An Adaptive Notch Filter with Improved Tracking Properties

Marina V. Dragošević, *Member, IEEE,* and Srdjan S. Stanković

***Abstract*—In this paper, an analysis of the properties of an adaptive notch filter (ANF) applied to time-varying frequency tracking is presented. Starting from the derivation of an expression for ANF output power, asymptotically optimal values for the pole contraction and forgetting factors are derived for recursive prediction error (RPE) type ANF algorithms. Based on the obtained results, a new ANF algorithm that includes adaptation of both pole contraction and forgetting factors is proposed. The given experimental results confirm the theoretical conclusions and show that the proposed algorithm is highly efficient in practice.**

I. Introduction

Two closely related problems have attracted a great deal of attention of researchers in the field of signal processing: The first one deals with eliminating (or extracting) a set of sine waves from noisy data, whereas the second one is concerned with the estimation of their unknown frequencies. One of the methodologically consistent approaches to these problems is based on adaptive notch filtering. Starting from frequency adaptation of pole-zero pairs with specific properties, numerous algorithms for frequency estimation have been proposed (e.g., [1]–[3]). Depending on the initial problem formulation, these algorithms can be considered to belong either to the class of stochastic gradient-type recursive schemes or to the generalized least-squares or approximate maximum likelihood procedures [4], [5], [1], [3]. Most of the existing analyses of adaptive notch filtering algorithms deal with the constant frequency case. Recently, some contributions have been made in the direction of time-varying frequency tracking [6]–[9]. It has been shown that the accuracy of recursive prediction error (RPE) type adaptive notch filters (ANF's) depends on two principal parameters: the forgetting factor of the estimation algorithm and the pole contraction factor of the notch filter. The analysis presented in [8] and [9] shows that an asymptotically optimal value of the forgetting factor can be derived on the basis of the assumption that the contraction factor is close to one but that the unknown frequency lies well within the ANF notch bandwidth.

Manuscript received December 18, 1992; revised February 14, 1995. This work was supported by the Science Foundation of the Republic of Serbia, project 1007. The associate editor coordinating the review of this paper and approving it for publication was Dr. H. Fan.

M. V. Dragošević was with the Computer System Design Laboratory, Vinča Institute, Belgrade, Yugoslavia. She is now with IOTEK, Inc. Dartmouth, Canada B3B 1L7.

S. S. Stanković is with the Faculty of Electrical Engineering, University of Belgrade, Belgrade, Yugoslavia.

IEEE Log Number 9413316.

The general motivation for the research presented in this paper has been twofold: i) to get a deeper insight into the properties of ANF in the case of time-varying frequencies and ii) to try to design a fully adaptive ANF, not requiring any *a priori* assumptions about the signal and noise characteristics.

In the first part of the paper, an attempt is made to provide a more detailed insight into the asymptotic properties of ANF based on RPE algorithms, starting from two performance criteria: the mean notch filter (NF) output power and the asymptotic mean-square error of frequency tracking. A derived new expression for the NF output power serves as a prerequisite for finding the asymptotically optimal values of both pole contraction and forgetting factors.

The derived optimality conditions for ANF depend on the signal and noise parameters. As a result of an effort to design an algorithm not relying on such an *a priori* knowledge and providing, at the same time, convergence to the optimal regime, in the second part of the paper, a new ANF is presented, based on the estimation of all the unknown NF parameters (both the notch frequency and the pole contraction factor). The algorithm is of the RPE type. Its asymptotic behavior is analyzed by using the ODE methodology, with an intention to demonstrate convergence to the optimal parameter values. Moreover, it is shown that the choice of the forgetting factor in the estimation algorithm, which exerts a domimant influence on the tracking accuracy, can be made data adaptive by using the pole contraction factor estimates.

Characteristic experimental results are also given. They confirm the derived theoretical conclusions and illustrate the main properties of the proposed algorithm. An extension to the multiple sinusoid case, based on the NF cascade form, is made evident experimentally.

II. Statement of the Problem

The following constrained form of the notch filter (CNF) has been considered in many papers [1]–[3], [5], [7], [12]

$$N(z)=\prod_{k=1}^{n}\frac{1-2\cos\omega_k z^{-1}+z^{-2}}{1-2\alpha_k\cos\omega_k z^{-1}+\alpha_k^2 z^{-2}};\quad 0<1-\alpha_k\ll 1 \tag{1}$$

where ω_k represents the notch frequencies and α_k the corresponding pole contraction factors defining the bandwidths of the notches, which are proportional to $1-\alpha_k$ [2], [7]. The entire cascade form defined by (1) is completely characterized by the individual properties of its second-order stages when the frequencies ω_k are well separated [1], [10]. Therefore, we shall pay attention mainly to the second-order single-stage CNF

$$N(z) = \frac{1 - 2\cos\omega z^{-1} + z^{-2}}{1 - 2\alpha\cos\omega z^{-1} + \alpha^2 z^{-2}}; \quad 0 < 1 - \alpha \ll 1. \quad (2)$$

If the input signal is in the form

$$y(i) = U\cos\omega_0 i + \sigma_2\xi(i) \quad (3)$$

where $\xi(i)$ is a zero-mean unit-variance white noise, then the output $e(i) = N(q)y(i)$, where q^{-1} is the unit delay operator, consists of a sinusoidal term, which is canceled for $\omega = \omega_0$, and a colored noise term. The effect of notch filtering can be seen from the average output power

$$P = \lim_{i\to\infty}\frac{1}{N}\sum_{i=1}^{N} E\{e(i)^2\}$$

represented as a function of ω and α. In the practically most important case when $|\ \delta\ |=|\ \omega-\omega_0\ | \ll 1-\alpha \leq 1$, it is possible to derive the following expression as a good approximation:

$$P = (\frac{\delta}{1-\alpha})^2\sigma_0^2 + (2-\alpha)\sigma_2^2 \quad (4)$$

where $\sigma_0^2 = \frac{U^2}{2}$; the derivation is given in Appendix A. This expression, describing the CNF behavior more precisely than in the related analyses shown, e.g., in [1], [8], and [7], is crucial for our further elaborations.

The derived formula can be used directly to find the optimal pole contraction factor α. After differentiating (4) w.r.t. α, one obtains the following condition for the minimum of P:

$$(1-\alpha)^3 = 2\delta^2\frac{\sigma_0^2}{\sigma_2^2} \quad (5)$$

assuming that δ can never be exactly equal to zero. This condition results from a tradeoff between the following two requirements: a) line suppression and b) preservation of the wideband spectral component of the input signal. In this sense, (5) defines the optimal CNF bandwidth for a given δ.

If the true frequency of the sinusoid in the input signal is not known, it can be estimated from the input data, and notch filtering can be done in an adaptive way. Adaptive notch filters derived from (2) are characterized, in general, by the time-varying operator $N(i,q)$, which is obtained from $N(q)$ in (2) by replacing the constant notch frequency ω by a time-varying estimate $\hat{\omega}(i)$ of the true frequency ω_0 of the input signal line component, which is obtained at the instant i on the basis of current signal measurements. There are many different successful adaptive notch filtering methods proposed in the literature [1]–[3], [11]. One of their common features is that the frequency estimation accuracy is better and notch filtering more effective when α is closer to 1 [12]. However, it has been noticed that the convergence rate can be very low if α is very close to 1 during the whole adaptation procedure [2], [10], [13]. Indeed, (5) indicates roughly that α should be gradually increased along with the improvement of the frequency estimate (if δ is considered as the frequency estimation error).

In the general case of time-varying frequency of the sinusoidal component, we have

$$\omega_0(i) = \omega_0(i-1) + \sigma_1\nu(i) \quad (6)$$

where $\omega_0(i)$ is the instantaneous frequency, and $\sigma_1\nu(i)$ a small disturbance. The input is then given by

$$y(i) = U\cos(\sum_{j=0}^{i}\omega_0(j)) + \sigma_2\xi(i) \quad (7)$$

with the power $\sigma_0^2 + \sigma_2^2$. In this situation, convergence to zero of the estimation error $\delta(i) = \hat{\omega}(i) - \omega_0(i)$, where $\hat{\omega}(i)$ is generated by a recursive algorithm of RPE type, cannot be achieved in general, i.e., the narrowband component will never be completely removed from the filter output $e(i) = N(i,q)y(i)$. The frequency tracking accuracy depends on the signal properties, as well as on the estimation algorithm itself, and the problem of choosing the contraction factor α becomes much more important (see [14] for a general discussion on time-varying parameter estimation).

In order to get a more precise insight into the performance of ANF in this situation, we shall use, together with the average output power P, the average mean-square error of frequency estimation

$$J = \lim_{i\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\{\delta(i)^2\}. \quad (8)$$

As we shall focus our attention to the steady-state regime, when all the transients can be neglected, we shall assume that the limit $\Pi = \lim_{i\to\infty}E\{\delta(i)^2\}$ exists and that $J = \Pi$ (a similar assumption has been made implicitely in [8] and [9] in an analogous context). Obviously, both P and Π depend on the properties of both the input signal (expressed through σ_0, σ_1 and σ_2) and the frequency estimation algorithm itself (expressed in the case of frequency tracking algorithms of RPE type through the pole contraction factor α and the forgetting factor ρ) [2], [1], [6], [4], [14]. Great attention has been paid to the choice of the forgetting factor ρ in the general context of time-varying parameter estimation since its influence to the tracking accuracy has been found to be fundamental [4], [14]. An asymptotically optimal value of ρ, which minimizes Π for the RPE frequency tracking algorithms, is derived in [8] and [9] as a function of σ_1, σ_2/σ_0, and α on the condition that $\delta(i) \ll 1-\alpha \ll 1$, i.e., that α is as high as possible, but such that the unknown instantaneous frequency still lies well within the ANF notch bandwidth.

We shall show that both criteria P and Π can be used simultaneously for determining the optimal values of α and ρ in a straightforward and consistent way in a more general context.

III. Asymptotically Optimal α and ρ

A. ANF Output Power

The ANF output signal $e(i)$ in the case of slowly time-varying frequency of the input signal (7) consists of two terms. If the frequency variations between successive samples of $\omega_0(i)$ are sufficiently small and the absolute frequency estimation error much smaller than $1-\alpha$, the first term can be approximately expressed as $u(i)\cos(\sum_{j=0}^{i}\omega_0(j))$, where $u(i)$ is the time-varying amplitude

$$u(i) \approx \frac{\delta(i)}{1-\alpha}U \tag{9}$$

obtained after neglecting the terms of order $\sigma_1\nu(i)$, $(1-\alpha)\delta(i)$, $\delta(i)^2/(1-\alpha)$ and $\delta(i)^3/(1-\alpha)^3$. The second term is due to noise. The ANF output power consists then approximately of the effective power due to the narrowband part and the power of the wideband noise and can be represented by the following expression, which is analogous to (4):

$$P = \frac{\Pi}{(1-\alpha)^2}\sigma_0^2 + (2-\alpha)\sigma_2^2. \tag{10}$$

A precise analysis is given in Appendix A.

Thus, on condition that the frequency changes are not too fast, the only statistical property of the estimation error relevant for choosing an optimal α minimizing P is the MSE. Of course, the MSE itself also depends on α. This dependence should be taken into account when trying to find an overall optimal value of α.

B. MSE for RPE Algorithms

Up to now, no specific assumptions have been made about the frequency estimation algorithm. For an RPE algorithm, the update of the ANF parameter $\hat{a}(i) = -2\cos\hat{\omega}(i)$ is given by (e.g., [4], [14], [2], [3])

$$\hat{a}(i+1) = \hat{a}(i) + \gamma_i R(i)^{-1}\psi(i)e(i) \tag{11}$$

$$R(i) = R(i-1) + \gamma_i(\psi(i)^2 - R(i-1)) \tag{12}$$

$$e(i) = N(i,q)y(i) = \frac{1+\hat{a}(i)q^{-1}+q^{-2}}{1+\alpha\hat{a}(i)q^{-1}+\alpha^2q^{-2}}\, y(i) \tag{13}$$

where the regressor $\psi(i)$ is an approximation of the negative prediction error gradient

$$\psi(i) = -\frac{\partial e(i)}{\partial a}\,|_{a=\hat{a}(i)} = \frac{-y(i-1)+\alpha e(i-1)}{1+\alpha\hat{a}(i)q^{-1}+\alpha^2q^{-2}} \tag{14}$$

where $\{\gamma_i\}$ is sequence of positive numbers, whereas

$$\rho_i = (1-\gamma_i)\gamma_{i-1}/\gamma_i \tag{15}$$

are the corresponding forgetting factors. Usually, constant values γ and ρ are selected for γ_i and ρ_i, satisfying $0 < \gamma = 1-\rho \ll 1$ [14].

The asymptotic mean-square error of the parameter estimates generated by (11)–(15) has been analyzed in [8] and [9] for the case of slowly time-varying frequencies obeying the random walk model. The basic assumptions have been that σ_1 and γ are sufficiently small and that $\nu(i)$ represents a zero-mean unit-variance white noise sequence, independent of the measurement noise $\xi(i)$. The following main conclusions related to this analysis are relevant for our further considerations:

- The asymptotic MSE consists of two terms, the first one resulting from the additive noise and the second one from the frequency tracking error, i.e.

$$\Pi = \gamma(1-\alpha)^2\frac{\sigma_2^2}{2\sigma_0^2} + \frac{\sigma_1^2}{2\gamma}. \tag{16}$$

- The asymptotic MSE Π is minimized with respect to γ when the above two terms are equal, giving the following optimal value

$$\gamma_{opt} = \frac{\sigma_0\sigma_1}{\sigma_2(1-\alpha)} \tag{17}$$

$$\Pi\,|_{\gamma=\gamma_{opt}} = \gamma_{opt}(1-\alpha)^2\frac{\sigma_2^2}{\sigma_0^2} = (1-\alpha)\frac{\sigma_1\sigma_2}{\sigma_0}. \tag{18}$$

- These results are valid if α satisfies the following relations:

$$\frac{\sigma_0\sigma_1}{\sigma_2} \ll 1-\alpha \ll 1 \tag{19}$$

$$\Pi \ll (1-\alpha)^2. \tag{20}$$

  When these inequalities are not satisfied, (16) is no longer a good approximation of the MSE, resulting in practice in poor filtering, mistracking, locking to false frequencies, or similar effects [8], [9], [18], [13].

In other words, if the adaptive notch filtering is basically successful, then the minimization of the frequency tracking MSE with respect to ρ is analytically possible. However, formal minimization of (16) with respect to α is contradictory to the basic assumptions in [8] and [9] so that (16) just shows that α should be as high as possible yet still in agreement with (19) and (20). Thus, such an approach does not provide any possibility to determine α optimally.

C. Jointly Optimal Tuning Variables

Our approach to the optimization of RPE algorithms for frequency tracking is based on the above introduced two criteria: P defined by (10) and Π defined by (16).

Replacement of Π in (10) by the optimal value defined by (18) gives

$$P\,|_{\gamma=\gamma_{opt}} = \frac{\sigma_0\sigma_1\sigma_2}{1-\alpha} + (2-\alpha)\sigma_2^2 \tag{21}$$

which is minimal for

$$\alpha_{opt} = 1 - \sqrt{\frac{\sigma_0\sigma_1}{\sigma_2}}. \tag{22}$$

Thus

$$P\,|_{\gamma=\gamma_{opt},\alpha=\alpha_{opt}} = \sigma_2^2 + 2\sqrt{\sigma_0\sigma_1\sigma_2} \tag{23}$$

and, by using (22), one comes to

$$\Pi\,|_{\gamma=\gamma_{opt},\alpha=\alpha_{opt}} = \sigma_1^2\sqrt{\frac{\sigma_2}{\sigma_0\sigma_1}} = (1-\alpha_{opt})^3\frac{\sigma_2^2}{\sigma_0^2}. \tag{24}$$

Comparing (18) and (24), one obtains that

$$\alpha_{opt} = \rho_{opt}. \tag{25}$$

A few remarks concerning the optimal value of α (22) should be made:

- As minimization of P is essentially concerned with correct notch filtering, the conditions (19) and (20) should be automatically satisfied for α_{opt}. Indeed, α_{opt} turns out to be, according to (22), the geometric mean of the bounds defined in (19). The second inequality (20) is certainly satisfied by α_{opt} since the optimal MSE is proportional to $(1-\alpha)^3$ (see (24)).
- The result (25) is not surprising since the roles of α and ρ in RPE-type ANF's are similar. They should be lower in the case of faster frequency changes and higher in the case of slowlier frequency changes, as expected on the basis of experimental evidences and a qualitative argumentation [6], [8], [9].
- An *a priori* knowledge on signal characteristics is required for the optimal selection of α and ρ. Usually, σ_1 and σ_2/σ_0 are unknown or time-varying, limiting possibilities of direct applications.

IV. ADAPTIVE α AND ρ

The above analysis allows us to determine the pole contraction factor α in the ANF in an adaptive way, using current measurements of the input signal. The inclusion of the adaptation of α can be considered to be a natural extension of the basic concept of the ANF. Indeed, the estimation of both free parameters (ω and α) in CNF represented by (2) leads to a fully adaptive second-order notch filter. Its application could be advantageous not only when there is no *a priori* information about the signal properties, but also, and especially, under unpredictable nonstationarities.

A. RPE Algorithm for Estimation of α

A straightforward way of making α data adaptive is to minimize the mean-square prediction error by a stochastic gradient-type procedure [4]. Starting from $e(i)$ expressed as a function of α, one comes to the following RPE-type recursive estimation algorithm:

$$\hat{\alpha}(i+1) = \hat{\alpha}(i) + \gamma_{\alpha_i} R_\alpha(i)^{-1}\psi_\alpha(i)e(i) \tag{26}$$

$$R_\alpha(i) = R_\alpha(i-1) + \gamma_{\alpha_i}(\psi_\alpha(i)^2 - R(i-1)) \tag{27}$$

where

$$\begin{aligned}\psi_\alpha(i) &= -\frac{\partial e(i)}{\partial \alpha}\Big|_{\alpha=\hat{\alpha}(i), a=\hat{a}(i)} \\ &= \frac{\hat{a}(i)q^{-1} + 2\hat{\alpha}(i)q^{-2}}{1+\hat{\alpha}(i)\hat{a}(i)q^{-1} + \hat{\alpha}(i)q^{-2}}e(i)\end{aligned} \tag{28}$$

and $\{\gamma_{\alpha_i}\}$ is a suitably chosen sequence of positive numbers. The above algorithm for estimating α can be obtained, together with the main algorithm for frequency estimation, from a general gradient-type RPE algorithm for estimating the vector $\{a, \alpha\}$ composed of both the unknown parameters in the notch filter (2). Both algorithms (11)–(15) and (26)–(28) can be obtained from this algorithm by decoupling, i.e., by neglecting the off-diagonal terms in the corresponding weighting matrix generated, according to the general methodology, by the vector of regressors $\{\psi(i), \psi_\alpha(i)\}$ [4] ($R(i)^{-1}$ and $R_\alpha(i)^{-1}$ represent the diagonal terms of this matrix). The decoupling effect can be emphasized by choosing $\gamma_{\alpha_i} < \gamma_i$, aiming at making the adaptation of α slower than that of the notch frequency.

The obtained ANF algorithm is, obviously, highly nonlinear and difficult for any rigorous analysis. An insight into its asymptotic behavior can be obtained, however, by using the ODE methodology [4]. This methodology has been used in [1] and [2] to determine possible convergence points of the RPE algorithm for constant frequency estimation (11)–(13), supposing that α is a constant parameter. In the above context, when the frequency is time varying, it makes sense to freeze the estimate $\hat{a}(i)$ and inspect possible convergence points α^* of $\hat{\alpha}(i)$ generated by (26)–(28). According to [4], we have, under mild conditions

$$f(\alpha^*) = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} E\{\bar{\psi}_\alpha(i)\bar{e}(i)\} = 0 \tag{29}$$

where $\bar{\psi}_\alpha(i)$ and $\bar{e}(i)$ are obtained from $\psi_\alpha(i)$ and $e(i)$ by freezing $\hat{a}(i)$ at some a and $\hat{\alpha}(i)$ at α^*. After a straightforward, but lengthy derivation, (29) gives rise to

$$(1-\alpha^*)^3 = 2\delta^{*2}\frac{\sigma_0^2}{\sigma_2^2} \tag{30}$$

where δ^* represents a frozen, sufficiently small value of the asymptotic tracking error. The complete derivation is given in Appendix B. A direct comparison with the expression (5) shows that the algorithm (26)–(28) asymptotically minimizes the mean-square prediction error with respect to α. This represents a justification of the proposed approach to ANF and makes a connection with the analysis given in the preceding section.

Characteristic features of the proposed method will be illustrated in the next section. It is to be noticed here that special care should be taken of the stability of filters generating $\psi(i)$ and $\psi_\alpha(i)$. An obvious way to ensure stability is to apply some projection rule providing $\alpha_{min} \le \hat{\alpha}(i) \le \alpha_{max}$, where $0 < \alpha_{min} < \alpha_{max} \approx 1$.

B. Adaptation of ρ

The introduction of an adaptive contraction factor α contributes, in principle, to frequency tracking, but it is not sufficient to ensure an overall high performance of the corresponding ANF. Numerous experiences have shown that the algorithm is highly sensitive to the choice of the sequence $\{\gamma_i\}$ (or of the forgetting factor ρ when $\gamma_i = \gamma = const$).

The important general problem of the forgetting factor determination in RPE algorithms for system identification has attracted the attention of many researchers. Numerous more or less efficient solutions have been analyzed; however, no general methodology has been proposed [4], [14]. In our specific situation, it is possible to start from the result derived in Section III-C, claiming that the optimal values of α and ρ are equal. This indicates that the choice of γ (or ρ) can be directly related to the sequence $\hat{\alpha}(i)$ of the

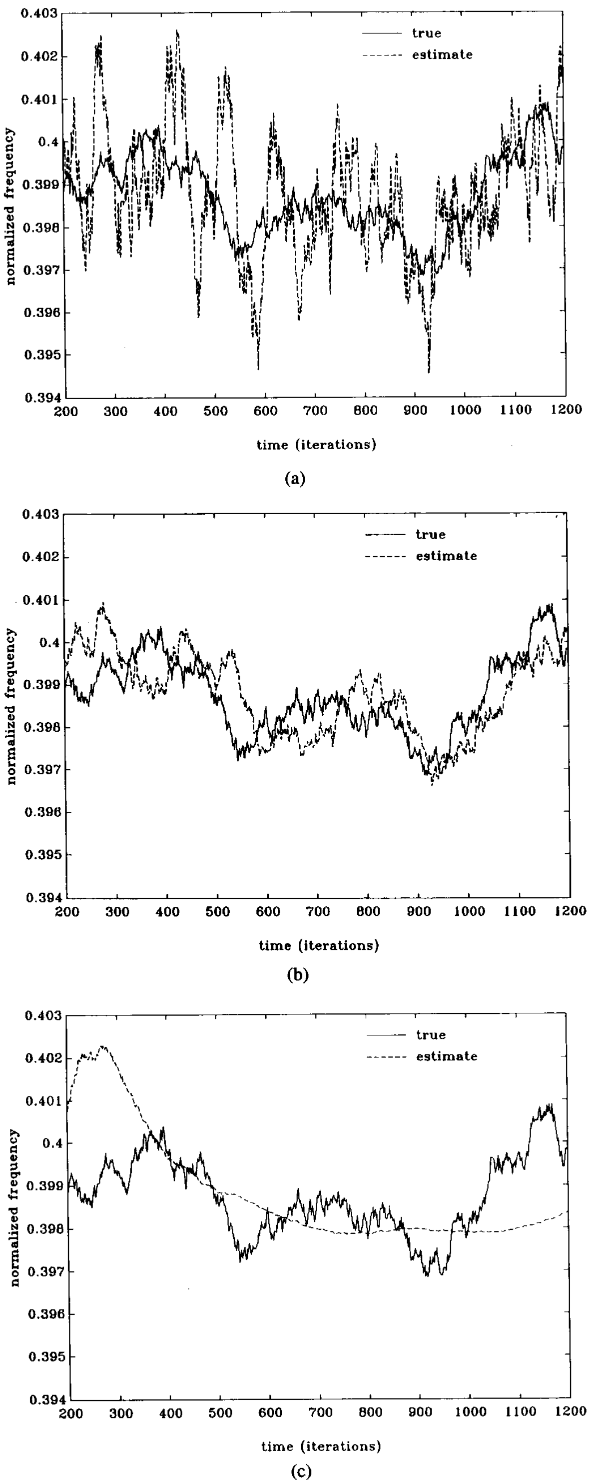


Fig. 1. Influence of α and ρ on frequency tracking: (a) $\alpha = \rho = 0.95$; (b) $\alpha = \rho = 0.975$ (optimal); (c) $\alpha = \rho = 0.995$.

pole contraction factor estimates. However, the idea to equate the estimates of α and ρ directly has not been found to be very effective as a consequence of a high sensitivity of the algorithm performance to the choice of ρ (keeping in mind that $\hat{\alpha}(i)$ may vary relatively too quickly). It has been found that a very good performance can be obtained simply by smoothing $\hat{\alpha}(i)$, retaining the main trends, and rejecting random spikes.

In such a way, an ANF algorithm with the adaptive frequency ω, contraction factor α, and forgetting factor ρ is obtained, requiring no *a priori* knowledge about the signal properties. The only parameter in the algorithm that should be chosen in advance is the forgetting factor $\rho_\alpha = 1 - \gamma_\alpha$ in the algorithm for the estimation of α. All the related experiences have shown that the sensitivity to this parameter is low, provided it is kept at a value close to 1 (which is justified by the essentially different time constants of the algorithms for frequency estimation and contraction factor adaptation).

Obviously, the resulting algorithm can be considered to be a special case of a general IIR ALE applicable to the time-varying case. It is essentially based on imposition of a specific constrained ANF structure and generation of all free parameters in an adaptive way. However, the main underlying motivation has not been to investigate possible generalizations in IIR ALE but, rather, to achieve an improvement of tracking capabilities of the existing popular ANF algorithms. Notice that tracking properties of general IIR ALE are far from being completely understood in the general time-varying case.

V. Computer Simulations

The above theoretical results, as well as the main properties of the proposed algorithm, will be illustrated on characteristic examples.

A. Single Sinusoid

In the first example, the input signal has been in the form of (7), where $U = 2\sqrt{2}$ and $\sigma_2 = 1$, whereas the time-varying frequency has been generated by the random walk model (6), where $\nu(i)$ has been a unit variance white noise with $\sigma_1 = \pi 10^{-4}$.

The first group of figures is presented as an illustration and experimental justification of the theoretical results concerning the influence of the parameters α and ρ to the performance of the frequency tracking RPE algorithm (11)–(13). The results presented in Fig. 1 have been obtained by (11)–(13) with different values of α and ρ, under the condition $\alpha = \rho$; in Fig. 1(a), $\alpha = \rho = 0.95$, in Fig. 1(b), $\alpha = \rho = 0.975$, and in Fig. 1(c), $\alpha = \rho = 0.995$. The best propertie are, obviously obtained for $\alpha = \rho = 0.975$, which represent the optimal values according to (22) and (24). Fig. 2 gives an illustration of the influence of α itself for a fixed value of ρ. Fig. 2(a) is obtained with $\alpha = 0.95$ and Fig. 2(b) with $\alpha = 0.995$; in both cases, ρ has been kept at its optimal value $\rho = 0.975$. In both cases, the performance is far inferior to that shown in Fig. 1(b). Notice that satisfactory tracking can be obtained only for α belonging to a small interval around its optimal value (even for the optimal value of ρ). This contradicts some claims that the performance of RPE frequency tracking algorithms is not very sensitive to the choice of α [6], [8], [9].

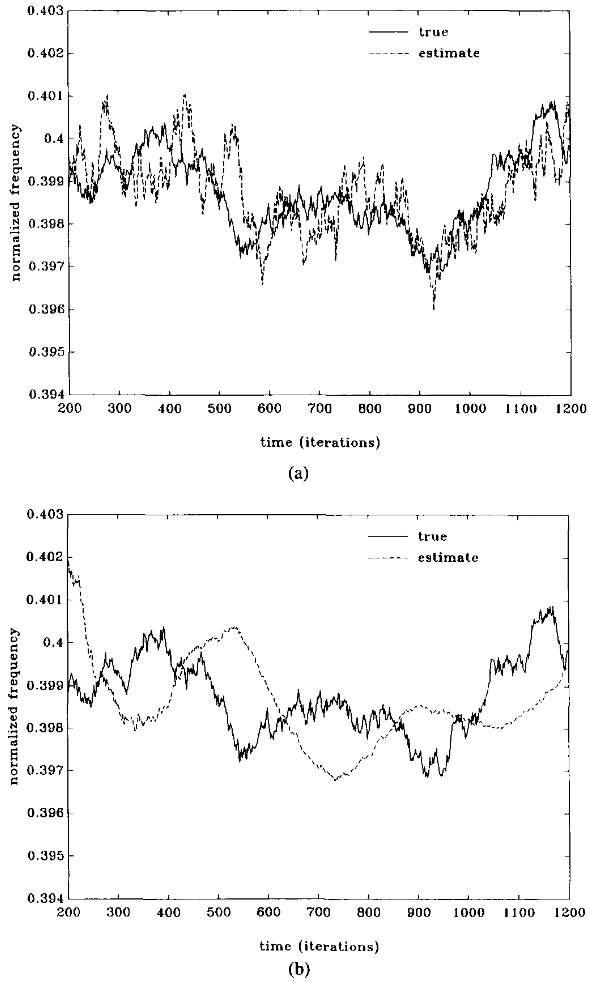


Fig. 2. Influence of α and ρ to frequency tracking for optimal $\rho = 0.975$: (a) $\alpha = 0.95$; (b) $\alpha = 0.995$.

Fig. 3 shows the properties of the proposed ANF with adaptive α and ρ. In Fig. 3(a), ρ has been kept at its optimal value $\rho = 0.975$, whereas α has been generated adaptively by the algorithm (26)–(28), with $\rho_\alpha = 0.99$. In Fig. 3(b), α has also been generated by (26)–(28), but ρ has been adapted to the data as well. Following the general conclusions exposed in Section IV-B, $\hat{\rho}(i)$ has been obtained from $\hat{\alpha}(i)$ by using a simple smoothing filter defined by

$$\hat{\rho}(i) = 0.995\hat{\rho}(i-1) + 0.005\hat{\alpha}(i). \tag{31}$$

It has been found that a more precise analytical treatment of the connections between $\hat{\alpha}(i)$ and $\hat{\rho}(i)$ is hardly feasible, keeping in mind, in the first place, that two different criteria are used in Sections III-B and C to derive α_{opt} and ρ_{opt}. In addition, no clear connections between $\hat{\alpha}(i)$ and the filter dynamics have been observed. First-order filters have given satisfactory results in almost all the experiments. Moreover, the choice of the pole value in (31) has not been found to be critical, as long

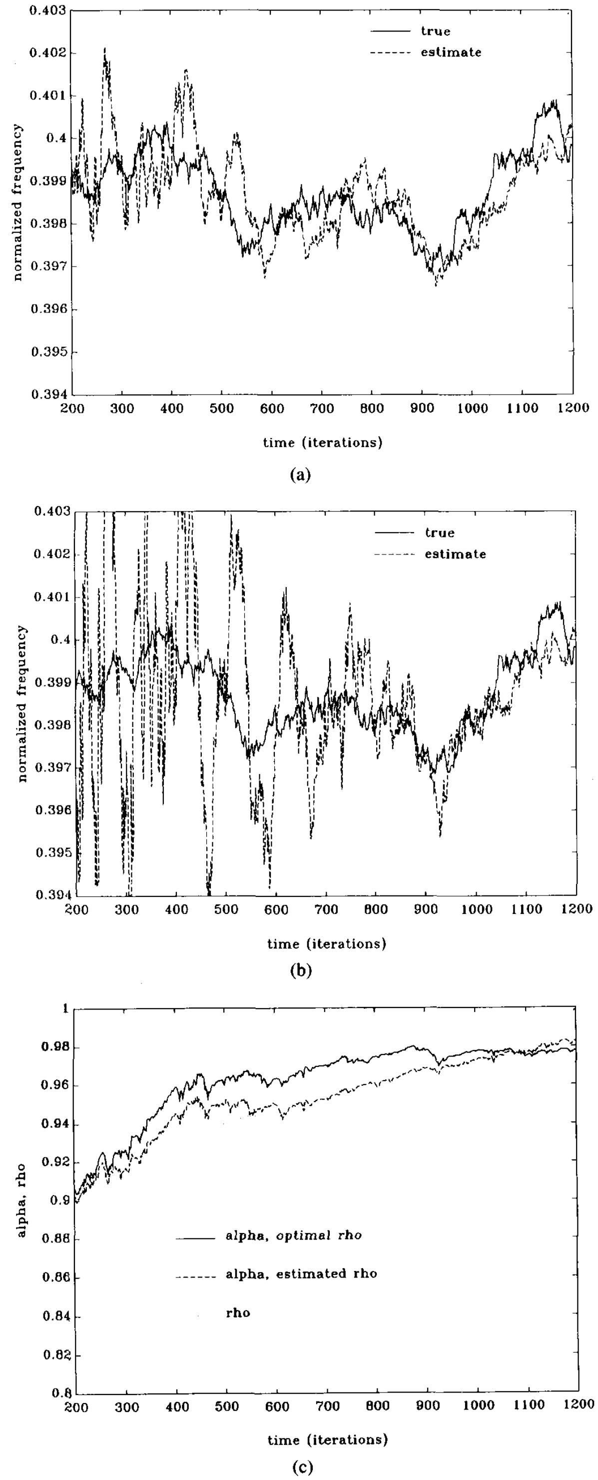


Fig. 3. Random walk frequency tracking: (a) $\rho = 0.975$; (b) ρ is adaptive; (c) estimates of α and ρ.

as $\hat{\rho}(i)$ follows smoothly the main trends of $\hat{\alpha}(i)$. Fig. 3(c) shows the corresponding estimates of α and ρ. In both cases, convergence of $\hat{\alpha}(i)$ to its optimal value $\alpha = 0.975$ is achieved

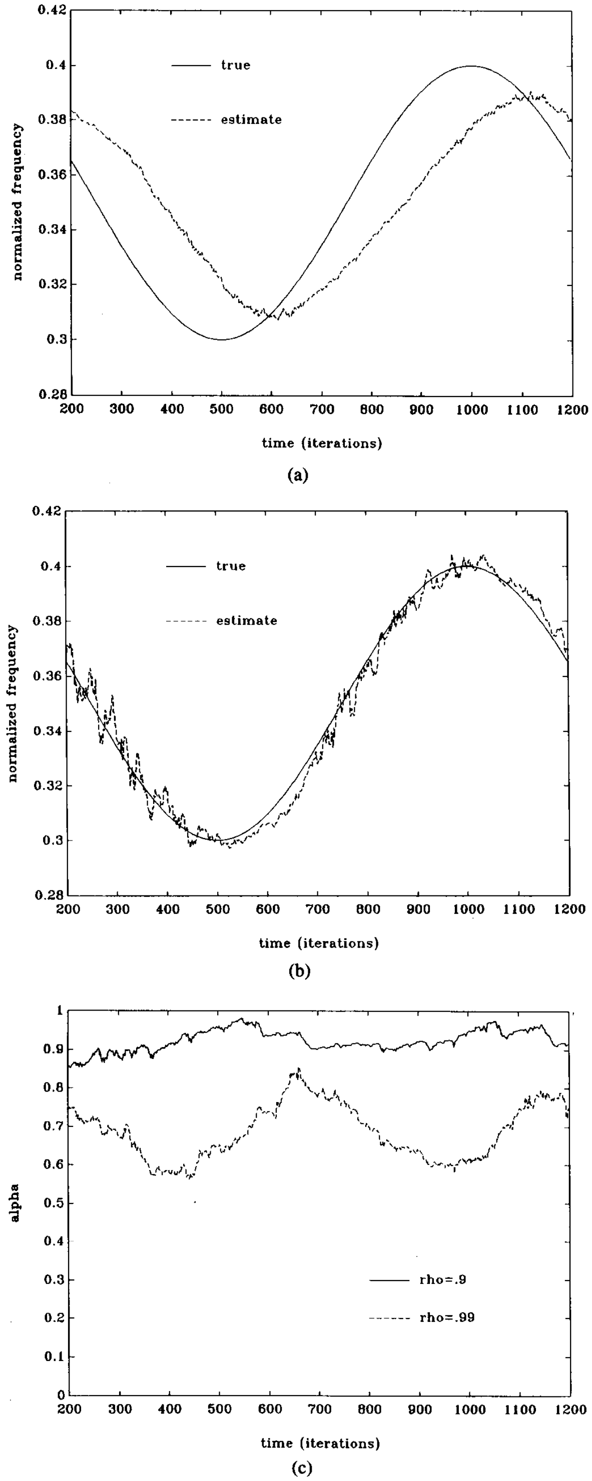


Fig. 4. FM signal frequency tracking with adaptive α: (a) $\rho = 0.99$; (b) $\rho = 0.9$; (c) estimates of α.

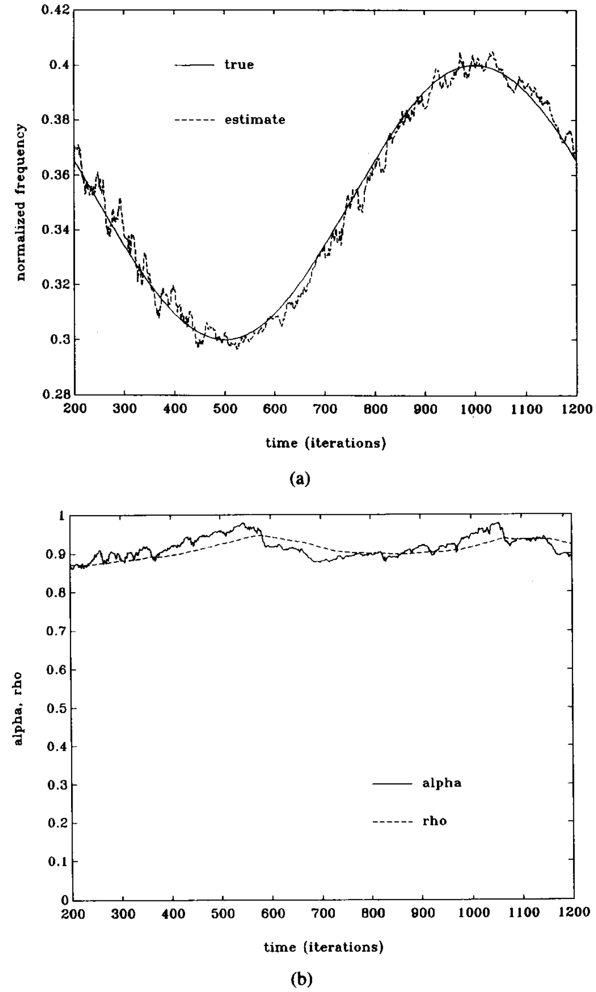


Fig. 5. FM signal frequency tracking with adaptive α and ρ.

exactly in accordance with the theoretical analysis presented in Section III-C. The adaptation period can be clearly seen in Fig. 3(a) and (b) (compare with Fig. 1(b), where both α and ρ have been optimal).

In the second example, the input signal has also been generated by (7), with $U = 2\sqrt{2}$ and $\sigma_2 = 1$, but the frequency $\omega_0(i)$ has followed a sinusoidal trajectory.

In Fig. 4(a), frequency tracking is shown when α has been generated by (26)–(28) with $\rho_\alpha = 0.995$, whereas ρ has been kept at $\rho = 0.99$; in Fig. 4(b), $\rho = 0.9$. Fig. 4(c) shows the corresponding estimates $\hat{\alpha}(i)$. It can be seen that $\hat{\alpha}(i)$ is lower on average for higher values of ρ when we try to speed up tracking. This effect can be explained by observing that for any given ρ, the optimal α_0 minimizing P is defined by

$$(1-\alpha_0)^3 = \frac{\sigma_1^2\sigma_0^2}{(1-\rho)\sigma_2^2}.$$

It can also be observed that $\hat{\alpha}(i)$ is well adapted to the current rate of frequency changes. In Fig. 5, the results obtained with adaptation of both α and ρ are shown (algorithm (11)–(13) has been applied together with (26)–(28)). The initial estimate of

ρ has been $\hat{\rho}(0) = 0.99$ (Fig. 4(a) is obtained with this value). The effect of adapting ρ is evident (compare with Fig. 4(a) and (c).The time variations of both $\hat{\alpha}(i)$ and $\hat{\rho}(i)$ contribute to the improvement of tracking performance since they reflect the rate of frequency changes.

The influence of the initial parameter estimates has also been carefully investigated. It has been found that the choice $\hat{\alpha}(0) = 0$ is appropriate for low SNR's, especially when the true frequency is close to 0 or π. However, in all the above examples, $\hat{\alpha}(0) = 0.8$ has been successfully applied.

A prevention from instability and an eventual locking at false frequencies has been achieved by applying simple projection rules consisting, e.g., of resetting $\hat{\alpha}(i)$ to 0.8 and 0.2 when it becomes greater than 1 and lower than 0, respectively. However, in typical examples, projection rules have hardly ever been activated. For extremely low SNR's, when $\hat{\alpha}(i)$ keeps exceeding 1, the upper bound of the projection region, being much more critical, has been decreased further, which has a beneficial effect of preventing locking at spurious frequencies.

B. Multiple Sinusoids

The cascade form of the CNF's represented in Section II has been extensively discussed in the literature [1], [15], [16], [13], [18], [10]. One of the main advantages of such a model in the above context lies in the possibility to implement adaptation of different stages separately. Separate adaptation of the notch frequencies has been proposed in [15] and [18]. Following the line of thought exposed above, it is possible to implement completely adaptive schemes for estimation of both contraction and forgetting factors of second-order ANF's within the cascade form. This becomes especially important when the frequencies in the input signal are time-varying, following different dynamics; one overall contraction factor for the whole denominator cannot be suitable in this case. Obviously, a multistage ANF for tracking multiple frequencies, which includes adaptation of both contraction and forgetting factors, is composed of a series of algorithms (11)–(13) and (26)–(28). In the simplest configuration, the prediction error of a given stage becomes the input of the following stage (see [17] and [15] for the discussion on possible algorithm structures). In order to demonstrate that the extension of the proposed approach to the case of tracking of multiple sinusoids is straightforward and effective by using the cascade form, we have generated a signal composed of one sinusoid with constant frequency, one chirp signal, and white noise (see Fig. 6(a)) and applied the proposed algorithm composed of two cascades, including adaptation of both contraction and forgetting factors. The obtained results are depicted in Fig. 6. It should be noticed that the cascade following the chirp signal generates α, tending to a value lower than that of the cascade estimating the constant frequency, where α tends to one. Such an effect is desirable and represents one of the main motives for decomposing the adaptation algorithm into stages. Notice also that the adaptation algorithms behave very regularly once a decision is made as to which frequency should be estimated within the given stage.

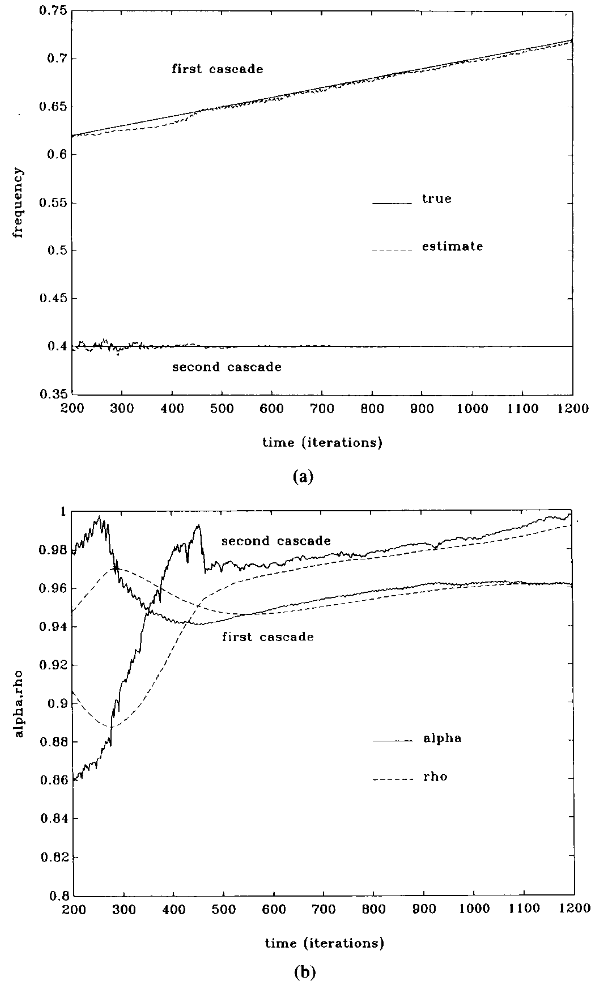


Fig. 6. Chirp signal frequency tracking by two-stage ANF with adaptive α and ρ.

VI. CONCLUSION

This paper presents two main contributions to the analysis and design of the techniques for time-varying frequency tracking based on adaptive notch filtering. The first is related to the analysis of ANF performance and optimization of its behavior with respect to its free design parameters: the pole contraction factor of the filter and the forgetting factor of the recursive prediction error algorithm for frequency estimation. The basic prerequisite has been the derivation of a new expression for the ANF output power. The results are in agreement with all practical experiences and encompass the existing results as special cases. The second contribution is the proposal of a new ANF algorithm based on the estimation of both notch frequency and pole contraction factor by an algorithm of recursive prediction error type. Moreover, the adaptation has been extended to the forgetting factor of the estimation algorithm on the basis of the given optimality study. An asymptotic analysis based on the ODE methodology confirms the validity of the approach. The resulting fully adaptive algorithm represents an efficient tool in nonstationary conditions when no *a priori* knowledge about signal characteristics is available.

A direct line of continuation of the described investigations can be oriented toward applying the obtained results to higher order ANF in the cascade form, as shown in the paper.

Appendix A
ANF Output Power

We shall start from the time-varying CNF model expressed as

$$(1+\alpha a(i)q^{-1}+\alpha^2q^{-2})e(i) = (1+a(i)q^{-1}+q^{-2})y(i) \quad (32)$$

where $e(i)$ is the output and $y(i)$ the input, and derive an expression for the output power P when $y(i) = U\cos(\sum_{j=0}^{i}\omega_0(i)) + \sigma_2\xi(i)$, where $\xi(i)$ is the unit variance zero-mean white noise. This result will be applicable to both constant and slowly time-varying frequency cases (see (4) and (10)).

When only the sinusoidal term is present in the input, the corresponding output can be represented in general as

$$e(i) = E_1(i)\cos(\sum_{j=0}^{i}\omega_0(j)) + E_2(i)\sin(\sum_{j=0}^{i}\omega_0(j)). \quad (33)$$

Inserting (33) into (32) and equating the corresponding terms, one obtains

$$E_1(i) \approx E_2(i)\frac{(1-\alpha)^2\cos\omega_0(i) + 2\alpha\delta(i)\sin\omega_0(i)}{(1-\alpha^2)\sin\omega_0(i)} \quad (34)$$

and (35), which appears at the bottom of the page, where $\delta(i) = \omega(i) - \omega_0(i)$, $\omega(i) = \arccos(-\frac{a(i)}{2})$, assuming that $\delta(i)^2$ and $\omega_0(i)-\omega_0(i-1)$ are small enough. After neglecting higher order terms in $1-\alpha$, one obtains $E_2(i) \approx \frac{\delta(i)}{1-\alpha}U$, and $E_1(i) \approx (\frac{\delta(i)}{1-\alpha})^2U$, and the first terms in (4) and (10) follow directly.

The response of the time-varying CNF to white noise will be expressed as $e(i) = \sum_{j=0}^{\infty} c_j(i)\xi(i-j)$, where $\{c_j(i); j = 0, 1, ...\}$ is the time-varying impulse response of ANF given by

$$c_j(i) = h_j(i) + a(i-j+1)h_{j-1}(i) + h_{j-2}(i) \quad (36)$$

$$h_j(i) = \begin{cases} 0 & j<0 \\ 1 & j=0\,. \\ -\alpha a(i-j+1)h_{j-1}(i) - \alpha^2 h_{j-2}(i) & j>0 \end{cases} \quad (37)$$

Then $E\{e(i)^2\} = \sigma_2^2\sum_{j=0}^{\infty}c_j(i)^2$ can be expressed as

$$E\{e(i)^2\} = \sigma_2^2 + \sigma_2^2(1-\alpha)^2(F(i) + 2(1+\alpha)G(i) + (1+\alpha)^2H(i)) \quad (38)$$

where

$$F(i) = \sum_{j=0}^{\infty} a(i-j)^2h_j(i)^2$$
$$G(i) = \sum_{j=0}^{\infty} a(i-j)h_j(i)h_{j-1}(i)$$
$$H(i) = \sum_{j=0}^{\infty} h_j(i)^2.$$

Multiplying (37) with $a(i-j+1)h_{j-1}(i)$ and summing from $j=1$ to ∞, one obtains for small variations of $a(i)$ that

$$G(i) \approx -\frac{\alpha}{1+\alpha^2}F(i). \quad (39)$$

On the other hand, squaring (37) and summing from $j = 1$ to ∞, one gets

$$H(i) - 1 = \alpha^2F(i) + 2\alpha^3G(i) + \alpha^4H(i) \quad (40)$$

or, taking into account (39)

$$(1+\alpha^2)^2H(i) - \alpha^2F(i) \approx \frac{1+\alpha^2}{1-\alpha^2}. \quad (41)$$

Substituting (39) into (38), one obtains

$$E\{e(i)^2\} \approx \sigma_2^2 + \sigma_2^2\frac{(1-\alpha)^2}{1+\alpha^2}\{\frac{(1+\alpha)^2}{1+\alpha^2}[(1+\alpha^2)^2H(i) - \alpha^2F(i)] + \frac{(1-\alpha^2)^2}{2\alpha}F(i)\}. \quad (42)$$

The last term in (42) is proportional to $1-\alpha$, keeping in mind that $F(i)$ is proportional to $\frac{1}{1-\alpha}$ due to the exponential stability of the filter (37) for $\alpha < 1$. The term in the square brackets is proportional to $\frac{1}{1-\alpha}$ due to (41) and, therefore, is dominant. As a result, one gets

$$E\{e(i)^2\} \approx \sigma_2^2[1 + (1-\alpha)\frac{1+\alpha}{1+\alpha^2}] \approx \sigma_2^2(2-\alpha). \quad (43)$$

In such a way, the second term in both (4) and (10) is obtained.

Appendix B
Convergence Points of $\hat{\alpha}(k)$

It follows from (13) and (28) that

$$\bar{e}(i) = N^*(q)y(i) \quad (44)$$

$$\bar{\psi}_\alpha(i) = \frac{aq^{-1} + 2\alpha^*q^{-2}}{1+\alpha^*aq^{-1}+\alpha^{*2}q^{-2}}\bar{e}(i) \approx -\frac{1}{1-\alpha^*}(1-N^*(q))N^*(q)y(i) \quad (45)$$

$$E_2(i) \approx U\frac{2\delta(i)(1-\alpha^2)\sin^2\omega_0(i)}{(1-\alpha^2)^2\sin^2\omega_0(i) + ((1-\alpha)^2\cos\omega_0(i) + 2\alpha\delta(i)\sin\omega_0(i))^2} \quad (35)$$

where $N^*(q)$ is obtained from (2) for $\alpha = \alpha^*$ and $\omega = \arccos(-\frac{a}{2})$. As $y(i)$ is composed of the sinusoidal term and noise, it follows that accordingly, $f(\alpha^*) = f_{sin}(\alpha^*) + f_{noise}(\alpha^*)$.

It is straightforward to conclude that

$$f_{sin}(\alpha^*) \approx -\frac{\delta^2\sigma_0^2}{(1-\alpha^*)^3} \tag{46}$$

keeping in mind that the amplitude of $\bar{e}(i)$ is approximately $\frac{\delta}{1-\alpha^*}U$ (see Appendix A) and that the transfer function $1-N^*(q)$ (line enhancer) changes neither amplitude nor phase of the input sinusoid for $\delta \ll 1-\alpha^*$.

The second term $f_{noise}(\alpha^*)$ will be derived directly after representing the transfer functions in (44) and (45) by their impulse responses, i.e., $\bar{e}(i) = \sum_{j=0}^{\infty} c_j\xi(i-j)$ and $\bar{\psi}_\alpha(i) = \sum_{j=0}^{\infty} d_j\bar{e}(i-j)$. From Appendix A, c_j is given by (36) for $a(i) = a$, whereas

$$d_j = ah_{j-1} + 2\alpha^* h_{j-2}. \tag{47}$$

Consequently

$$\begin{aligned} f_{noise}(\alpha^*) &= E\{\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} d_j c_k c_l \xi(i-j-k)\xi(i-l)\} \\ &= \sigma_2^2 \sum_{j=1}^{\infty} d_j(c_j + \sum_{k=1}^{\infty} c_k c_{j+k}). \end{aligned} \tag{48}$$

After replacing c_j and d_j, one obtains

$$\begin{aligned} &f_{noise}(\alpha^*) \\ &= \sigma_2^2(1-\alpha^*)((2\alpha^*(1+\alpha^*)+a^2)H_0 + (1+3\alpha^*)aH_1) \\ &+ \sigma_2^2(1-\alpha^*)^2\sum_{j=1}^{\infty}(ah_{j-1}+2\alpha^*h_{j-2}) \\ &(((1+\alpha^*)^2+a^2)H_j + (1+\alpha^*)aH_{j-1} + (1+\alpha^*)aH_{j+1}) \end{aligned} \tag{49}$$

where $H_k = \sum_{j=0}^{\infty} h_j h_{j+k}$. Notice that H_0 and H_1 are related to $F(i)$, $G(i)$, and $H(i)$ defined in Appendix A by

$$H_0 = H(i) = \frac{F(i)}{a^2} = \frac{1+\alpha^{*2}}{1-\alpha^{*2}}\frac{1}{(1+\alpha^{*2})^2 - \alpha^{*2}a^2} \tag{50}$$

$$H_1 = \frac{G(i)}{a} = -\frac{\alpha^* a}{1+\alpha^{*2}}H_0 \tag{51}$$

for $a(i) = a$ and $\alpha = \alpha^*$ (see (39) and (41)). For $k \geq 2$, one has the following recursion:

$$H_k = -\alpha^* aH_{k-1} - \alpha^{*2}H_{k-2} \tag{52}$$

which follows from (37) after multiplying by h_{j+k} and summing from $j = 0$ to ∞. The resulting expressions show that $\{H_k\}$ is the impulse response of the following transfer function:

$$H_0\frac{1+\frac{\alpha^{*3}a}{1+\alpha^{*2}}q^{-1}}{1+\alpha^* aq^{-1} + \alpha^{*2}q^{-2}}$$

where from we have

$$H_k = H_0(h_k + \frac{\alpha^{*3}a}{1+\alpha^{*2}}h_{k-1}) \tag{53}$$

since $\{h_k\}$ is the impulse response of $(1+\alpha^* aq^{-1} + \alpha^{*2}q^{-2})^{-1}$. It follows that

$$\sum_{j=0}^{\infty} h_j H_{j+k} = H_0(H_k + \frac{\alpha^{*3}a}{1+\alpha*2}H_{k-1}). \tag{54}$$

Replacing (54) and (50) in (49), one obtains an expression where all the factors proportional to $1-\alpha^*$ are canceled by the factors proportional to $(1-\alpha^*)^{-1}$, which means simply, for $\alpha^* \approx 1$, that $f_{noise}(\alpha^*) = \frac{\sigma_2^2}{2}$. Equation (30), defining the convergence points of $\hat{\alpha}(i)$, now follows directly.

References

[1] D. V. B. Rao and S. Y. Kung, "Adaptive notch filtering for the retrieval of sinusoids in noise," *IEEE Trans. Acoust, Speech, Signal Processing*, vol. ASSP-32, pp. 791–802, 1984.

[2] A. Nehorai, "A minimal parameter adaptive notch filter with constrained poles and zeros," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 983–996, 1985.

[3] T. S. Ng, "Some aspects of adaptive digital notch fiter with constrained poles and zeros," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 158–161, 1987.

[4] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification.* Cambridge, MA: MIT Press, 1983.

[5] M. Dragošević, S. Stanković and M. Čarapić, "An approach to recursive estimation of time-varying spectra," in *Proc. IEEE ICASSP '82*, Paris, 1982, pp. 2080–2083.

[6] B. D. Rao and R. Peng, "Tracking characteristics of the constrained IIR adaptive notch filter," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1466–1479, 1988.

[7] R. V. R. Kumar and R. U. Pal, "Tracking of bandpass signals using center-frequency adaptive filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1710–1721, 1990.

[8] P. Händel and A. Nehorai, "Tracking analysis of a constrained adaptive notch filter," in *Proc. IEEE ICASSP 1991*, Toronto, 1991, pp. 3209–3212.

[9] ——, "Tracking analysis of an adaptive notch filter with constrained poles and zeros,"*IEEE Trans. Signal Processing*, vol. 42, pp. 281–291, 1994.

[10] J. M. Travassos-Romano and M. Béllanger, "Fast adaptive notch filtering in the cascade form," in *Proc. EUSIPCO 1988*, Grenoble, 1988, pp. 567–569.

[11] J. F. Chicharo and T. S. Ng, "Gradient-based adaptive IIR notch filtering for frequency estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 769–777, 1990.

[12] P. Stoica and A. Nehorai, "Performance analysis of an adaptive notch filter with constrained poles and zeros," *IEEE Trans. Acoust., Speech, Signal Processing*, pp. 911–919, 1988.

[13] M. Dragošević and S. Stanković, "Time-varying frequency tracker with adaptive pole contraction," in *Proc. IEEE ICASSP 1992*, San Francisco, 1992, pp. IV 181–184.

[14] L. Ljung and S. Gunnarsson, "Adaptation and tracking in system identification—A survey," *Automatica*, vol. 26, pp. 7–21, 1990.

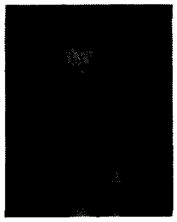
[15] N. K. M'Sirdi and H. A. Tjokronegoro, "Cascaded adaptive notch filters: an RML estimation algorithm," in *Proc. EUSIPCO 1988*, Grenoble, 1988, pp. 1409-1412.

[16] N. K. M'Sirdi, H. R. Tjokronegoro, and I. D. Landau, "Adaptive comb filters implementation for harmonic signals," in *Proc. EUSIPCO 1988*, Grenoble, 1988, pp. 195–198.

[17] W. J. Shyn and J. Tsao, "A ring structured adaptive notch filter," *Proc. IEEE ICASSP 1992*, San Francisco, 1992, pp. IV 313-316.

[18] M. Dragošević, S. Jovanović, and S. Stanković, "Reduced order modeling of harmonic signals," in *Proc. 12th Colloque GRETSI*, Juan-les-Pins, 1989, pp. 157–160.

Marina V. Dragošević (M'95) was educated in Belgrade, Yugoslavia, where she received B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Faculty of Electrical Engineering, University of Belgrade, in 1979, 1984, and 1990, respectively. From 1980 to 1993, she was a research assistant, research associate, and scientific associate at Vinča Institute, Belgrade, Yugoslavia. Her research activities were in the areas of communications, digital and statistical signal processing, and pattern recognition. She was involved in the development of algorithms and computer systems for analysis and automatic recognition of radio signals. Since 1993, she has been with Iotec Inc., Dartmouth, Canada, where she is involved in radar and sonar signal processing. Her main research interests are in the area of parameter and state estimation and adaptive algorithms.



Srdjan S. Stanković was born in Belgrade, Yugoslavia, on April 27, 1945. He received the B.Sc., M.Sc., and Ph.D. degrees from the University of Belgrade, Belgrade, Yugoslavia, in 1968, 1971, and 1976, respectively. From 1968 to 1972, he was a research asssociate in the Electronics Laboratory, Vinča Insitute, Belgrade, Yugoslavia. From 1971 to 1972, he was a research fellow at the Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, the Netherlands. In 1972, he joined the Faculty of Electrical Engineering, University of Belgrade, where he is presently full professor, teaching graduate and postgraduate courses in automatic control and signal processing. During 1986, he was a research associate and visiting professor at the Santa Clara University, Santa Clara, CA. Besides a number of scientific papers and chapters in monographs, he has published a textbook entitled *Nonlinear Automatic Control Systems*. He has also been the leader of numerous projects for industry. His current research interests include aspects of stochastic and adaptive control, signal processing, and neural networks.