literature—see, e.g., [1] and references therein. Recursive estimation of signal parameters—its amplitude

and frequency—can be achieved in many different ways using the so-called adaptive notch filtering (ANF) algorithms. When comparing different approaches to adaptive notch filtering—and the number of different ANF algorithms proposed over the past 30 years is truly overwhelming—one needs a good benchmark problem. The case which is referred to in a majority of analytical studies [2]–[6], and hence pays a role of such a benchmark, is retrieval of a cisoid with frequency drifting according to the random-walk (RW) equation

$$\omega(t) = \omega(t-1) + w(t) \quad (2)$$

where

- A2) the process $\{w(t)\}$, independent of $\{v(t)\}$, is a zero-mean white noise with variance σ_w^2 .

Even though from the practical viewpoint the RW model can be criticized as rather naive, it has several advantages. First, it allows one to objectively compare tracking capabilities of various algorithms for different signal-to-noise ratios $\text{SNR} = b^2/\sigma_v^2$ and different rates of frequency variations σ_w^2 . Second, under A2) the properties of many adaptive notch filters can be studied analytically. Finally, and perhaps most importantly, under the random-walk hypothesis one can establish the posterior Cramér–Rao lower tracking bound, which limits performance of any frequency tracking scheme [7]. This allows one to judge upon the statistical efficiency of different algorithms, i.e., to evaluate their performance in absolute, rather than relative, terms.

Almost all adaptive notch filtering algorithms proposed so far are causal estimation schemes, which means that the instantaneous frequency estimators are functions of the current and past measurements only: $\hat{\omega}(t) = h[\mathcal{Y}_-(t)]$, where $\mathcal{Y}_-(t) = \{y(1), \dots, y(t)\}$. While in the real-time applications causality is an obvious requirement, in many offline processing tasks, such as elimination of a sinusoidal interference from a prerecorded signal, estimation can be based on both past and “future” measurements, leading to $\tilde{\omega}(t) = h[\mathcal{Y}(N)]$, where $\mathcal{Y}(N) = \{\mathcal{Y}_-(t), \mathcal{Y}_+(t+1)\}$, $\mathcal{Y}_+(t) = \{y(t), \dots, y(N)\}$ and N denotes the number of available (recorded) data samples. When appropriately designed, such noncausal estimators, which incorporate smoothing, yield smaller estimation errors than their causal counterparts. Despite this obvious advantage, noncausal solutions are surprisingly absent from the theory and practice of adaptive notch filtering. One of the possible reasons of this situation is that noncausal filters are often, and rather deceptively, called *unrealizable*, which has usually a negative connotation.

A simple adaptive notch smoother (ANS), based on compensation of estimation delay arising in the frequency tracking loop of an adaptive notch filter, was proposed in [8]. In this case

Manuscript received August 2, 2007; revised February 18, 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Alper Tunga Erdogan.

The author is with the Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control, Gdańsk University of Technology, Gdańsk 80-952, Poland (e-mail: maciekn@eti.pg.gda.pl).

Digital Object Identifier 10.1109/TSP.2008.921781

the smoothed (noncausal) frequency estimate $\tilde{\omega}(t)$ is obtained by simply delaying the causal estimate $\hat{\omega}(t)$ provided by ANF: $\tilde{\omega}(t - \tau_o) = \hat{\omega}(t)$, where τ_o is an appropriately chosen (analytically computable) estimation delay. The approach pursued in this paper is more sophisticated.

The paper is organized as follows. The adaptive notch smoothing algorithm is derived and analyzed in Section II. The problem of evaluation of the Cramér–Rao-type lower frequency tracking bound (known) and frequency smoothing bound (new) is discussed in Section III. Section IV shows the results of three numerical studies. Finally, Section V concludes.

II. ADAPTIVE NOTCH SMOOTHING ALGORITHM

The proposed smoothing procedure is two-pass. During the first, forward-time pass, a conventional adaptive notch filtering algorithm is used to obtain preliminary (biased) frequency estimates. The second, backward-time pass, is needed to perform filtering of the trajectory of frequency estimates yielded by ANF. The transfer function of the smoothing (lowpass) filter depends on the frequency tracking characteristics of ANF employed in the first stage of processing and will be determined analytically.

A. Filtering

The forward-time frequency tracking will be realized using the signal-oriented version of the generalized adaptive notch filter (GANF) proposed and analyzed in [6], [9], and [10]. Generalized adaptive notch filters are identification algorithms designed to track quasi-periodically varying dynamic systems. They can be considered an extension, to the system case, of classical adaptive notch filters and they reduce to ANFs in the special signal case. The adaptive notch filter described in [10] combines the exponentially weighted least squares approach to amplitude tracking with gradient approach to frequency tracking. The rate of amplitude adaptation is controlled by the gain μ , $0 < \mu \ll 1$, and the rate of frequency adaptation is controlled by another gain γ , $0 < \gamma \ll \mu$. The algorithm can be summarized as follows:

$$\begin{aligned} \varepsilon(t) &= y(t) - e^{j\hat{\omega}(t)}\hat{s}(t-1) \\ \hat{s}(t) &= e^{j\hat{\omega}(t)}\hat{s}(t-1) + \mu\varepsilon(t) \\ \hat{\omega}(t+1) &= \hat{\omega}(t) - \gamma \text{Im} \left[\frac{\varepsilon^*(t)e^{j\hat{\omega}(t)}}{\hat{s}^*(t-1)} \right]. \end{aligned} \quad (3)$$

The tracking properties of this algorithm are well understood [6], which will be helpful when designing the smoothing filter. Using the approximating linear filter (ALF) technique, developed by Tichavský and Händel [4], one can establish the following steady-state relationship between the frequency estimation error $\Delta\hat{\omega}(t) = \hat{\omega}(t) - \omega(t)$, the one-step frequency changes $w(t) = \omega(t) - \omega(t-1)$, and the scaled measurement noise $z(t) = \text{Im}[v(t)/s(t)]$:

$$\begin{aligned} \Delta\hat{\omega}(t) &\cong \frac{F(q^{-1}) - 1}{1 - q^{-1}}w(t) + (1 - q^{-1})F(q^{-1})z(t) \\ F(q^{-1}) &= \frac{(1 - \delta)q^{-1}}{1 - (\lambda + \delta)q^{-1} + \lambda q^{-2}} \end{aligned} \quad (4)$$

where q^{-1} denotes the backward shift operator and $\lambda = 1 - \mu$, $\delta = 1 - \gamma$.

Note that under A1) $\{z(t)\}$ is a real-valued white noise with variance $\sigma_z^2 = \sigma_v^2/(2b^2)$. Under A1) and A2) the mean-squared frequency estimation error can be expressed in explicit form as [6]

$$\text{E} \left\{ [\hat{\omega}(t) - \omega(t)]^2 \right\} \cong \frac{\gamma^2}{4b^2\mu}\sigma_v^2 + \left[\frac{\mu}{2\gamma} + \frac{1}{2\mu} \right] \sigma_w^2. \quad (5)$$

The approximation in (5) is tight for sufficiently small values of μ and γ .

The minimum value of the frequency tracking error is achieved for

$$\mu = \mu_{\text{opt}} = \sqrt[4]{8\kappa}, \quad \gamma = \gamma_{\text{opt}} = \sqrt{2\kappa}$$

where

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

is a scalar coefficient that can be regarded as a measure of signal nonstationarity. As shown in [6], under Gaussian assumptions, the minimum tracking error obtained for the optimal settings

$$\text{E} \left\{ [\hat{\omega}(t) - \omega(t)]^2 \mid \mu_{\text{opt}}, \gamma_{\text{opt}} \right\} \cong \sqrt[4]{2\kappa^{-1}}\sigma_w^2 \quad (6)$$

attains its lower limit known as the posterior Cramér–Rao bound—see Section III below for more details. This means that in the case considered the optimally tuned ANF algorithm (3) is a *statistically efficient* procedure for tracking randomly drifting frequency. It is interesting to note that the algorithm (3) is not the only known efficient frequency tracker—at least four other ANF algorithms with the same property were pointed out in the literature [4], [5].

B. Smoothing

To obtain a smoothed estimate of $\omega(t)$, further denoted by $\tilde{\omega}(t)$, we will pass the estimates yielded by the “pilot” ANF algorithm (3) through an appropriately designed noncausal filter $G(q^{-1}) = \dots + g_{-1}q^{-1} + g_0 + g_1q^1 + \dots$

$$\tilde{\omega}(t) = G(q^{-1})\hat{\omega}(t). \quad (7)$$

For causal estimators, such a “postfiltering” technique was advocated in [11].

As demonstrated in [8], a very simple form of smoothing can be realized by regarding $\hat{\omega}(t)$ as an estimate of $\omega(t - \tau_o)$, rather than as an estimate of $\omega(t)$, where $\tau_o = \text{int}[\mu/\gamma]$ denotes a nominal (low-frequency) delay of the filter $F(q^{-1})$. This is equivalent to setting $G(q^{-1}) = q^{\tau_o}$. The approach described below is more sophisticated. The filter $G(q^{-1})$ will be designed so as to minimize the mean-squared frequency matching error $\text{E}[(\Delta\tilde{\omega}(t))^2]$, where $\Delta\tilde{\omega}(t) = \tilde{\omega}(t) - \omega(t)$. Combining (4) and (7), one arrives at

$$\Delta\tilde{\omega}(t) \cong \frac{X(q^{-1}) - 1}{1 - q^{-1}}w(t) + (1 - q^{-1})X(q^{-1})z(t) \quad (8)$$

where

$$X(q^{-1}) = F(q^{-1})G(q^{-1}). \quad (9)$$

Since the processes $\{w(t)\}$ and $\{z(t)\}$ are wide-sense stationary and mutually uncorrelated it holds that

$$\begin{aligned} \mathbb{E}\{[\Delta\tilde{\omega}(t)]^2\} &\cong \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{\Delta\tilde{\omega}}(\xi) d\xi \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{X(e^{-j\xi}) - 1}{1 - e^{-j\xi}} \right|^2 S_w(\xi) d\xi \\ &\quad + \frac{1}{2\pi} \int_{-\pi}^{\pi} |(1 - e^{-j\xi})X(e^{-j\xi})|^2 S_z(\xi) d\xi \end{aligned}$$

where, to avoid confusion with ω , the standard Fourier-domain angular frequency variable is denoted here by ξ . Furthermore, since $S_w(\xi) = \sigma_w^2$, $S_z(\xi) = \sigma_z^2$, $\forall \xi \in [-\pi, \pi]$, one arrives at

$$\begin{aligned} \mathbb{E}\{[\Delta\tilde{\omega}(t)]^2\} &\cong \frac{1}{2\pi} \int_{-\pi}^{\pi} f[X(e^{-j\xi})] d\xi \\ f[X] &= c_1(X - 1)(X^* - 1)H_1H_1^* + c_2XX^*H_2H_2^* \end{aligned} \quad (10)$$

where $c_1 = \sigma_w^2$, $c_2 = \sigma_z^2$, $H_1(e^{-j\xi}) = 1/(1 - \xi^{-1})$, $H_2(e^{-j\xi}) = 1 - \xi^{-1}$ and $*$ denotes the complex conjugation.

Minimization of (10) is pretty straightforward—the problem can be solved by minimizing $f[X(e^{-j\xi})]$ for every value of $\xi \in [-\pi, \pi]$. Using Wirtinger (complex-real) calculus, applicable to nonanalytic functions, such as $f(\cdot)$, and setting

$$\left. \frac{\partial f}{\partial X^*} \right|_{X=X_{\text{opt}}} = 0$$

one obtains $X_{\text{opt}} = c_1H_1H_1^*/(c_1H_1H_1^* + c_2H_2H_2^*)$ or equivalently

$$X_{\text{opt}}(q^{-1}) = \frac{c_1/c_2}{c_1/c_2 + (1 - q^{-1})^2(1 - q)^2}. \quad (11)$$

Since $c_1/c_2 = 2\sigma_w^2b^2/\sigma_v^2 = 2\kappa$, one can rewrite (11) in the form

$$X_{\text{opt}}(q^{-1}) = \frac{1}{a_0 + a_1(q^{-1} + q) + a_2(q^{-2} + q^2)} \quad (12)$$

where

$$a_0 = \frac{3 + \kappa}{\kappa}, \quad a_1 = -\frac{2}{\kappa}, \quad a_2 = \frac{1}{2\kappa}.$$

Observe that

$$F(q^{-1})F(q) = \frac{1}{b_0 + b_1(q^{-1} + q) + b_2(q^{-2} + q^2)} \quad (13)$$

where

$$\begin{aligned} b_0 &= \frac{1 + (\lambda + \delta)^2 + \lambda^2}{(1 - \delta)^2} \\ b_1 &= -\frac{(1 + \lambda)(\lambda + \delta)}{(1 - \delta)^2} \\ b_2 &= \frac{\lambda}{(1 - \delta)^2}. \end{aligned}$$

To find the values of λ and δ which allow one to factorize $X_{\text{opt}}(q^{-1})$ as $F(q^{-1})F(q)$ we will request that $a_0 = b_0$, $a_1 = b_1$ and $a_2 = b_2$. This leads to the following two conditions:

$$\frac{\gamma^2}{1 - \mu} = 2\kappa \quad (14)$$

$$\frac{\mu^2}{2 - \mu} = \gamma \quad (15)$$

which can be easily resolved given that the adaptation gain μ is sufficiently small. Actually, when $\mu \ll 1$, so that $1 - \mu \cong 1$ and $2 - \mu \cong 2$, (14) and (15) turn into $\gamma^2 \cong 2\kappa$ and $\mu^2 \cong 2\gamma$, respectively. This leads to

$$\gamma \cong \sqrt{2\kappa} = \gamma_{\text{opt}}, \quad \mu \cong \sqrt[4]{8\kappa} = \mu_{\text{opt}}. \quad (16)$$

The approximations are tight provided that $\mu_{\text{opt}} \ll 1$. Assuming that the term “much smaller” (\ll) can be interpreted as “at least ten times smaller,” one can rewrite this condition in a more explicit form as

$$\kappa \leq 10^{-5}. \quad (17)$$

As argued in [6], (17) can be regarded a condition of “satisfactory signal tracking” by the ANF filter (3).

According to (16) one arrives at

$$X_{\text{opt}}(q^{-1}) = F_{\text{opt}}(q^{-1})F_{\text{opt}}(q) \quad (18)$$

which, given that the forward-time ANF algorithm is optimally tuned, leads to [cf. (9)]

$$\begin{aligned} G_{\text{opt}}(q^{-1}) &= \frac{X_{\text{opt}}(q^{-1})}{F_{\text{opt}}(q^{-1})} = F_{\text{opt}}(q) \\ &= \frac{(1 - \delta_{\text{opt}})q}{1 - (\lambda_{\text{opt}} + \delta_{\text{opt}})q + \lambda_{\text{opt}}q^2} \\ \lambda_{\text{opt}} &= 1 - \mu_{\text{opt}}, \quad \delta_{\text{opt}} = 1 - \gamma_{\text{opt}}. \end{aligned} \quad (19)$$

Since the filter $G_{\text{opt}}(q^{-1})$ is *anticausal*, the smoothed estimate $\tilde{\omega}(t)$ can be obtained by means of backward-time filtering of the estimates yielded by the ANF algorithm (3). Such backward-time filtering can be performed recursively using the following equations:

$$\begin{aligned} \tilde{\omega}(N) &= \hat{\omega}(N) \\ \tilde{\omega}(N - 1) &= \hat{\omega}(N - 1) \\ \tilde{\omega}(t) &= (\lambda + \delta)\tilde{\omega}(t + 1) - \lambda\tilde{\omega}(t + 2) \\ &\quad + (1 - \delta)\hat{\omega}(t + 1), \quad t = N - 2, \dots, 1 \end{aligned} \quad (20)$$

where the optimal gains λ_{opt} and δ_{opt} , usually not known *a priori*, were replaced with λ and δ , respectively—the gains used in the tracking algorithm. Making such a choice is equivalent to adopting $G(q^{-1}) = F(q)$.

Once the smoothed frequency trajectory is available, one can use it to improve estimation results by running—as a follow up to (3)—another algorithm that incorporates the smoothed frequency estimates

$$\begin{aligned} \tilde{\varepsilon}(t) &= y(t) - e^{j\hat{\omega}(t)}\tilde{s}(t-1) \\ \tilde{s}(t) &= e^{j\hat{\omega}(t)}\tilde{s}(t-1) + \mu\tilde{\varepsilon}(t). \end{aligned} \quad (21)$$

Note that the “frequency guided” ANF filter (21), originally proposed in [8] for the purpose of fixed-lag smoothing, does not estimate the instantaneous frequency on its own—it relies on frequency estimates obtained from the smoothing filter (20).

Remark 1: Even though derivation and analysis of the ANF/ANS algorithms was carried out under the assumption that the signal amplitude a is unknown but constant, both filters cope favorably with slow frequency *and* amplitude changes.

Remark 2: The computational cost of the proposed ANS algorithm is very low. Careful examination of ANS recursions shows that they require only 20 real multiplications, one real division and evaluation of two complex exponentials per time update, offering in return both frequency and signal estimates.

C. Extension to the Multiple Frequencies Case

Using the framework described in [10, Sec. 3c], the proposed ANS algorithm can be easily extended to the multiple frequencies case, where

$$s(t) = \sum_{i=1}^k s_i(t), \quad s_i(t) = a_i(t)e^{j\sum_{l=1}^t \omega_i(l)}$$

and k denotes the number of cisoids embedded in noise.

The resulting algorithm is a parallel estimation scheme combining k “local” ANS filters. The component filters are designed to track different frequency components of $s(t)$ and are driven by “global” prediction errors $\varepsilon(t)$ and $\tilde{\varepsilon}(t)$ (to add some extra design flexibility, we have equipped each subalgorithm with independently assigned adaptation gains μ_i and γ_i):

pilot filter:

$$\begin{aligned} \varepsilon(t) &= y(t) - \sum_{i=1}^k e^{j\hat{\omega}_i(t)}\hat{s}_i(t-1) \\ \hat{s}_i(t) &= e^{j\hat{\omega}_i(t)}\hat{s}_i(t-1) + \mu_i\varepsilon(t) \\ \hat{\omega}_i(t+1) &= \hat{\omega}_i(t) - \gamma_i \text{Im} \left[\frac{\varepsilon^*(t)e^{j\hat{\omega}_i(t)}}{\hat{s}_i^*(t-1)} \right] \\ & \quad i = 1, \dots, k \\ \hat{s}(t) &= \sum_{i=1}^k \hat{s}_i(t) \\ & \quad t = 1, \dots, N \end{aligned} \quad (22)$$

smoothing filter:

$$\begin{aligned} \tilde{\omega}_i(N) &= \hat{\omega}_i(N) \\ \tilde{\omega}_i(N-1) &= \hat{\omega}_i(N-1) \\ \tilde{\omega}_i(t) &= (\lambda_i + \delta_i)\tilde{\omega}_i(t+1) - \lambda_i\tilde{\omega}_i(t+2) \\ & \quad + (1 - \delta_i)\hat{\omega}_i(t+1) \\ & \quad i = 1, \dots, k, \\ & \quad t = N-2, \dots, 1 \end{aligned} \quad (23)$$

frequency-guided filter:

$$\begin{aligned} r\tilde{\varepsilon}(t) &= y(t) - \sum_{i=1}^k e^{j\tilde{\omega}_i(t)}\tilde{s}_i(t-1) \\ \tilde{s}_i(t) &= e^{j\tilde{\omega}_i(t)}\tilde{s}_i(t-1) + \mu_i\tilde{\varepsilon}(t) \\ & \quad i = 1, \dots, k \\ \tilde{s}(t) &= \sum_{i=1}^k \tilde{s}_i(t) \\ & \quad t = 1, \dots, N \end{aligned} \quad (24)$$

where $\lambda_i = 1 - \mu_i$ and $\delta_i = 1 - \gamma_i$.

The efficient initialization procedure, which can be used to identify the number of frequency modes k , and to determine initial conditions—both frequencies and amplitudes—that guarantee smooth start, i.e., start without initialization transients, of the ANS algorithm (22)–(24) was presented in [12] (see also [13]).

Remark: When $k > 1$, the ALF analysis carried out above is valid as long as the estimated frequencies remain well-separated. When frequency tracking is disabled ($\gamma = 0$, $\hat{\omega}(0) = \omega_o$), the first-order steady-state parameter tracking properties of (3) are characterized by the relationship $E[\hat{s}(t)] = T(q^{-1})s(t)$ where $T(q^{-1}) = \mu/[1 - (1 - \mu)e^{j\omega_o}q^{-1}]$ is a narrowband extraction filter centered at the frequency ω_o , with bandwidth $B_{3\text{dB}} \cong 2\mu$. Therefore, the frequency separation condition for (22) can be approximately expressed in the form

$$|\omega_i(t) - \omega_j(t)| > \mu_i + \mu_j \quad \forall i \neq j, \forall t. \quad (25)$$

When the above condition is not fulfilled, the bandwidths of the extraction filters partially overlap, making the behavior of the entire structure difficult to predict. It should be stressed, however, that the proposed algorithm works pretty well even if the frequency separation condition is violated.

III. POSTERIOR CRAMÉR–RAO BOUNDS

When the estimated frequency is time-invariant ($\omega(t) = \omega_o$, $\forall t$), the measurement noise is circular Gaussian and the signal amplitude a is known, the estimation accuracy of any unbiased estimator $\hat{\omega}(N) = h[\mathcal{Y}(N)]$ of ω_o is limited by the following Cramér–Rao inequality:

$$\begin{aligned} E \left[(\hat{\omega}(N) - \omega_o)^2 \right] &\geq \frac{6\sigma_v^2}{b^2N(N^2 - 1)} = \text{CRB} \\ &\cong \frac{6}{N^3\text{SNR}} \end{aligned} \quad (26)$$

where CRB denotes the Cramér–Rao lower frequency estimation bound [14], [15]. A similar bound, also proportional to $1/N^3$, can be derived in the unknown amplitude case.

When the instantaneous frequency is a stochastic variable, the classical Cramér–Rao inequality does not apply. A bound that is similar to the CRB, and can be applied to signals with randomly drifting frequency, was derived by Tichavský [7], based on a more general result due to van Trees [16]. This bound was called in [7] the posterior Cramér–Rao bound (PCRB). Denote by $\hat{\omega} = [\hat{\omega}(1), \dots, \hat{\omega}(N)]^T$ any estimator (possibly biased) of the vector of instantaneous frequencies $\omega = [\omega(1), \dots, \omega(N)]^T$, based on the entire observation history $\mathcal{Y}(N)$. Suppose that, in addition to A1) and A2), the following two assumptions are fulfilled:

- A3) the processes $\{v(t)\}$ and $\{w(t)\}$ are Gaussian;
- A4) the complex amplitude a is known and the prior distribution of $\omega(1)$ is noninformative: $\pi(\omega(1)) = 1/(2\pi)$ for $\omega(1) \in (-\pi, \pi]$.

Then it is possible to show that [7]

$$\mathbf{E}[(\hat{\omega} - \omega)(\hat{\omega} - \omega)^T] \geq \mathbf{F}_N^{-1} \quad (27)$$

where \mathbf{F}_N denotes the so-called posterior Fisher information matrix

$$\mathbf{F}_N = \frac{1}{\sigma_w^2} [\mathbf{G}_N + 2\kappa \mathbf{H}_N]$$

$$\mathbf{G}_N = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ -1 & 2 & -1 & \dots & 0 & 0 \\ 0 & -1 & 2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 2 & -1 \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}$$

$$\mathbf{H}_N = \begin{bmatrix} N & N-1 & N-2 & \dots & 2 & 1 \\ N-1 & N-1 & N-2 & \dots & 2 & 1 \\ N-2 & N-2 & N-2 & \dots & 2 & 1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 2 & 2 & 2 & \dots & 2 & 1 \\ 1 & 1 & 1 & \dots & 1 & 1 \end{bmatrix}$$

and averaging is carried out over all realizations of the measurement noise sequence $\mathcal{V} = \{v(t), 1 \leq t \leq N\}$ and all realizations of the frequency trajectory $\Omega = \{\omega(t), 1 \leq t \leq N\}$.

Note that (27) entails

$$\mathbf{E}[(\hat{\omega}(i) - \omega(i))^2] \geq [\mathbf{F}_N^{-1}]_{ii} \quad (28)$$

where \mathbf{A}_{ij} denotes the (i, j) th element of the matrix \mathbf{A} . Denote by $\mathbf{A}^{(i,j)}$ the matrix obtained after removing from \mathbf{A} the i th row and the j th column. Then the right-hand side of (28) can be expressed in the form

$$[\mathbf{F}_N^{-1}]_{ii} = \frac{\det(\mathbf{F}_N^{(i,i)})}{\det(\mathbf{F}_N)}. \quad (29)$$

Let $N = 2n + 1$. Based on (28) Tichavský computed the asymptotic (steady-state) value of PCRB for any causal

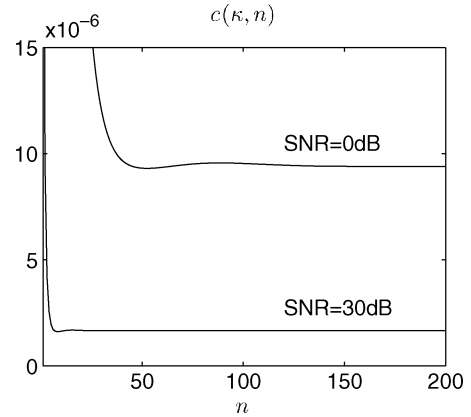


Fig. 1. Convergence of the estimation bound $c(n, \kappa)$ to the corresponding limiting values for $\sigma_w = 0.001$ and for two SNR values: 0 dB ($\kappa = 10^{-6}$) and 30 dB ($\kappa = 10^{-3}$).

estimator of $\omega(t)$. This bound, further referred to as the lower tracking bound (LTB), can be obtained from

$$\begin{aligned} \text{LTB} &= \lim_{t \rightarrow \infty} \inf_{\hat{\omega}(\cdot)} \mathbf{E} \left\{ [\hat{\omega}(t) - \omega(t)]^2 \right\} \\ &= \lim_{n \rightarrow \infty} [\mathbf{F}_{2n+1}^{-1}]_{2n+1, 2n+1}. \end{aligned} \quad (30)$$

The analogous quantity, limiting the steady-state performance of *any* frequency estimation scheme (including all noncausal estimators), will be called the lower smoothing bound (LSB). It is given by

$$\begin{aligned} \text{LSB} &= \lim_{t \rightarrow \infty} \inf_{\tilde{\omega}(\cdot)} \mathbf{E} \left\{ [\tilde{\omega}(t) - \omega(t)]^2 \right\} \\ &= \lim_{n \rightarrow \infty} [\mathbf{F}_{2n+1}^{-1}]_{n+1, n+1}. \end{aligned} \quad (31)$$

Note that in the asymptotic case the estimate $\hat{\omega}(t)$ in (30) is a function of an infinite past observation history $\mathcal{Y}_{-}^{\infty}(t) = \{y(i), -\infty < i \leq t\}$ while the estimate $\tilde{\omega}(t)$ in (31) is a function of both an infinite past and infinite future observation history $\mathcal{Y}^{\infty} = \{\mathcal{Y}_{-}^{\infty}(t), \mathcal{Y}_{+}^{\infty}(t+1)\}$, where $\mathcal{Y}_{+}^{\infty}(t) = \{y(i), t \leq i < \infty\}$.

As shown in [7], the lower tracking bound can be computed analytically

$$\text{LTB} = \sigma_w^2 g_{\text{LTB}}(\kappa) \quad (32)$$

where

$$g_{\text{LTB}}(\kappa) = -1 + \sqrt{1 + 4u^{-1}}, \quad u = \kappa + \sqrt{\kappa^2 + 8\kappa}.$$

Even though it is difficult to derive similar expressions for the lower smoothing bound, it is clear from (29) and (31) that it must take the form

$$\text{LSB} = \sigma_w^2 g_{\text{LSB}}(\kappa).$$

The lower smoothing bound can be evaluated numerically as $[\mathbf{F}_{2n+1}^{-1}]_{n+1, n+1} = c(\kappa, n)$ provided that n is sufficiently large. Fig. 1 illustrates convergence of $c(\kappa, n)$ to the corresponding limiting values $c(\kappa, \infty) = \text{LSB}$ for $\sigma_w = 0.001$ and for two

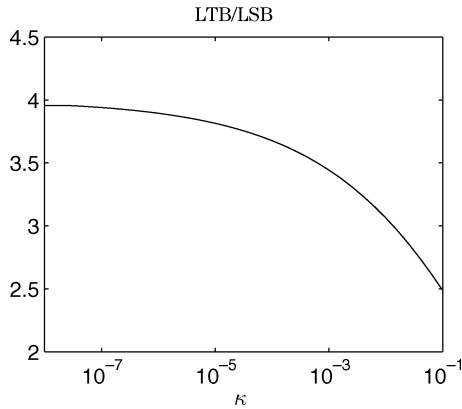


Fig. 2. Dependence of the limiting tracking versus smoothing variance ratio on the degree of signal nonstationarity κ .

SNR values: 0 dB ($\kappa = 10^{-6}$) and 30 dB ($\kappa = 10^{-3}$). Generally, the smaller the value of κ , the larger n should be to guarantee good approximation.

Since, in the case of smoothing, the amount of data used for estimation is effectively doubled compared to tracking, it holds that $LSB < LTB$, i.e., the limiting accuracy of noncausal estimators exceeds accuracy of their causal counterparts. The possible margin of improvement LTB/LSB , which is a function of κ , is depicted in Fig. 2. Note the difference between the time-invariant case and the time-varying case. When the number of data points N is doubled in (26), the limiting frequency estimation variance is reduced by the factor of eight.¹ In the randomly drifting frequency scenario, the analogous reduction rate approaches 4 for small values of κ .

When $\sqrt{\kappa} \ll 1$, which is a less stringent condition than (17), it holds that $u \cong \sqrt{8\kappa}$ and $g_{LTB}^2 \sqrt{4u^{-1}} \cong \sqrt[4]{2\kappa^{-1}}$, leading to the following approximation:

$$LTB \cong \sigma_w^2 \sqrt[4]{2\kappa^{-1}}. \quad (33)$$

Comparison of (6) with (33) confirms that when the degree of signal nonstationarity κ is sufficiently small, the optimally tuned ANF filter (3) attains the lower tracking bound, i.e., it is—under assumptions A1)–A4)—a statistically efficient frequency estimation procedure. Numerical test reported in the next section clearly demonstrate that the proposed ANS algorithm is also statistically efficient, i.e., when optimally tuned it reaches the lower smoothing bound, which limits performance of any noncausal frequency estimation scheme.

IV. NUMERICAL STUDY

Four numerical experiments were performed to check performance of the proposed scheme.

Example 1: Fig. 3 shows a comparison of the theoretical values of the lower smoothing bound with experimental results obtained for the signal (1), (2), for three different speeds of frequency variation σ_w (0.01, 0.001, and 0.0001) and 16 different SNR values, ranging from 0 to 30 dB. The mean-squared frequency estimation errors were evaluated (for the optimally tuned ANS algorithm) by means of joint time and ensemble

¹This obviously means that the amount of information relevant for frequency estimation does not grow linearly with N .

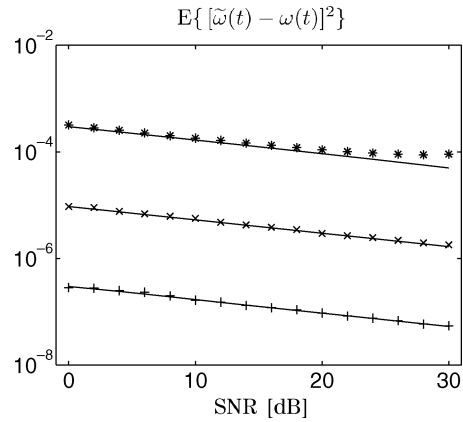


Fig. 3. Comparison of the theoretical values of the lower smoothing bound (solid lines) with experimental results obtained for the signal with randomly drifting frequency for three different speeds of frequency variation: $\sigma_w = 0.01$ (*), $\sigma_w = 0.001$ (×), $\sigma_w = 0.0001$ (+), and 16 different SNR values.

averaging. First, for each realization of the measurement noise sequence and each realization of the frequency trajectory, the mean-squared errors were computed from 200 iterations of the ANS filter (after the algorithm has reached its steady state). The obtained results were next averaged over 200 realizations of $\{v(t)\}$ and 200 realizations of $\{w(t)\}$.

Note the good agreement between the theoretical curves and the results of computer simulations. The worst fit can be observed for the fastest frequency changes ($\sigma_w = 0.01$), which is understandable since in this case the associated degrees of signal nonstationarity κ range from 10^{-4} (for SNR = 0 dB) to 10^{-1} (for SNR = 30 dB), i.e., they violate the “satisfactory tracking” condition (17).

Example 2: In our second experiment the average rate of frequency change and the signal-to-noise ratio were fixed ($\sigma_w = 0.003$, SNR = 10 dB) whereas the adaptation gain μ was changed. To reduce the number of design degrees of freedom the second adaptation gain was set to $\gamma = \mu^2/2$. This rule was inspired by the fact that, according to (16), $\gamma_{opt} = \mu_{opt}^2/2$. All remaining simulation details were the same as those described in Example 1. Fig. 4 shows how the mean-squared frequency estimation errors depend on μ for the ANF and ANS algorithms, respectively. Note that the ANS filter yields uniformly better (approximately four times better) results than its ANF “pilot” prototype, for the entire range of adaptation gains, and in particular, for the gains that differ from the optimal settings.

Example 3: Our third experiment aimed at checking how the proposed algorithm behaves under nonstandard conditions. Fig. 5 shows results obtained for a signal with sinusoidally varying frequency

$$\omega(t) = 0.5 + 0.2 \sin \frac{2\pi t}{3000}.$$

To ensure that the ANF and ANS algorithms reach their steady-state behavior the evaluation interval, covering 3000 samples (i.e., one period of frequency changes), was placed in the middle of a much wider analysis interval. All mean-squared errors were evaluated for 200 realizations of the measurement

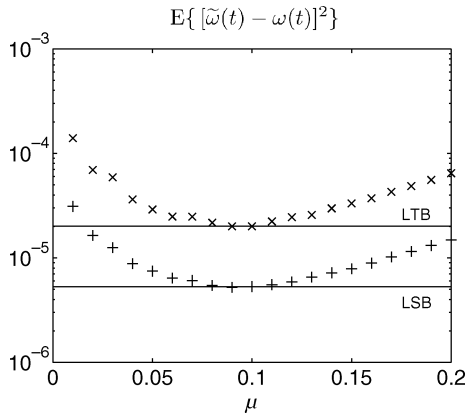


Fig. 4. Comparison of the results obtained for the ANF algorithm (x) and the ANS algorithm (+) for different values of the adaptation gain μ ($\gamma = \mu^2/2$). The LTB and the LSB are indicated by horizontal lines.

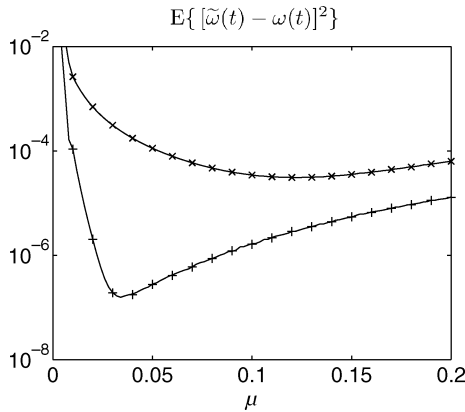


Fig. 5. Comparison of the results obtained for the ANF algorithm (x) and the ANS algorithm (+) for different values of the adaptation gain μ ($\gamma = \mu^2/2$) for the signal with sinusoidally varying frequency. Both plots (solid lines) were evaluated on a grid of 100 equidistant values of μ .

noise (SNR = 10 dB). Similarly as in the previous example, the ANS algorithm yielded uniformly better results than the ANF algorithm. It is interesting to note that the margin of improvement, offered by smoothing, is in this deterministic case much larger than in the stochastic case—the peak-to-peak (or, more adequately, bottom-to-bottom) variance reduction is equal 1.76×10^2 , i.e., the smallest achievable mean-squared smoothing errors are two orders of magnitude smaller than the corresponding tracking errors.

Example 4: The purpose of our last experiment was to check properties of the multiple frequency smoother (22), (25), and (26). Figs. 6 and 7 show results obtained for a signal consisting of two components with well-separated linearly time-varying frequencies

$$\omega_1(t) = 0.1 + 0.0001t, \quad \omega_2(t) = 0.3 + 0.0001t.$$

The structure made up of two ANF/ANS filters ($k = 2$) was implemented. The same settings were adopted for both subalgorithms: $\mu = \mu_1 = \mu_2$, $\gamma = \gamma_1 = \gamma_2$. Similarly as before, γ was set to $\mu^2/2$.

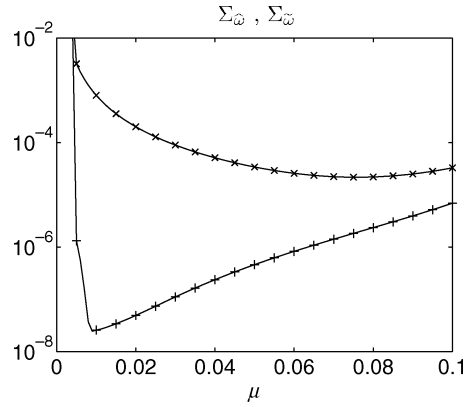


Fig. 6. Comparison of frequency estimation errors obtained for the ANF algorithm (x) and the ANS algorithm (+) for different values of the adaptation gain μ ($\gamma = \mu^2/2$) for a signal consisting of two components with well-separated linearly varying frequencies. Both plots (solid lines) were evaluated on a grid of 100 equidistant values of μ .

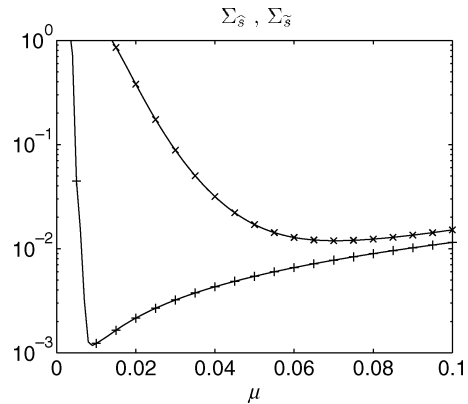


Fig. 7. Comparison of signal reconstruction errors obtained for the ANF algorithm (x) and the ANS algorithm (+) for different values of the adaptation gain μ ($\gamma = \mu^2/2$) for a signal consisting of two components with well-separated linearly varying frequencies. Both plots (solid lines) were evaluated on a grid of 100 equidistant values of μ .

For each realization of measurement noise the steady-state mean-squared estimation errors, given by

$$\Sigma_{\hat{\omega}} = \frac{1}{2000} \sum_{t=2001}^{4000} \min \left\{ [\hat{\omega}_1(t) - \omega_1(t)]^2 + [\hat{\omega}_2(t) - \omega_2(t)]^2, \right. \\ \left. [\hat{\omega}_1(t) - \omega_2(t)]^2 + [\hat{\omega}_2(t) - \omega_1(t)]^2 \right\}$$

$$\Sigma_{\hat{s}} = \frac{1}{2000} \sum_{t=2001}^{4000} |\hat{s}(t) - s(t)|^2$$

were computed in the evaluation interval [2001,4000] (placed inside a wider analysis interval [1,6000]). The frequency smoothing and signal smoothing errors $\Sigma_{\hat{\omega}}$ and $\Sigma_{\hat{s}}$ were computed in an analogous way. All results were next averaged over 200 realizations of noise (for both signal components SNR was equal to 10 dB). From the qualitative viewpoint, the situation is similar to that discussed in the previous, single-frequency case—smoothing offers significant improvements over filtering for a wide range of adaptation gains $\mu \in [0.001, 0.1]$.

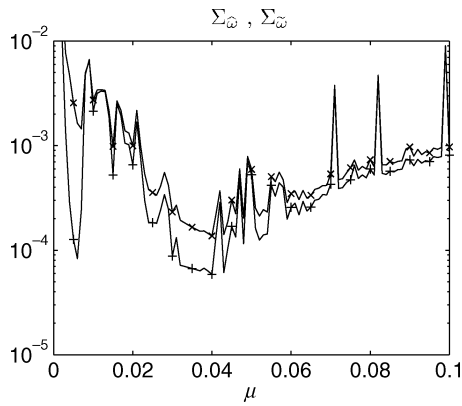


Fig. 8. Comparison of frequency estimation errors obtained for the ANF algorithm (x) and the ANS algorithm (+) for different values of the adaptation gain μ ($\gamma = \mu^2/2$) for a signal consisting of two components with poorly separated linearly varying frequencies. Both plots (solid lines) were evaluated on a grid of 100 equidistant values of μ .

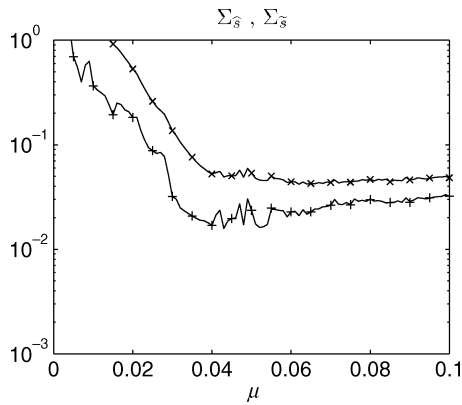


Fig. 9. Comparison of signal reconstruction errors obtained for the ANF algorithm (x) and the ANS algorithm (+) for different values of the adaptation gain μ ($\gamma = \mu^2/2$) for a signal consisting of two components with poorly separated linearly varying frequencies. Both plots (solid lines) were evaluated on a grid of 100 equidistant values of μ .

To check what happens when the separation condition (25) is not fulfilled, another linearly chirped signal was generated using

$$\omega_1(t) = 0.1 + 0.0001t, \quad \omega_2(t) = 0.7 - 0.0001t.$$

Note that the corresponding frequency trajectories intersect in the middle of the analysis interval. This means that in the vicinity of the crossover point ($t = 3000$) the condition (25) is violated for every value of $\mu \in [0.001, 0.1]$. Results obtained in this case are summarized in Figs. 8 and 9. The ruggedness of the plots is caused by outliers due to the frequency switching effect, observed (rarely) when the separation condition is not fulfilled. When passbands of two ANF filters overlap, the corresponding estimates, say $\hat{\omega}_1(t)$ and $\hat{\omega}_2(t)$, may occasionally “switch” between $\omega_1(t)$ and $\omega_2(t)$. This means that in some time intervals $\hat{\omega}_1(t)$ follows $\omega_1(t)$ and $\hat{\omega}_2(t)$ follows $\omega_2(t)$, while in some other intervals $\hat{\omega}_1(t)$ follows $\omega_2(t)$ and $\hat{\omega}_2(t)$ follows $\omega_1(t)$. Frequency switching produces jitter which can be seen at both error plots. Note that even under such harsh estimation conditions, smoothing yields consistently better results than filtering (although the achievable rate of improvement is much reduced).

V. CONCLUSION

We have shown that accuracy of frequency estimates yielded by the adaptive notch filtering (ANF) algorithm can be considerably improved by means of “postfiltering.” The characteristics of the smoothing filter, operated backward in time, can be derived analytically based on the linearized description of ANF. We have demonstrated that when the time-varying signal frequency drifts according to the random-walk model, the proposed adaptive notch smoothing (ANS) algorithm is statistically efficient, i.e., it attains the Cramér–Rao type lower smoothing bound. Since ANS is a noncausal estimation scheme, its use is restricted to offline processing tasks, such as extraction or elimination of nonstationary sinusoidal signals from sequences of prerecorded noisy measurements. The algorithm is computationally very attractive and can be easily extended to the multiple frequencies case. The frequency smoothing framework described in the paper can be extended to other frequency estimation schemes, such as the multiple frequency tracker [17].

REFERENCES

- [1] P. A. Regalia, *Adaptive IIR Filtering in Signal Processing and Control*. New York: Marcel Dekker, 1995.
- [2] P. Händel and A. Nehorai, “Tracking analysis of an adaptive notch filter with constrained poles and zeros,” *IEEE Trans. Signal Process.*, vol. 42, no. 2, pp. 281–291, Feb. 1994.
- [3] N. I. Cho and S. U. Lee, “Tracking analysis of an adaptive lattice notch filter,” *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 42, no. 3, pp. 186–195, Mar. 1995.
- [4] P. Tichavský and P. Händel, “Two algorithms for adaptive retrieval of slowly time-varying multiple cisoids in noise,” *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1116–1127, May 1995.
- [5] P. Tichavský and A. Nehorai, “Comparative study of four adaptive frequency trackers,” *IEEE Trans. Signal Process.*, vol. 45, no. 6, pp. 1473–1484, Jun. 1997.
- [6] M. Niedźwiecki and P. Kaczmarek, “Tracking analysis of a generalized adaptive notch filter,” *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 304–314, Jan. 2006.
- [7] P. Tichavský, “Posterior Cramér–Rao bound for adaptive harmonic retrieval,” *IEEE Trans. Signal Process.*, vol. 45, no. 5, pp. 1473–1484, May 1995.
- [8] M. Niedźwiecki and A. Sobociński, “A simple way of increasing estimation accuracy of generalized adaptive notch filters,” *IEEE Signal Process. Lett.*, vol. 14, no. 3, pp. 217–220, Mar. 2007.
- [9] M. Niedźwiecki and P. Kaczmarek, “Generalized adaptive notch filters,” in *Proc. 2004 IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, Montreal, Canada, 2004, pp. II-657–II-660.
- [10] M. Niedźwiecki and P. Kaczmarek, “Generalized adaptive notch filter with a self-optimization capability,” *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4185–4193, Nov. 2006.
- [11] M. Niedźwiecki, “Identification of time-varying systems using combined parameter estimation and filtering,” *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, no. 4, pp. 679–686, Apr. 1990.
- [12] M. Niedźwiecki and P. Kaczmarek, “Adaptive notch filters based on combined parametric and nonparametric approach,” in *Proc. IFAC Workshop Adaptation Learning in Control Signal Processing*, Yokohama, Japan, 2004, pp. 439–444.
- [13] M. Niedźwiecki and P. Kaczmarek, “Identification of quasi-periodically varying systems using the combined nonparametric/parametric approach,” *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 4588–4598, Dec. 2005.
- [14] D. C. Rife and R. R. Boorstyn, “Single-tone parameter estimation from discrete-time observations,” *IEEE Trans. Inf. Theory*, vol. 20, no. 5, pp. 591–598, Sep. 1974.

- [15] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood and Cramér–Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.
- [16] H. L. van Trees, *Detection, Estimation and Modulation Theory*. New York: Wiley, 1968.
- [17] M. Niedźwiecki, "From the multiple frequency tracker to the multiple frequency smoother," in *Proc. 2008 IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Las Vegas, NV, Mar.–April 2008, pp. 3549–3552.



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.

From 1986 to 1989, he was a Research Fellow with the Department of Systems Engineering, Australian National University. From 1990 to 1993, he served as a Vice-Chairman of Technical Committee on Theory of the International Federation of Automatic Control

(IFAC). He is a Professor and Head of the Department of Automatic Control, Faculty of Electronics, Telecommunications and Computer Science, Gdańsk University of Technology. He is the author of the book *Identification of Time-Varying Processes* (Wiley, 2000). His main areas of research interests include system identification, signal processing, and adaptive systems.