Tracking Analysis of an Adaptive Notch Filter with Constrained Poles and Zeros

Peter Händel, Member, IEEE, and Arve Nehorai, Fellow, IEEE

Abstract—This paper analyzes the asymptotic tracking properties of an adaptive notch filter (ANF) with pole-zero constraints [1] for the cancellation or retrieval of multiple time-varying sine waves in additive noise. The asymptotic mean square error (MSE) is analyzed using the methods of Ljung and Gunnarsson [2] when the variations in the underlying frequencies are assumed to be sufficiently small. Closed-form expressions for the MSE are derived as functions of the tuning variables of the algorithm. The results give insight into the operational properties of the algorithm and are used in order to minimize the MSE with respect to the tuning variables. Computer simulations confirm the validity of the derived results.

I. INTRODUCTION

In many applications of signal processing, it is desirable to eliminate or extract sine waves from observed data or to estimate their unknown frequencies. Since the frequencies often vary with time, it is useful to apply adaptive notch filters (ANF’s) that adapt their notch frequencies as a function of the observed time series. The subject of notch filtering has been widely studied, and the main references can be found, for example, in [1].

The ANF algorithm proposed in [1] uses a minimal number of parameters equal to the number of narrow-band signal components and has been analyzed for stationary sine waves in [3]. The results of [3] confirm the excellent properties of this algorithm for stationary processes. In particular, in this case, it was proved analytically that the accuracy of the algorithm estimates is nearly optimal, i.e., attains the Cramér-Rao bound asymptotically. In this article, see also [5], we study the properties of the ANF for nonstationary sine waves in noise, where the steady-state squared error is analyzed using the ideas of Ljung and Gunnarsson [2]; see also [4]. The two main difficulties in using the results of [2] to our problem are that the model is not of the linear-regression type and that the frequencies are not linearly related to the filter coefficients. The tracking problem of another constrained ANF has also been studied in [6], where the mean of the parameter estimates was studied using the ODE approach. However, the results of [6] cannot be used to calculate and minimize the mean square error (MSE).

Deriving approximate and simple closed-form results for the asymptotic MSE gives insight into the properties of the algorithm. The results will naturally depend on the properties of the observed time series and the user-chosen variables of the ANF. With a priori information about the observed data, it is possible to tune the filter for optimal steady-state operation. The ANF considered has two tuning variables: the forgetting factor \( \lambda \) arising from the algorithm and the pole contraction factor \( \rho \) arising from the special model structure used. In the case of stationary narrow-band signals, it is useful to let \( \lambda(t) \) and \( \rho(t) \) grow exponentially from user-chosen initial values to the steady-state values \( \lambda = 1 \) and \( \rho = 0.995 \), which yield the most accurate results [1]. However, in the case of tracking nonstationary parameters, it is common to choose a constant \( \lambda < 1 \) and change the value of \( \rho \) in order to increase the notch bandwidth of the ANF. The choice of \( \lambda \) and \( \rho \) is a tradeoff between noise sensitivity and tracking error, and the optimal way to choose them in order to minimize scalar measures of the MSE is investigated in this paper.

The outline of the article is as follows. Section II states the preliminaries and gives the background information. In Section III, the signal is modeled as a random walk, and the relation to the parameter vector is derived. Section IV deals with the recursive prediction error algorithm, and Section V presents the MSE analysis. First, the case of multiple sinewaves is considered, and an expression for the asymptotic MSE is derived. In the general case, the MSE is matrix valued, and therefore, a scalar quality measure is introduced and studied. In particular, the single sinewave case is considered and studied in detail. Afterwards, the validity of the present analysis is studied. In Section VI, numerical results that confirm the present analysis are presented. Finally, Section VII concludes the paper.

II. PROBLEM DEFINITION

Let the measured data consist of a known number of sinewaves in additive white noise, i.e.,

\[
    x(t) = \sum_{k=1}^{m} \alpha_k \sin(\Phi_k(t) + \phi_k) \quad (2.1)
\]

\[
    y(t) = x(t) + e(t) \quad t = 1, \cdots, N \quad (2.2)
\]

where in (2.1), the amplitudes \( \{\alpha_k\} \) are unknown constants, and the phases \( \{\phi_k\} \) are assumed to be unknown random variables uniformly distributed over the interval \([0, 2\pi]\). The instantaneous frequencies \( \{\omega_k(t)\} \) of the signal are defined as

\[
    \omega_k(t) \triangleq \Phi_k(t) - \Phi_k(t-1) \quad k = 1, \cdots, m \quad (2.3)
\]
where \( \{\omega_k(t)\} \) are unknown and time varying. Here, \( \omega_k \) is the normalized radian frequency \( \omega_k = 2\pi f_k/f_s \), where \( f_k \) is the true frequency, and \( f_s \) is the sampling frequency. The measurement noise \( \{\epsilon(t)\} \) is a zero-mean white noise of variance \( \sigma^2 \) and is assumed to be independent of the phases \( \{\phi_k\} \).

In the case of time-invariant frequencies, i.e., \( \Phi_k(t) = \omega_k t \), it is well known that the measurement \( y(t) \) obeys the degenerate ARMA equation

\[
A(q^{-1}, \theta_0)y(t) = A(q^{-1}, \theta_0)e(t)
\]

(2.4)

where \( q^{-1} \) is the unit delay operator \( q^{-1}y(t) = y(t - 1) \), etc. The monic polynomial \( A(q^{-1}, \theta_0) \) is of order \( n = 2m \), and its coefficients have a mirror symmetric form

\[
A(q^{-1}, \theta_0) \triangleq \sum_{k=1}^{m} \left( 1 - 2 \cos \omega_k q^{-1} + q^{-2} \right) \\
= 1 + a_1 q^{-1} + \cdots + a_m q^{-m} + \cdots + a_m q^{-n+1} + q^{-n}.
\]

(2.5)

It follows from (2.5) that the conjugated pairs of zeros \( \{\eta_l\} \) of \( A(z, \theta_0) \) are located on the unit circle at the sinewave frequencies

\[
\eta_l = \left\{ \begin{array}{c}
e^{j\omega_l} & l = 1, \ldots, m \\
e^{-j\omega_{m-l+1}} & l = m+1, \ldots, n. 
\end{array} \right.
\]

(2.6)

The parameter vector \( \theta \) is of minimal order and consists of the unique polynomial coefficients

\[
\theta = (a_1, \ldots, a_m)^T.
\]

(2.7)

The vector of true parameters is denoted \( \theta_0 \).

Consider the problem of whitening the data \( y(t) \). Since the roots of \( A(z, \theta_0) \) are located on the unit circle, the system description (2.4) must be slightly modified by a small contraction of the poles in order to achieve the minimally parameterized stable approximate whitening filter of \( y(t) \) [1], [7]

\[
e(t, \theta) = \frac{A(q^{-1}, \theta)}{A(q^{-1}, \theta_0)} y(t).
\]

(2.8)

The poles of the filter in (2.8) are slightly displaced towards the origin by the contraction factor \( \rho \), which is close to but smaller than unity. It was shown in [8] that the filter output (2.8) for the true parameters \( \theta_0 \) approximates the white noise \( e(t) \) up to an order \( O(1 - \rho) \), that is, \( E[e(t, \theta_0) - e(t)]^2 = O(1 - \rho) \), where \( O(x) \) denotes a term such that \( |O(x)/x| \) is bounded when \( x \to 0 \). As desired, the magnitude of the transfer function of this filter is approximately one everywhere, except at the true sinewave frequencies, where it is zero. The bandwidth (BW) (in radians) of the complex notches created by each pole-zero pair is approximately given by [1]

\[
\text{BW} = \pi (1 - \rho).
\]

(2.9)

Using the above properties, the time series \( y(t) \) can be written as an asymptotically stable one step predictor and nearly white additive noise, i.e.,

\[
y(t) = \hat{y}(t|\theta_0) + \epsilon(t, \theta_0)
\]

(2.10)

where it is clear that the predictor in (2.10) is given by

\[
\hat{y}(t|\theta_0) = \frac{A(q^{-1}, \theta_0) - A(q^{-1}, \theta_0)}{A(q^{-1}, \theta_0)} y(t).
\]

(2.11)

This time-invariant description of the sine-waves-in-noise process is later used in Section V for the small error analysis. Note that the class of models (2.10) does not include exactly the true description of the sine-waves-in-noise process, and therefore, estimates of the frequencies are asymptotically (for \( t \to \infty \)) slightly biased.

III. TIME-VARYING SIGNAL MODELING

This section deals with a time-varying description of the signal (2.1). Assume that the instantaneous frequencies are slowly varying according to the random walk model

\[
\omega_k(t) = \omega_k(t - 1) + \gamma_k v_k(t), \quad k = 1, \ldots, m
\]

(3.1)

where the scalars \( \{\gamma_k\} \) are assumed to be small (\( \ll 1 \)), and \( \{v_k(t)\} \) is a zero-mean white noise of unit variance and independent of \( \{v_j(t)\} \), \( j \neq k \), and independent of the phases \( \{\phi_k\} \) and the measurement noise \( \{\epsilon(t)\} \). The assumption of small \( \gamma_k \) implies small variations in the frequencies with time.

Each frequency \( \omega_k(t) \) influences all the elements of the parameter vector in a nontrivial way, but since the changes are small, a first-order expansion around the time instant \( t - 1 \) gives

\[
\theta_0(t) = \theta_0(t - 1) + F(t - 1) \gamma v(t)
\]

(3.2)

where we define the vector \( v(t) = (v_1(t), \ldots, v_m(t))^T \) and the scaling matrix \( \gamma = \text{diag}(\gamma_1, \ldots, \gamma_m) \). The Jacobian matrix \( F(t) \) is the gradient of \( \theta \) with respect to the frequencies \( \{\omega_k\} \) evaluated at the time-instant \( t \), i.e.,

\[
F(t) \triangleq \frac{\partial \theta}{\partial \omega(t)} = \left[ \begin{array}{c}
\frac{\partial a_1}{\partial \omega_1} & \cdots & \frac{\partial a_1}{\partial \omega_m} \\
\cdots & \cdots & \cdots \\
\frac{\partial a_m}{\partial \omega_1} & \cdots & \frac{\partial a_m}{\partial \omega_m}
\end{array} \right]_{\omega = \omega(t)}.
\]

(3.3)

In order to calculate \( F(t) \), we use the results presented in [9], where it is shown that the gradient for the \( k \)th polynomial coefficient \( a_k \) of (2.5) with respect to the \( l \)th root (2.6) is given by the recursion formula (for \( k = 1, \ldots, m \) and \( l = 1, \ldots, n \))

\[
\frac{\partial a_k}{\partial \eta_l} = - \sum_{p=0}^{k-1} \eta_l^{p-1} a_p \quad a_0 = 1.
\]

(3.4)

It is straightforward to see that the gradient of the \( l \)th root with respect to the \( j \)th frequency is given by the relation, cf. (2.6)

\[
\frac{\partial \eta_l}{\partial \omega_j} = i \eta_l \delta_{lj} - i \eta_l \delta_{l,m+j}
\]

(3.5)

where \( i = \sqrt{-1} \) and where \( \delta_{lj} \) is the Kronecker delta. Using the chain rule, a closed-form expression is achieved for the
elements in (3.3)
\[ \frac{\partial a_k}{\partial \omega_j} = \sum_{l=1}^{n} \frac{\partial a_k}{\partial \eta_l} \frac{\partial \eta_l}{\partial \omega_j} = - \sum_{l=1}^{n} \sum_{p=0}^{k-1} \eta_l \left( \left( i\eta_l \delta_{l,j} - i\eta_l \delta_{l,m+j} \right) e^{-i\omega_j(p-k-p)} - e^{i\omega_j(p-k-p)} \right) = 2 \sum_{p=0}^{k-1} a_p \sin(k-p) \omega_j \]
\[ a_0 = 1. \quad (3.6) \]

Note that (3.4)–(3.5) hold for a general polynomial with conjugated pairs of zeros, whereas (3.6) holds only for polynomials whose roots are on the unit circle.

IV. THE RECURSIVE ALGORITHM

The ANF is based on the model (2.8) and the recursive prediction error (RPE) algorithm [10], which minimizes the criterion
\[ V_t = \sum_{s=1}^{t} \lambda^{t-s} e^2(s) \quad (4.1) \]
where \( \lambda \) is the forgetting factor and where the prediction error \( e(t) \) is defined below. Thus, the parameter update \( \hat{\theta}(t) \) obeys the recursions
\[ \hat{\theta}(t) = \hat{\theta}(t-1) + L(t)e(t) \quad (4.2) \]
\[ e(t) = y(t) - \hat{y}(t|\hat{\theta}(t-1)) \quad (4.3) \]
\[ L(t) = P(t)\psi(t) \quad (4.4) \]
\[ P(t) = \frac{1}{\lambda} P(t-1) - \frac{P(t-1)\psi(t)\psi^T(t)P(t-1)}{\lambda + \psi^T(t)P(t-1)\psi(t)} \quad (4.5) \]

where \( \psi(t) \) approximates the gradient of \( \hat{y}(t|\theta) \) with respect to the set of parameters \( \theta \), i.e.,
\[ \psi(t) \approx \frac{\partial \hat{y}(t|\theta)}{\partial \theta} \bigg|_{\theta=\hat{\theta}(t-1)}. \quad (4.6) \]

An expression for \( \psi(t) \) as a function of \( \theta \) is derived in [1] and is given in Appendix A.

The error output of the algorithm (4.2)–(4.5) depends, among other things, on the forgetting factor \( \lambda \) and on the pole contraction factor \( \rho \). A common choice of \( \lambda \) in the tracking scenario is a constant somewhat smaller than unity, which implies that old prediction errors only contribute marginally to the cost function (4.1). The role of \( \rho \) in the algorithm is well described in [1]. Recall that \( \rho \) determines the bandwidth of the notch filter, see (2.9). Thus, its value has to be small enough to give a sufficiently large notch bandwidth and allow good tracking performance to the input sinusoids. On the other hand, it is desirable to have \( \rho \) as close to one as possible for the best accuracy. Thus, both \( \lambda \) and \( \rho \) have to be chosen as a compromise between tracking sensitivity and accuracy.

The properties of the algorithm for time-varying processes, when \( \lambda \) and \( \rho \) are constant and close to unity and the gain \( L(t) \) becomes small, can be investigated as follows. Assuming small errors, the description of the sine-wave-in-noise process (2.10) can be approximated by the first-order Taylor expansion around \( \hat{\theta}(t-1) \)
\[ \hat{y}(t|\theta_0(t-1)) \approx \hat{y}(t|\hat{\theta}(t-1)) + [\theta_0(t-1) - \hat{\theta}(t-1)]^T \psi(t). \quad (4.7) \]

Introducing the known (using past data) variable
\[ z(t) = y(t) - \hat{y}(t|\hat{\theta}(t-1)) + \hat{\theta}^T(t-1)\psi(t) \quad (4.8) \]
the prediction error (4.3) can be written as the linear regression
\[ \epsilon(t) = z(t) - \hat{\theta}^T(t-1)\psi(t). \quad (4.9) \]

The linear regression expression (4.9) implies that the asymptotic theory of tracking considered in [2] can be applied. In particular, the use of approximations similar to those presented above are studied in the Appendix of [2].

V. MSE ANALYSIS

This section analyzes the MSE of the ANF [1] with time-varying sine-waves in noise and the assumptions above. First, a difference equation for the parameter error is derived. The corresponding MSE matrix can then be obtained using the recent results of Ljung and Gunnarsson [2].

Using (3.2), (4.2), (4.3), and (4.7), we find that the estimation error \( \hat{\theta}(t) \) defined below obeys a linear time-varying difference equation. Straightforward calculations give
\[ \hat{\theta}(t) \triangleq \hat{\theta}(t) - \theta_0(t) \]
\[ = \hat{\theta}(t-1) + L(t)e(t) - \theta_0(t-1) - F(t-1)\gamma(t) \]
\[ = \hat{\theta}(t-1) + L(t)[y(t) - \hat{y}(t|\hat{\theta}(t-1))] - F(t-1)\gamma(t) \]
\[ = \hat{\theta}(t-1) + L(t)[y(t) - \hat{y}(t|\theta_0(t-1))] - \psi^T(t)\hat{\theta}(t-1) - F(t-1)\gamma(t) \]
\[ = \hat{\theta}(t-1) + L(t)[y(t) - \hat{y}(t|\theta_0(t-1))] - \psi^T(t)\hat{\theta}(t-1) - F(t-1)\gamma(t). \quad (5.1) \]

It follows from the assumption of small \{\gamma_k\} that the Jacobian matrix \( F(t) \) is approximately constant. For simplicity, it is hereafter approximated with the constant matrix \( F = F(0) \) for some \( t_0 \).

In addition, since the system is slowly varying, we use the fact that the changes in the parameter vector are small, and thus, the one-step prediction of the signal is approximately independent of the parameter variations, viz.
\[ \hat{y}(t|\theta_0(t-1)) \approx \hat{y}(t|\theta_0(t_0)). \quad (5.2) \]

for some constant parameter vector \( \theta_0(t_0) \). Thus, we can apply the small-error result of [8], yielding
\[ y(t) - \hat{y}(t|\theta_0(t_0)) \approx e(t) + O(\sqrt{1-\rho}). \quad (5.3) \]
Inserting (5.3) into (5.1) and neglecting the term of order \(O(\sqrt{1 - \rho})\) give

\[
\hat{\theta}(t) = [I - L(t)\psi^T(t)]\hat{\theta}(t - 1) + L(t)e(t) - F\gamma_1(t)
\]

(5.4)

which should be compared with (49) in [2]. In [2], it is shown that the behavior of the MSE matrix

\[
\Pi(t) = E[\hat{\theta}(t)\hat{\theta}^T(t)]
\]

(5.5)

when \(t \to \infty\) is well described by the matrix \(\hat{\Pi}\), which is given by

\[
\hat{\Pi} = \frac{(1 - \lambda)rG(t_0)^{-1}}{2} + \frac{F\gamma_1 F^T}{2(1 - \lambda)}
\]

(5.6)

where

\[
G(t_0) = E[\psi(t)\psi^T(t)].
\]

(5.7)

The small error covariance matrix \(G(t_0)\) in (5.7) is calculated using a stationary stochastic process to generate \(\psi(t)\), and a constant value of the system parameters denoted \(\theta(t_0)\), equal to the current estimate \(\hat{\theta}(t_0)\). See [2] for a more detailed discussion on this subject.

Due to (5.6), it is possible to solve a static optimization problem in order to tune the user-chosen variables of the algorithm \(\lambda\) and \(\rho\), where the dependence on \(\rho\) is embedded in \(G(t_0)\). In the next parts of this section, explicit expressions for \(G(t_0)\) will be derived in order to study the MSE.

A. The General Case

The small error covariance matrix \(G(t_0)\) corresponding to \(\psi(t)\) has two components: one due to the narrow-band signals and the other due to the additive noise. Since the gradient vector \(\psi(t)\) is given by the filtering of the data through \(1/A(\rho_0 t_0^-1, \hat{\theta})\), it follows that near the optimum, the narrow-band signals get amplified and play a dominant role in \(G(t_0)\). The calculations of the covariance matrix \(G(t_0)\) are presented in Appendix A, where it is shown that \(G(t_0)\) can be written as

\[
G(t_0) = \sum_{k=1}^{m} \alpha_k^2 \left| \frac{1}{2} A(\rho e^{j\omega_k}, \theta(t_0)) \right|^2 C(\omega_k)C^T(\omega_k) + O((1 - \rho)^{-3/2})
\]

(5.8)

where the first term is of order \(O((1 - \rho)^{-2})\). The vector \(C(\omega_k)\) is of length \(m\) and is given by

\[
C(\omega_k) = (2\cos(m - 1)\omega_k, \ldots, 2\cos\omega_k, 1)^T
\]

(5.9)

where \(\{\omega_k\}\) are the constant frequencies corresponding to \(\theta(t_0)\).

In Appendix B, it is shown that when the instantaneous frequencies are in the interior of \((0, \pi)\) and well separated, i.e.,

\[
\Delta\omega_{k,j}(t) \triangleq |\cos\omega_k(t) - \cos\omega_j(t)| \gg 1 - \rho
\]

(5.10)

the denominator in (5.8) can be approximated with the relation

\[
|A(\rho e^{j\omega_k}, \theta(t_0))|^2 \approx 4^m (1 - \rho)^2 (1 - \cos^2\omega_k) \cdot \prod_{j=1,j \neq k}^{m} \Delta\omega_{k,j}(t_0).
\]

(5.11)

Due to (5.11), the inverse of \(G(t_0)\) can be formally written as

\[
G(t_0)^{-1} = (1 - \rho)^2 \hat{G}(t_0)^{-1}
\]

(5.12)

where, from (5.8) and (5.11), it follows that

\[
\hat{G}(t_0) = \sum_{k=1}^{m} \alpha_k^2 \left| \frac{1}{2} A(\rho e^{j\omega_k}, \theta(t_0)) \right|^2 C(\omega_k)C^T(\omega_k).
\]

(5.13)

From (5.13), we note that \(G(t_0)\) is independent of \(\rho\).

Combining (5.8) with the MSE formula (5.6) gives a closed-form expression for the MSE. However, in order to investigate the behavior of the parameter MSE with respect to \(\lambda\) and \(\rho\), we use (5.12) and achieve the explicit expression

\[
\hat{\Pi} = (1 - \lambda)(1 - \rho)^2 \frac{\hat{G}(t_0)^{-1}}{2} + \frac{F\gamma_1 F^T}{2(1 - \lambda)}.
\]

(5.14)

In (5.14), the MSE is matrix valued so that a scalar quality measure

\[
\sigma = \text{trace}[Q\hat{\Pi}]
\]

(5.15)

is introduced. The quality measure (5.15) can be minimized for different user-chosen weighting matrices \(Q\) with respect to the tuning variables \(\lambda\) and \(\rho\). It is straightforward to show that the minimum value of (5.15) with respect to \(\lambda\) is achieved for

\[
\lambda_{opt} = 1 - \frac{(\text{trace}[QF\gamma_1 F^T])^{1/2}}{(1 - \rho)(\text{trace}[QG(t_0)^{-1}])^{1/2}}.
\]

(5.16)

The optimal value of \(\hat{\Pi}\) with respect to \(\rho\) is at the maximum value chosen so that the notch bandwidth (2.9) of the filter is sufficiently large. This will be studied in detail in the next section.

The weighting matrix

\[
Q = (FF^T)^{-1}
\]

(5.17)

corresponds to a quality measure (5.15) that gives equal weight to the different frequencies. To see this, let \(\hat{\Pi}_{\omega}\) describe the asymptotic frequency MSE

\[
\hat{\Pi}_{\omega} \approx \lim_{t \to \infty} E[\hat{\omega}(t)\hat{\omega}^T(t)].
\]

(5.18)

In (5.18), the frequency error vector \(\hat{\omega}(t)\) is given by

\[
\hat{\omega}(t) = \begin{bmatrix} \hat{\omega}_1(t) - \omega_1(t) \\ \vdots \\ \hat{\omega}_m(t) - \omega_m(t) \end{bmatrix}
\]

(5.19)

where the frequency estimate at time \(t\) is denoted by \(\hat{\omega}_k(t), k = 1, \ldots, m\). A first-order expansion for the relation
between the error in the parameters and the error in the frequencies gives

$$\dot{\theta}(t) - \theta_0(t) = F(t)\dot{\omega}(t) \approx F\omega(t)$$  \hspace{1cm} (5.20)

where the Jacobian matrix is defined in (3.3). Now, substituting (5.17) into (5.15) and using the expansion (5.10) give (F is square)

$$\sigma = \text{trace} \left[ (FF^T)^{-1} F\Pi_\omega F^T \right] = \text{trace} [\Pi_\omega].$$  \hspace{1cm} (5.21)

In addition, inserting (5.17) and (5.14) into (5.15) gives

$$\sigma = (1 - \lambda)(1 - \rho)^2 \frac{r\text{trace} \left[ (FF^T)^{-1} \tilde{G}(t_0)^{-1} \right]}{2} + \frac{\sum_{k=1}^{m} \gamma_k^2}{2(1 - \lambda)}.$$  \hspace{1cm} (5.22)

Finally, inserting (5.17) into (5.16) gives

$$\lambda_{\text{opt}} = 1 - \frac{(1 - \rho)(r\text{trace} \left[ (FF^T)^{-1} \tilde{G}(t_0)^{-1} \right])^{1/2}}{(1 - \rho)^{1/2}}.$$  \hspace{1cm} (5.23)

Note that in (5.23), \(\tilde{G}(t_0)\) is entirely determined by the signal \(x(t)\), and thus, the denominator is proportional to \((1 - \rho)^{1/2} \text{SNR}^{-1/2}\), where SNR denotes the signal-to-noise ratio. The numerator of the second term in (5.23) describes the (average) system variability.

Note that the random walk modeled frequency changes do not exclude that the frequencies cross each other. One way to handle this case is to extrapolate where the frequencies are going to be later and reinitialize the algorithm at that point. In the intermediate time, a lower order ANF can be used. However, we did not encounter any problems of this kind in our numerical simulations.

In the single sinewave case, the MSE expression is given in the following example.

Example—The Single Sinewave Case: In this case, the parameter vector \(\theta\) is reduced to the scalar \(a = -2\cos \omega\), the Jacobian element (3.3) is given by \(F = 2\sin \omega\), and the squared Jacobian element is given by \(F^2 = 4 - a^2\). Further, the small error covariance (5.13) is reduced to

$$\tilde{G}(t_0) = \alpha^2 \frac{1}{2} \frac{1}{4 - a_0(t_0)^2}. $$  \hspace{1cm} (5.24)

Here, \(\Pi_\omega\) is a scalar quantity. Thus, inserting (5.24) into (5.22) gives

$$\Pi_\omega = \frac{(1 - \lambda)(1 - \rho)^2}{2\text{SNR}} + \frac{\gamma^2}{2(1 - \lambda)} $$  \hspace{1cm} (5.25)

where the SNR is defined as \(\text{SNR} = \alpha^2/2\pi\). The minimal value of (5.25) is obtained with \(\lambda\) given by (see (5.23))

$$\lambda_{\text{opt}} = 1 - \frac{\gamma\sqrt{\text{SNR}}}{1 - \rho}.$$  \hspace{1cm} (5.26)

Inserting (5.26) into (5.25) gives

$$\Pi_{\omega|\lambda = \lambda_{\text{opt}}} = \frac{\gamma}{\sqrt{\text{SNR}}} \frac{1}{1 - \rho}. $$  \hspace{1cm} (5.27)

In light of the example, the following remarks are now in order:

The MSE formula (5.25) has the basic feature of a tradeoff between noise sensitivity and tracking error. In particular, the error variance depends on the SNR, and the tracking error depends on the frequency variation \(\gamma\). Note also that the MSE is independent of the actual frequency \(\omega\).

For the stationary sinewave case, it was shown in [3] that the bias term in the MSE could dominate the variance for extremely large values of data. In the tracking scenario, this is not the case since the cost function of the algorithm contains a forgetting factor \(\lambda < 1\). The use of \(\lambda < 1\) prevents the error variance from decreasing to zero as time tends to infinity, as seen from (5.25), where the error variance is of order \(O((1 - \lambda)(1 - \rho)^2)\). The asymptotic squared bias derived in [3] is of order \(O((1 - \rho)^{10})\), which is negligible compared with the error variance for all \(\lambda < 1\) in practical use.

By assumption and in practice, the forgetting factor \(\lambda\) is close to unity. Therefore, in order to use (5.26), we must require that the second term in (5.26) is sufficiently small, that is

$$\sqrt{\text{SNR} \gamma} \ll 1 - \rho.$$  \hspace{1cm} (5.28)

From (5.27), it follows that the MSE using \(\lambda = \lambda_{\text{opt}} < 1\) is of order \(O(1 - \rho)\). Thus, the optimal value of \(\rho\) (for minimizing the MSE) is as close to unity as possible, i.e., at the maximum value chosen so that the frequency estimate is inside the notch. This maximum value of \(\rho\) is studied in detail in the next section.

B. The Pole Contraction Factor

In the tracking scenario, the estimated parameter vector differs from the "true" vector for the following two reasons. The use of a forgetting factor \(\lambda < 1\) implies that the error variance is always present. The second reason is due to the nonstationarity of the input signal, which gives rise to a tracking error. However, we can expect that the parameter error of the ANF will become small after a transient period, and therefore, the present small error results can be applied. We will see later by numerical examples that this is indeed the case.

In order to analyze the validity of the derived MSE expressions, the small error assumption has to be investigated in detail. For this purpose, we study the single sinewave case and derive the covariance \(E[\psi(t)\psi^T(t)]\) with constant but unequal values of \(a_0(t_0)\) and \(\dot{a}(t_0)\). This case gives insight into the choice of \(\rho\) that cannot be seen directly from the result (5.6)-(5.7). In Appendix C, an analytical expression for \(E[\psi(t)\psi^T(t)]\) is presented, and in particular, it is shown that \(E[\psi(t)\psi^T(t)]\) can be written as

$$E[\psi(t)\psi^T(t)] = T_e + T_x$$  \hspace{1cm} (5.29)

where \(T_e\) is entirely determined by the measurement noise \(e(t)\), whereas \(T_x\) is determined by the signal component \(x(t)\). It is also shown in Appendix C, that the terms in (5.29) are of order, respectively,

$$T_e \sim O((1 - \rho)^{-1}).$$  \hspace{1cm} (5.30)
\[
T_x \sim \begin{cases} \frac{O(1 - \rho)}{(1 - \rho)^2}, & |\hat{a}(t_0) - a_0(t_0)| \ll 1 - \rho \\ \frac{O(1 - \rho)}{(1 - \rho)^2}, & |\hat{a}(t_0) - a_0(t_0)| \gg 1 - \rho. \end{cases}
\]

(5.31)

Since the small error performance is studied, the relation
\[
|\hat{a}(t) - a_0(t)| \ll 1 - \rho
\]

(5.32)
in (5.31) must be fulfilled for any time instant, especially \(t_0\).
Taking expectation of the square of the term on the left-hand side of (5.32) and using (5.20) gives
\[
\hat{\Pi} \approx F^2 \hat{\Pi}_w \ll (1 - \rho)^2.
\]

(5.33)

Using (5.27) gives
\[
(1 - \rho) \gg (4 - a_0(t_0)^2) \frac{\gamma}{\sqrt{\text{SNR}}}. \tag{5.34}
\]

From (5.34), we can expect the small error assumption to be fulfilled for larger values of \(\rho\) when the SNR increases or the signal variation \(\gamma\) decreases. This latter statement is consistent with the analyses in [1] and [3].

A more heuristically-based motivation of the possible values of \(\lambda\) and \(\rho\) is as follows. Recall that \(\lambda\) determines the forgetting rate in the algorithm, and \(\rho\) determines the notch bandwidth of the filter. Essentially, in this type of algorithm, a constant \(\lambda < 1\) corresponds to an exponential decay time constant or memory length given by [10]
\[
T_\lambda = \frac{1}{1 - \lambda}. \tag{5.35}
\]

To ensure that the frequency variations are small enough compared with the notch bandwidth of the ANF, we choose \(\lambda\) and \(\rho\) such that the standard deviation of the frequency variation over the time interval \(T_\lambda\) is small compared with the notch bandwidth of the filter. For this purpose, recall that the random walk noise \(\nu(t)\) for the frequency changes is of unit variance, and thus, the variance of the frequency changes over the period \(T_\lambda\) is, cf. (3.1)
\[
E[\omega(t + T_\lambda) - \omega(t)]^2 = E \left[ \sum_{k=n+1}^{r+T_\lambda} \gamma \nu(k) \right]^2 = \gamma^2 T_\lambda. \tag{5.36}
\]

Inserting (5.35) into (5.36) and using (2.9) for the bandwidth of the notch filter gives the relation
\[
\frac{\gamma^2}{1 - \lambda} \ll \pi^2 (1 - \rho)^2. \tag{5.37}
\]

Further, using (5.26) gives
\[
(1 - \rho)^2 \gg \frac{1}{\pi^2} \frac{\gamma^2}{1 - \lambda |\lambda = \lambda_{\text{opt}}| = \frac{1}{\pi^2} \frac{\gamma}{\sqrt{\text{SNR}}} (1 - \rho) \tag{5.38}
\]
or
\[
(1 - \rho) \gg \frac{1}{\pi^2} \frac{\gamma}{\sqrt{\text{SNR}}}, \tag{5.39}
\]

which is essentially the same relation as (5.34).

In light of (5.34) and (5.39), it is possible to formally write the value of \(\rho\) as
\[
\rho = 1 - \frac{c \gamma}{\sqrt{\text{SNR}}} \tag{5.40}
\]

for some value of the scalar \(c\). Combining (5.28) and (5.40) gives a lower bound of the scalar \(c\), i.e., \(c \gg \text{SNR}\). On the other hand, an upper bound of \(c\) follows from the fact that \(\rho\) is close to unity. Thus, the second term on the right-hand side of (5.40) is small, i.e., \(c \ll \sqrt{\text{SNR}}/\gamma\). Combining these bounds gives
\[
\text{SNR} \ll c \ll \sqrt{\text{SNR}}/\gamma. \tag{5.41}
\]

The quantity \(c\) is used in the next section in order to compare the derived theoretical MSE expressions with the simulation results.

VI. NUMERICAL EXAMPLES

In this section, the tracking ability of the algorithm is compared with the derived theoretical small error performance by illustrations of typical simulation runs.

Since the asymptotic tracking properties of the ANF are studied, no effort is made to achieve fast convergence at the start of data processing in the simulations. For simplicity and to achieve a short transient time, the initializations \(\hat{a}(0) = -2 \cos \omega(0)\) and \(P(0) = 1.0 \cdot 10^{-2}\) are used in the single sinewave case. The algorithm is applied with constant values of the user-chosen variables \(\lambda\) and \(\rho\). However, in practical use, it is advisable to apply the algorithm with smaller values of \(\rho\) (i.e., wider notches) and \(\lambda\) at the start of the data processing in order to speed up the convergence rate and eliminate the use of stability monitoring [1].

A. Tracking Illustrations

The first example illustrates the properties of the algorithm when estimating noisy and/or time-varying signals. For this purpose, a realization of length \(N = 10000\) is generated for a single sinewave \((m = 1)\) with amplitude \(\alpha = 1\) and with initial phase randomly chosen in the interval [0, 2\(\pi\)]. The initial frequency is \(\omega(0) = 0.3\). In this example, three cases are treated:

\begin{align*}
\text{case 1} & \quad \gamma = 0 & r = 0.5 \\
\text{case 2} & \quad \gamma = 1.0 \cdot 10^{-3} & r = 0 \\
\text{case 3} & \quad \gamma = 1.0 \cdot 10^{-3} & r = 0.5
\end{align*}

where \(\gamma\) is the scaling factor for the frequency drift, and \(r\) is the variance of the measurement noise. In particular, in the first case, we study the behavior of the algorithm for a time-invariant sinewave in additive noise. The second case studies the tracking properties of the algorithm when the measurements are noise free and consist of a single time-varying sinewave. Finally, in the third case, the general situation with both tracking error and error due to the measurement noise is studied. Since the MSE is a tradeoff between noise sensitivity (case 1) and tracking error (case 2), we expect that the sum of MSE’s from case 1 and case 2 is approximately equal to the MSE in case 3. In the runs, the values of \(\lambda\) and \(\rho\) are constant and given by \(\lambda = 0.98\) and \(\rho = 0.95\), which for case 3 is the optimal value for \(\lambda\) (see (5.26)) and, for \(\rho\), corresponds to a value of \(c = 50\) (see (5.40)). Note that relation (5.41) is fulfilled.
The tracking properties of the ANF algorithm is now illustrated by a simulation run for case 1–case 3. The aim is to illustrate the tradeoff between error variance and tracking error and to verify the theoretical expression for the MSE. The results are presented in Figs. 1–3, which depict the estimate \( \hat{\omega}(t) \) and the true frequency \( \omega(t) \) versus time. Here, \( \hat{\omega}(t) \) is calculated from the parameter estimate \( \hat{\alpha}(t) \) from the relation

\[ \hat{\alpha} = -2\cos \omega. \]

Since the runs are based on a large number of data, only every 20th sample is shown in the diagrams.

In the diagram of Fig. 1 we note the constant true frequency at \( \omega = 0.3 \) and an estimate that fluctuates around it. Calculating the normalized sum of squared prediction errors gives \( 1/N\Sigma(t)^2 \approx 0.55 \), which is approximately the same as the variance of the measurement noise \((\tau = 0.5)\). This implies that the term of order \( O(\sqrt{1-\rho}) \) in (5.3) is small and can be neglected. Fig. 2 shows the tracking properties of the algorithm for noise-free measurements, and we note that the algorithm is able to track the variations in the input frequency. Since no measurement noise is present, \( 1/N\Sigma(t)^2 \approx 7.1 \cdot 10^{-3} \) is the prediction error due to the time-varying input frequency. Comparing the values of \( 1/N\Sigma(t)^2 \) for the first two cases imply that the dependence of the one-step prediction of the signal on parameter variations is small compared with the parameter error, and thus, (5.2) is a valid approximation. In Fig. 3, the tracking properties of the algorithm for noisy measurements are illustrated.

In order to verify the theoretical expression for the frequency MSE, the sum square error (SSE) is evaluated as

\[
SSE = \frac{1}{9000} \sum_{t=1001}^{10000} (\hat{\omega}(t) - \omega(t))^2.
\]  

The results from evaluating (6.1) and from evaluating \( \hat{\Pi}_w \) in (5.25) are presented in Table I for the three different cases, respectively. Comparing the values in Table I shows that the theoretical and empirical MSE agree reasonably well.

<table>
<thead>
<tr>
<th>CASE</th>
<th>( \hat{\Pi}_w )</th>
<th>SSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.5 \cdot 10^{-5}</td>
<td>1.1 \cdot 10^{-5}</td>
</tr>
<tr>
<td>2</td>
<td>2.5 \cdot 10^{-5}</td>
<td>3.5 \cdot 10^{-5}</td>
</tr>
<tr>
<td>3</td>
<td>5.0 \cdot 10^{-5}</td>
<td>4.8 \cdot 10^{-5}</td>
</tr>
</tbody>
</table>

**B. The Influence of \( \lambda \) and \( \rho \) on MSE**

In this example, we study the effect on \( \hat{\Pi}_w \) with respect to the user-chosen variables \( \lambda \) and \( \rho \), which determine the asymptotic tracking ability of the ANF. From (5.25), we can expect a minima in the MSE with respect to the forgetting factor \( \lambda \) for a value slightly smaller than unity and given by (5.26). It is also expected that the MSE decreases when the pole contraction factor \( \rho \) increases towards unity. However, the analysis is based on the assumption that the notch bandwidth of the filter is large enough to ensure small errors, and therefore, in practical use, we cannot have \( \rho \) too close to unity. This fact was studied in detail in Section V-B. Especially, in Section V-B, the pole contraction factor was written in the form (5.40), where \( c \) was bounded by relation (5.41).
A realization of length \( N = 20000 \) is generated for a single time-varying sine wave \((m = 1)\) with amplitude \( \alpha = 1 \) and with random initial phase. The initial frequency is \( \omega(0) = \pi / 3 \), and its variation is \( \gamma = 1.0 \cdot 10^{-3} \). The variance of the noise is \( \tau = 0.5 \). For the analysis, two cases are treated:

\[
\begin{align*}
\text{case 4} & \quad \lambda \in \{0.970, \ldots, 1\} \quad \rho = 0.95 \\
\text{case 5} & \quad \lambda = 0.98 \quad \rho \in \{0.93, \ldots, 0.9999\}.
\end{align*}
\]

For case 4, we can note that the optimal value (i.e., minimizing the MSE) of \( \lambda \) is given by \( \lambda_{\text{opt}} = 0.980 \); see (5.26). In addition, this particular value of \( \rho \) corresponds to \( c = 50 \), which clearly fulfills (5.41). For case 5, Fig. 4 depicts the corresponding values of \( c \) with respect to \( \rho \), and we note that the value of \( c \) decreases with increasing \( \rho \). The dotted line in the diagram is the lower bound for tracking in (5.41). The upper bound for tracking in (5.41) equals \( 10^{3} \) and is beyond the range of the scale of Fig. 4. We can note that the relation between \( c \) and the lower bound, i.e., \( c / \text{SNR} \), equals 10 for \( \rho = 0.99 \). We will later see that this in practice gives an approximative upper bound on \( \rho \).

From the simulation runs, the SSE is computed similar to (6.1) using data \( t = 10001, \ldots, 20000 \). In Figs. 5 and 6, the MSE’s \( \langle \hat{\mathbf{I}}_w \rangle \) and SSE) are presented versus \( \lambda \) (case 4) and \( \rho \) (case 5), respectively.

In Fig. 5, there is an excellent agreement between the theoretical and estimated MSE. Notice that the minimum around the theoretical optimal value of \( \lambda \) is very shallow and coincides with the minimum for the SSE.

In Fig. 6, there is reasonable agreement for \( \rho < 0.99 \) (i.e., \( c / \text{SNR} > 10 \)). For larger values of the pole contraction factor \( \rho \), the SSE increases significantly as the algorithm loses tracking ability. Recall that \( \rho \) has to be small enough to give a sufficiently large notch bandwidth compared with the variations in the signal. In addition, when the value of \( \rho \) becomes too small, the discrepancy between the theoretical analysis and the simulation result increases. This latter discrepancy originates from the assumption used that \( \rho \) is sufficiently close to unity.

### C. Frequency Separation

This example studies the quality measure \( \sigma \) (5.15) using the special weighting matrix (5.17) for the case of two slowly time-varying sinewaves in noise. In this case, the measure \( \sigma \) equals \( \sigma = \text{trace} [ \hat{\mathbf{I}}_w ] \). In this example, the forgetting factor equals \( \lambda = 0.98 \). Two values of the contraction factor are considered: \( \rho = 0.9 \) and \( \rho = 0.95 \), respectively. The noise variance is \( \tau = 0.5 \). Fig. 7 depicts \( \sigma \) versus the instantaneous second frequency \( \omega_2(t) \) for the case (6.9)

\[
\begin{align*}
\alpha_1 = \alpha_2 = 1 & \quad \omega_1(t) = 0.3 \\
\gamma = \begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix} & = 1.0 \cdot 10^{-3} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.
\end{align*}
\]

From Fig. 7, we note the following. The quality measure \( \sigma \) increases rapidly when \( \omega_2(t) \) is close to the boundaries \( \omega_2(t) = 0 \) and \( \omega_2(t) = \pi \), and when the two instantaneous
frequencies are closely spaced. Thus, we can expect problems with resolving the two frequencies from each other when the instantaneous frequency separation is too small. To overcome this, it is necessary to make the notches of the filter narrower, that is, increase the value of $\rho$. However, there is a bound when the notches become too narrow and the ANF fails to track the frequency variations, which is described by (5.10).

Notice also that for well-separated instantaneous frequencies in the intervals of $(0, \pi)$, the sum of frequency MSE's obeys $\sigma \approx 2\pi\omega_c$, where $\pi\omega_c$ is calculated using the single sinusoidal result (5.25).

VII. CONCLUSIONS

We have studied an ANF algorithm [1] for tracking time-varying sinusoidal signals in additive measurement noise. Closed-form expressions for the asymptotic frequency MSE have been derived and studied under the assumption that the frequencies are slowly time varying. The result gives insight into the behavior of the algorithm and has been used in order to adjust the tuning variables ($\lambda$ and $\rho$) in an optimal way to minimize the MSE.

In the general case, the MSE is given by (5.14), and the optimal value of $\lambda$ is given by (5.16). In particular, in the case of a single time-varying sinusoidal in noise, the theoretical frequency MSE is given by (5.25), and the optimal value of the algorithm's forgetting factor $A$ is explicitly given by (5.26).

In the tracking scenario, the notch bandwidth of the filter must be sufficiently large in order to enable good tracking. The notch bandwidth is entirely determined by the user-chosen pole contraction factor $\rho$ arising from the special model structure used. The connection between the behavior of the system and the value of $\rho$ is described by relations (5.40) and (5.41) in the single sinusoidal case and in part by relation (5.10) in the case of multiple frequencies.

In addition, in this article we presented numerical examples that illustrate and compare the theory with simulation runs. The comparison shows that there is a good agreement between the theoretical and simulation results.
Known results (see, for example, [12]) give that the covariance matrix can be written as

\[
R = \sum_{k=1}^{m} \frac{\alpha_k^2}{2} \frac{1}{|A(\rho e^{i\omega_k}, \theta_0(t_0))|^2} \begin{pmatrix}
1 & \cdots & \cos(n-2)\omega_k \\
\vdots & \ddots & \vdots \\
\cos(n-2)\omega_k & \cdots & 1
\end{pmatrix}^T \\
\cdot \text{Re} \{ \mathcal{C}(\omega_k) \mathcal{C}^H(\omega_k) \}. \tag{A.12}
\]

In (A.12), the vector \( \mathcal{C}(\omega_k) \) of dimension \((n-1)(1)\), is given by

\[
\mathcal{C}(\omega_k) = (1, e^{-i\omega_k}, \cdots, e^{-i\omega_k(n-2)})^T \tag{A.13}
\]

where the frequencies corresponding to \( \theta_0(t_0) \) are denoted \( \omega_k \).

In order to rewrite (A.12), note the relation

\[
\forall \mathcal{C}(\omega_k) \triangleq \begin{pmatrix}
1 & 0 & \cdots & 1 \\
0 & \ddots & \vdots & \vdots \\
0 & 1 & \cdots & 0 \\
e^{-i\omega_k} & \cdots & e^{-i\omega_k(n-2)} & 2\cos(m-1)\omega_k \\
& \ddots & \vdots & \vdots \\
& & 2\cos\omega_k & 1
\end{pmatrix}
\]

Thus, it follows that

\[
\forall \text{Re} \{ \mathcal{C}(\omega_k) \mathcal{C}^H(\omega_k) \}^T = \mathcal{C}(\omega_k) \mathcal{C}^T(\omega_k) \tag{A.14}
\]

where \( \mathcal{C}(\omega_k) \) is the vector of dimension \((m)(1)\), given by (5.9).

Using (A.12) and (A.14), it follows that the covariance matrix \( G(t_0) \) satisfies

\[
G(t_0) = \forall \mathcal{R} \mathcal{T} + O((1-\rho)^{-3/2})
\]

\[
= \sum_{k=1}^{m} \frac{\alpha_k^2}{2} \frac{1}{|A(\rho e^{i\omega_k}, \theta_0(t_0))|^2} \cdot \mathcal{C}(\omega_k) \mathcal{C}^T(\omega_k) + O((1-\rho)^{-3/2}) \tag{A.15}
\]

which is the same as relation (5.8).

**APPENDIX B**  
**PROOF OF (5.11)**

In the case that the frequencies are well separated (see (5.10), the denominator in (A.15) or (5.8) can be approximated as follows, cf. (A.25):

\[
|A(\rho e^{i\omega_k}, \theta_0(t_0))|^2 = \prod_{j=1}^{m} [1 - 2\rho \cos \omega_j e^{-i\omega_k} + \rho^2 e^{-2i\omega_k}]^2 \\
\approx 4(1-\rho)^2(1-\rho \cos^2 \omega_k) \\
\cdot \prod_{j=1, j \neq k}^{m} (-2 \cos \omega_j + 2 \cos \omega_k)^2 \\
= 4^m (1-\rho)^2 (1-\rho \cos^2 \omega_k) \\
\cdot \prod_{j=1, j \neq k}^{m} (\cos \omega_j - \cos \omega_k)^2 \\
= 4^m (1-\rho)^2 (1-\rho \cos^2 \omega_k) \\
\cdot \prod_{j=1, j \neq k}^{m} \Delta^2 \omega_{k,j} \tag{A.16}
\]

where \( \{\omega_k\}_{k=1}^{m} \) are the frequencies corresponding to \( \theta_0(t_0) \). The behavior of \(|A(\cdot)|^2\) when \( \omega_k \) is in the interior of \((0, \pi)\) is then well described by

\[
|A(\rho e^{i\omega_k}, \theta_0(t_0))|^2 \approx 4^m (1-\rho)^2 \\
\cdot (1-\cos^2 \omega_k) \\
\cdot \prod_{j=1, j \neq k}^{m} \Delta^2 \omega_{k,j} \tag{A.17}
\]

which verifies (5.11).

**APPENDIX C**  
**THE SINGLE SINEWAVE CASE**

In the single sinewave case, it is possible to derive an analytical expression for the covariance \( E[\psi(t)\psi^T(t)] \) with constant but unequal values of \( \tilde{\alpha} = \tilde{\alpha}(t_0) \) and \( \alpha_0 = \alpha_0(t_0) \). Here, the stationary gradient output of the algorithm is given by [1]

\[
\psi(t) = \frac{1}{A(pq^{-1}, \tilde{\alpha})} [-y(t-1) + \rho \psi(t-1)]. \tag{A.18}
\]

\[
E[\psi(t)\psi^T(t)] = E\left[ \begin{pmatrix} 1 \\
A(pq^{-1}, \tilde{\alpha}) \\
A(pq^{-1}, \tilde{\alpha}) - 1 \end{pmatrix} y(t) \right]^2 \\
= E\left[ \begin{pmatrix} 1 \\
A(pq^{-1}, \tilde{\alpha}) \\
A(pq^{-1}, \tilde{\alpha}) - 1 \end{pmatrix} e(t) \right]^2 \\
+ E\left[ \begin{pmatrix} 1 \\
A(pq^{-1}, \tilde{\alpha}) \\
A(pq^{-1}, \tilde{\alpha}) - 1 \end{pmatrix} x(t) \right]^2 \\
\equiv T_c + T_f. \tag{A.19}
\]
First, note (A.19), which appears at the bottom of the previous page. Using residue calculus, we get

$$T_e = (1 - \rho)^2 E \left[ \left( \frac{1 - \rho q^{-2}}{A^2(\rho q^{-1}, \hat{a})} c(t) \right)^2 \right]$$

$$= (1 - \rho)^2 \frac{r}{2\pi i} \oint \mathcal{F}(z) \frac{dz}{z} = r(1 - \rho)^2 I(\hat{a}, \rho) \quad (A.20)$$

where in (A.20), \(r\) is the variance of the measurement noise, and \(\mathcal{F}(z)\) is given by

$$\mathcal{F}(z) = \frac{z^4 - \rho z^2}{z^4 + 2\hat{a} \rho z^3 + \rho^2 (2 + \hat{a}^2) z^2 + 2\hat{a} \rho^2 z + \rho^4}. \quad (A.21)$$

It can be shown that for \(\rho\) close to unity

$$I(\hat{a}, \rho) = \frac{1}{2(1 - \rho)^2 (4 - \hat{a}^2)} + O((1 - \rho)^{-2}). \quad (A.22)$$

After inserting (A.22) into (A.20), the part of (A.19) arising from the noise can be written, up to the order \(O(1)\), as

$$T_e = \frac{r}{2(1 - \rho)(4 - \hat{a}^2)} \quad (A.23)$$

which verifies (5.30). The signal-dependent part \(T_s\) is the squared transfer function, cf. (A.19), evaluated at the instantaneous frequency \(\omega = \omega(t_0)\). Thus

$$T_s = (1 - \rho)^2 \frac{|1 - e^{-2i\omega}|^2}{A^2(\rho e^{-i\omega}, \hat{a})^2} \frac{\alpha^2}{2} \quad (A.24)$$

Note that for \(\alpha = \frac{\omega}{2} = -2 \cos \omega\), it holds true that

$$\cos 2\omega = \frac{1}{2}(\alpha^2 - 2)$$

and that

$$|A(\rho e^{-i\omega}, \hat{a})|^2 = |A(\rho e^{-i\omega}, \hat{a})|^2 |A(\rho e^{i\omega}, \hat{a})|^2.$$  

Thus, for \(\rho\) close to unity

$$|A(\rho e^{-i\omega}, \hat{a})|^2 = (1 - \rho)^2 + \rho(\hat{a}^2 \rho - (1 + \rho^2)\hat{a} a_0 + \rho a_0^2)$$

$$\approx \begin{cases} 
\frac{(4 - \rho a_0^2)(1 - \rho^2)}{\hat{a} a_0} & |\hat{a} - a_0| \ll 1 - \rho \\
|\hat{a} - a_0|^2 & |\hat{a} - a_0| \gg 1 - \rho
\end{cases} \quad (A.25)$$

and

$$|1 - e^{-2i\omega}|^2 = 1 - \rho(\hat{a}_0^2 - 2) + \rho^2$$

$$= (1 + \rho)^2 - \rho a_0^2 \approx 4 - \rho a_0^2 \quad (A.26)$$

where the last equalities in (A.25) and (A.26) follow from a first-order expansion around \(\rho = 1\). Substituting (A.25) and (A.26) into (A.24), the following expression for the covariance \(T_x\) is achieved

$$T_x = \begin{cases} 
\frac{\alpha^2}{2} & \frac{1}{(4 - \rho a_0^2)(1 - \rho)} \ |\hat{a} - a_0| \ll 1 - \rho \\
\frac{1}{\alpha^2} & \frac{1}{(4 - \rho a_0^2)(1 - \rho)} |\hat{a} - a_0| \gg 1 - \rho
\end{cases} \quad (A.27)$$

which yields (5.31).

REFERENCES


Peter Händel (S’88–M’94) was born in Uppsala, Sweden, in 1962. He received the M.Sc. degree in engineering physics in 1987 and the Ph.D. degree in automatic control in 1993, both from the Institute of Technology, Uppsala University, Uppsala, Sweden. During 1987, he was at The Svedberg Laboratory, Uppsala, and from 1988 to 1993, he was with the Systems and Control Group, Uppsala University. Since 1993, he has been with Ericsson Radio Systems AB, Stockholm, Sweden. His research interests are in the fields of system identification, time series analysis, and signal processing.

Arye Nehorai (S’80–M’83–SM’90–F’94), for a photograph and biography please see page 398 of this issue.