Akademie věd České republiky
ÚSTAV TEORIE INFORMACE A AUTOMATIZACE

Academy of Sciences of the Czech Republic
INSTITUTE OF INFORMATION THEORY AND AUTOMATION

RESEARCH REPORT

P. Tichavský A. Nehorai

Comparative Study of Four Adaptive Frequency Trackers

No. 1845 August 1995

ÚTIA AV ČR P. O. Box 18, 182 08 Prague, Czech Republic
Telex: 122018 atom c, Fax: (+42)(2)6641 4903
e-mail: utia@utia.cas.cz

This report constitutes an unrefereed manuscript which is intended to be submitted for publication. Any opinions and conclusions expressed in this report are those of the author(s) and do not necessarily represent the views of the Institute.

Comparative Study of Four Adaptive Frequency Trackers

Petr Tichavský and **Arye Nehorai**

Abstract

We study and compare four algorithms for adaptive retrieval of slowly time varying multiple cisoids in noise: adaptive notch filter, [1, 2], multiple frequency tracker, [3, 4], adaptive estimation scheme in [5], and hyperstable adaptive line enhancer, [6, 7]. The local behavior of the algorithms in a neighborhood of their equilibrium state (assuming high signal-to-noise ratio (SNR) and large data sample) for a two–cisoid signal is treated in a similar way to the linear filter approximation technique used for a single cisoid case, [4]. Next, behavior of the four algorithms with respect to global convergence and the SNR threshold are examined by aid of computer simulations.

I. INTRODUCTION

The recovery of noisy sinusoidal signals with slowly varying parameters is important in many engineering applications. In this paper, we study the complex-valued version of the problem, that is the recursive estimation of the parameters of multiple *cisoids* in noise. This signal model slightly differs from the traditional one with real-valued sinusoids. However, most methods solving the problem for real-valued sinusoids can be easily modified to cisoids and vice versa. Also, performance of the same algorithm in the real-valued and complex-valued version is usually similar. The main reason why we prefer to deal with the complex-valued version of the problem is the lower complexity of the analysis; for example, a notch filter for tracking or canceling p real-valued sinusoids must have the order $2p$, while the same task for complex-valued signal with p cisoids is performed by a notch filter of the order p.

The most frequently studied algorithm for solving the above problem is the adaptive notch filter (ANF), see [1, 2]. The algorithm is based on the general recursive least square prediction error method, [8], and exists in a number of variants such as adaptive lattice notch filter or the ANF in cascade implementation, [9]. In this paper we treat the ANF modified for complex-valued signals in [4], and compare its tracking performance with those of three other methods: (i) "Multiple Frequency Tracker" (MFT), [4], (ii) "Adaptive IIR Structure for separation, enhancement and tracking 2" (AIS-2)[3], [5], and (iii) "Hyperstable Adaptive Line Enhancer" (HALE), [6]. The MFT algorithm is based on recursive estimation of instantaneous amplitudes and phases of the cisoids, combined with update of the instantaneous frequencies employing the estimated phase differences. The AIS-2 is an ad-hoc

[1]This work was supported by the Air Force Office of Scientific Research under Grant no. F49620-93-1-0096, the Office of Naval Research under Grant no. N00014-91-J-1298, and the U.S. Information Agency under the J.W. Fulbright Fellowship.

[2]Arye Nehorai is with Yale University, Dept. of Electrical Engineering, P.O. Box 208284, New Haven, CT 06520-8284, U.S.A.

[3]The number 2 is appended in order to distinguish the algorithm from ANF, which is an "adaptive IIR structure" as well.

method whose main advantage is its computational simplicity. The HALE employs a resonator-based IIR filter and its adaptation law is derived from general theory of hyperstability, [10, 11]. In this paper we study a complex-valued version of the algorithm proposed for tracking of parameters of real sinusoids in [6, 7].

All the four methods depend on two design variables that influence the tradeoff between the tracking ability and noise rejection. The performance of the methods (measured by mean-square estimation error) can be optimized by suitable choice of the design variables, which depends on properties of the processed signal, namely on the signal-to-noise ratio (SNR) and rate of evolution of the signal parameters. For example, the evolution of the signal parameters can be modeled by random walk, i.e. by a Gaussian random process with independent random increments, and the rate of evolution can have the meaning of the variance of the increments.

Basically, if the rate of evolution of the sinusoidal frequencies is low compared to 1/SNR, it is judicious to take pole contraction factor of ANF or forgetting factors of MFT close to 1, and in the opposite case take them small. The optimum choice of the design variables of these two algorithms has been found for the single cisoid case in [4]. In [12] it is shown that for this choice of variables the algorithms achieve the least mean-square estimation error, given by the Posterior Cramér-Rao Bound, what is a lower bound suited for random parameters, [13].

The optimization of the choice of the design variables can be performed adaptively (recursively) using the processed data, [21]. For simplicity, this paper considers only a constant choice of the design parameters.

The analysis of the algorithms performed in [4] for $p = 1$, referred as the *linear filter approximation* technique, is "local" in a certain sense. It is based on the assumptions that all internal (state-space) parameters of the algorithms fluctuate in a neighborhood of the corresponding equilibrium state (ES), [4]. The ES means the ideal values of the state-space parameters that correspond to processing of the stationary noiseless signal with true frequencies, amplitudes and initial phases of the cisoids. The small size of the fluctuations around the ES are assured by the assumptions of high SNR, slow evolution of the signal parameters, local stability of the algorithms and proper initialization of the methods, or large N assumption, provided the algorithms are globally convergent.

The ES of a procedure, if known, allows a smooth start of the tracking, without any initial transients. The true ES is rarely available in practice; however, if an initial estimate of the sinusoidal frequencies is available, it can be used together with an initial segment of the data to estimate the initial amplitudes and phases of the cisoids and consequently the ES, so that the transients can be minimized. (If the initial frequencies are not known, they can be estimated from the data by some standard frequency estimation method, also.)

Besides the local behavior of the algorithms, some other characteristics are important as well. The first is the *global convergence*, which means the ability of the procedures to approach their ES when processing a stationary noiseless signal, after being arbitrarily initialized. This property is very desirable if the signal parameters may have large abrupt changes in time, or if the algorithms are not started from the ES. The ANF and MFT are known to exhibit very good global convergence, but analytic proof is not known yet. The only procedure with proven global convergence (at least for $p \leq 2$) is HALE.

Another important attribute of the tracking procedures is the SNR threshold. This means the limit value of the SNR, below which the algorithms lose their tracking ability because of high variance of the estimates. In this paper, two ways of determining the SNR threshold are considered. In the first case the design variables of the algorithms are chosen constant, and in the latter case the design variables of the procedures are taken adaptively w.r.t. the SNR.

Finally, we discuss the computational complexity, considering recursive estimation of parameters (frequencies, in particular) of p cisoids. We show that the computational complexity of ANF, MFT, and AIS-2 is $O(p^2)$ operations per data sample[4] , and the complexity of HALE is $O(p)$ per data sample, but the last order does not include operations needed for finding the instantaneous frequencies at each time by rooting of a $p-$th order polynomial. We note that the above mentioned complexity of the procedures is only asymptotic, valid for large p, and usually it is not of practical interest to track more than several slowly varying cisoids.

A number of authors have studied the performance of different variants of ANF. The tracking properties of ANF have been analyzed in [14], using the ordinary differential equation (ODE) approach, and in [15] using the Ljung and Gunnarsons's approach, [16]. Another recent paper dealing with tracking analysis of the (lattice) ANF is [20].

The local properties of the MFT and ANF for a model with a single cisoid with slowly varying frequency were studied in [4], using so called *Approximating Linear Filter* (ALF). The ALF characterizes the relation between the sequence of the frequency estimation errors and sequencies of the (normalized) additive noise and of the one step changes of the true frequency, provided that the algorithms operate in a neighborhood of their equilibrium states.

In this paper, the ALF is calculated for the AIS-2 and HALE algorithms. It is shown that the approximating filter for these algorithms has the same form as the one for ANF and MFT; it follows that under the used approximation the performance of the algorithms is equal. Comparison of the coefficients of the ALF allows to relate the parameters of the different methods.

In the case of multiple cisoids we restrict ourselves to the question of local stability of the methods. We introduce a *local convergence factor* that characterizes the rate of the local convergence of the algorithms to their steady-state values, provided that the incoming signal is approximately stationary and the SNR is high.

The signal under consideration has the form

$$y_n = \sum_{k=1}^{p} A_{kn} + v_n \qquad n = 0, 1, 2, \ldots \tag{1}$$

where A_{kn} represents the k-th cisoid at the time instant n, v_n is the noise and p is the (known) number of the cisoids. Assume that A_{kn} evolves according to

$$A_{kn} = e^{i\omega_{kn}} A_{k,n-1}(1 + h_{kn}) \qquad k = 1, \ldots, p. \tag{2}$$

In (2), h_{kn} is implicitly defined as the relative increment of the magnitude of the k-th cisoid and ω_{kn} is the k-th instantaneous angular frequency, given as the angle increment of A_{kn},

$$\omega_{kn} \triangleq \mathrm{Arg}\left[\frac{A_{kn}}{A_{k,n-1}}\right]. \tag{3}$$

The stationary case is given by (1) and (2) with constant $\{\omega_{kn}\}$ and $\{h_{kn} \equiv 0\}$.

We assume that the noise $\{v_n\}$ is zero mean and white with

$$E[v_k\, v_\ell^*] = \sigma^2 \delta_{k,\ell} \qquad E[v_k\, v_\ell] = 0 \qquad k, \ell = 0, 1, 2, \ldots. \tag{4}$$

In (4), "E" denotes the expectation, the asterix " *" denotes the complex conjugate and $\delta_{k,\ell}$ the Kronecker delta.

[4]In [5] it is erroneously claimed that complexity of AIS-2 is $O(p)$.

Further, introduce the notation

$$
\begin{aligned}
A_n &= (A_{1n}, \ldots, A_{pn})^T && (5)\\
\Omega_n &= (\omega_{1n}, \ldots, \omega_{pn})^T && (6)
\end{aligned}
$$

where the superscript "^{T}" denotes the transpose. The vectors in (5) and (6) together contain the complete information about the signal at time n. The former vector will be called in the sequel the *signal decomposition* vector, considering the formula (1). The main problem solved by the algorithms is the estimation and tracking of the instantaneous frequencies, $\{\omega_{kn}\}$.

The rest of this paper is organized as follows. In Section II, the MFT, ANF, AIS-2 and HALE algorithms are formulated. Section III summarizes some analytical results concerning the MFT and ANF for a single cisoid, [4], and completes them by computing the linear filter approximation for the AIS-2 and HALE.

Differences between the algorithms are revealed in Section IV, where the ALF approximation is extended to the case of $p = 2$ cisoids. Section V includes some simulation results that demonstrate the global convergence properties and the SNR thresholds of the algorithms. Section VI summarizes the results and draws conclusions.

II. FORMULATION OF THE ALGORITHMS

Basically, each tracking algorithm is described by a set of internal parameters (state-space parameters, vector Θ_n) and a rule (update law) describing how these parameters are updated using the previous state and next available value of the signal at each time instant n,

$$\Theta_n = \mathcal{F}(\Theta_{n-1}, y_n) \qquad (7)$$

The mapping $\mathcal{F}$ representing the update law is usually nonlinear and depends on some user chosen design variables. The estimated parameters of the signal (namely frequencies) at time n are either directly included in Θ_n or given as a function of Θ_n.

In this section, the full description of the four algorithms is presented, including the following characteristics: (i) user chosen variables and their range, (ii) state-space parameters, i.e. elements of Θ_n, (iii) update law representing the mapping $\mathcal{F}$ in (7), and (iv) the ES values of the parameters, denoted by a bar. The ES of Θ_n in (7) obeys the equation

$$\bar{\Theta}_t = \mathcal{F}(\bar{\Theta}_{t-1}, \bar{y}_t)\,; \qquad t = n, n \pm 1, n \pm 2, \ldots \qquad (8)$$

where $\{\bar{y}_t\}$ is a stationary noiseless signal that corresponds to the true parameters A_n, Ω_n of the given signal at time instant n.

A. Adaptive Notch Filter

Design variables: $\lambda, \rho \in (0, 1]$.
State-space parameters: the estimated frequencies $\hat{\Omega}_n = (\hat{\omega}_{1n}, \ldots, \hat{\omega}_{pn})^T$, prediction error ε_n and its $p-1$ preceding values $\varepsilon_{n-1}, \ldots, \varepsilon_{n-p+1}$, an error gradient ψ_n and its $p-1$ preceding values $\psi_{n-1}, \ldots, \psi_{n-p+1}$, and a $(p \times p)$ matrix R_n (or the real part of R_n, see below the recursion).

Update law:

$$\varepsilon_n = y_n + \sum_{k=1}^{p} a_k(\hat{\Omega}_{n-1}) \left[y_{n-k} - \lambda^k \varepsilon_{n-k}\right] \tag{9}$$

$$\psi_n = -\sum_{k=1}^{p} \left\{b_k(\hat{\Omega}_{n-1}) \left[y_{n-k} - \lambda^k \varepsilon_{n-k}\right] + \lambda^k a_k(\hat{\Omega}_{n-1})\psi_{n-k}\right\} \tag{10}$$

$$\hat{\Omega}_n = \hat{\Omega}_{n-1} + (\mathrm{Re}\{R_n\})^{-1}\,\mathrm{Re}\{\psi_n^* \varepsilon_n\} \tag{11}$$

$$R_n = \rho R_{n-1} + \psi_n \psi_n^H \tag{12}$$

where $a_k(\hat{\Omega}_{n-1})$ denotes the coefficients of the polynomial determined by its roots $e^{i\hat{\omega}_{k,n-1}}$, $k = 1, \ldots, p$:

$$\prod_{k=1}^{p} (1 - e^{i\hat{\omega}_{k,n-1}} q^{-1}) = 1 + a_1(\hat{\Omega}_{n-1})q^{-1} + \ldots + a_p(\hat{\Omega}_{n-1})q^{-p}. \tag{13}$$

and

$$b_k(\Omega) = \frac{d}{d\Omega}\, a_k(\Omega) \qquad (p \times 1) \qquad k = 1, \ldots, p. \tag{14}$$

For $p > 1$, the scalars $a_k(\Omega)$ and vectors $b_k(\Omega)$ can be evaluated recursively with respect to the number of cisoids p. Put

$$\mathcal{A}_p = [a_1(\Omega_p), \ldots, a_p(\Omega_p)] \qquad (1 \times p) \tag{15}$$

$$\mathcal{B}_p = [b_1(\Omega_p), \ldots, b_p(\Omega_p)] \qquad (p \times p) \tag{16}$$

where Ω_p is Ω of dimension p, $p = 2, 3, \ldots$. It can easily be proved that

$$\mathcal{A}_{k+1} = [\mathcal{A}_k, 0] - e^{i\omega_{k+1}}\ [1, \mathcal{A}_k] \tag{17}$$

$$\mathcal{B}_{k+1} = \begin{bmatrix} [\mathcal{B}_k, 0_k] - e^{i\omega_{k+1}}\ [0_k, \mathcal{B}_k] \\ -ie^{i\omega_{k+1}}\ [1, \mathcal{A}_k] \end{bmatrix} \tag{18}$$

where $k = 1, \ldots, p-1$ and 0_k stands for the $k-$dimensional zero vector.

This procedure has a computational complexity of $\mathrm{O}(p^3)$ operations per data sample, due to the matrix inversion in (11) and the recursive scheme in (18). We claim that the number of the required operations can be reduced to $\mathrm{O}(p^2)$. The matrix $(\mathrm{Re}\ \{R_n\})^{-1}$ that appears in (11) can be updated in $\mathrm{O}(p^2)$ operations by aid of the matrix inversion lemma, as shown in [4], and the vector coefficients $b_1(\Omega_p), \ldots, b_p(\Omega_p)$ can be calculated in $\mathrm{O}(p^2)$ operations by the method given in [18][1].

Equilibrium state, [4]: $\bar{\psi}_n = i\, D_n Q$, and $\bar{R}_n = D_n \bar{R} D_n^*$, where

$$D_n = \mathrm{diag}\ (A_{1n}, \ldots, A_{pn}) \tag{19}$$

$$Q = [q_1, \ldots, q_p]^T \tag{20}$$

$$q_k = \frac{-\sum_{\ell=1}^{p} \ell\, a_\ell e^{-i\ell\omega_k}}{1 + \sum_{\ell=1}^{p} a_\ell\, \lambda^\ell e^{-i\ell\omega_k}} \approx \frac{1}{1-\lambda} \qquad k = 1, \ldots, p \tag{21}$$

"diag$(\cdot)$" means a diagonal matrix with the shown diagonal elements, and $\bar{R}$ is the $(p \times p)$ with elements

$$r_{jk} = \frac{q_j q_k^*}{1 - \rho e^{i(\omega_k - \omega_j)}} \qquad j, k = 1, \ldots, p. \tag{22}$$

The approximation in (21) is valid for λ close to 1.

[1] Our practical experience with the latter method is that it has lower complexity than the scheme (18) only for $p \geq 4$.

B. Multiple Frequency Tracker

Design variables: $\lambda, \rho \in (0, 1]$.
State-space parameters: the signal decomposition vector A_n, vector of the estimated frequencies $\hat{\Omega}_n = (\hat{\omega}_{1n}, \ldots, \hat{\omega}_{pn})^T$, and an auxilinary $(p \times p)$ matrix G_n.
Update Law:

$$
\begin{aligned}
G_n &= J\, J^T + \lambda\, \hat{F}_{n-1}\, G_{n-1}\, \hat{F}_{n-1}^H && (23)\\
\hat{A}_n &= \hat{F}_{n-1} \hat{A}_{n-1} + G_n^{-1} J \left[y_n - J^T \hat{F}_{n-1} \hat{A}_{n-1} \right] && (24)\\
\hat{\omega}_{kn} &= \hat{\omega}_{k,n-1} + (1-\rho)\, \mathrm{Arg} \left[\frac{\hat{A}_{kn}}{\hat{A}_{k,n-1} e^{i\hat{\omega}_{k,n-1}}} \right] \qquad k = 1, \ldots, p && (25)
\end{aligned}
$$

where

$$
\begin{aligned}
J &= (1, \ldots, 1)^T \qquad\qquad (p \times 1) && (26)\\
\hat{F}_n &= \mathrm{diag}\, (e^{i\hat{\omega}_{1n}}, \ldots, e^{i\hat{\omega}_{pn}}) && (27)
\end{aligned}
$$

This version of the procedure has computational complexity of $\mathrm{O}(p^3)$ operations per data sample, due to the matrix inversion in (24). For a large p, this number of operations can be reduced to $\mathrm{O}(p^2)$ by some factorization algorithm for update of the matrix G_n^{-1}, see Chapter 6 of [8].
Equilibrium state, [4] : $\bar{G}_n$ is a matrix independent of n, with entries

$$
\bar{g}_{jk} = \frac{1}{1 - \lambda\, e^{i(\omega_j - \omega_k)}}; \qquad j, k = 1, \ldots, p. \tag{28}
$$

C. Adaptive IIR Structure 2

Design variables: $\lambda \in (0, 1]$, and $\mu_1, \ldots, \mu_p \in (0, \infty)$.
State-space parameters: estimated poles $\hat{Z}_n = (\hat{z}_{1n}, \ldots, \hat{z}_{pn})^T$ of the filter, prediction error ε_n and its $p-1$ preceding values $\varepsilon_{n-1}, \ldots, \varepsilon_{n-p+1}$, and vector $Y_n = (y_{1n}^f, \ldots, y_{pn}^f)^T$ that contains the separated cisoids.
Update Law:

$$
\begin{aligned}
\varepsilon_n &= y_n + \sum_{k=1}^{p} a_k(\hat{Z}_{n-1}) \left[y_{n-k} - \lambda^k \varepsilon_{n-k} \right] && (29)\\
y_{kn}^f &= \varepsilon_n + \hat{z}_{k,n-1} y_{k,n-1}^f \qquad k = 1, \ldots, p && (30)\\
\hat{z}_{kn} &= \hat{z}_{k,n-1} + \mu_k (y_{k,n-1}^f)^* \varepsilon_n \qquad k = 1, \ldots, p && (31)
\end{aligned}
$$

where $a_k(\hat{Z}_{n-1}), k = 1, \ldots, p$ are the coefficients of the polynomial

$$
\prod_{k=1}^{p} (1 - \hat{z}_{k,n-1} q^{-1}) = 1 + a_1(\hat{Z}_{n-1}) q^{-1} + \ldots + a_p(\hat{Z}_{n-1}) q^{-p}. \tag{32}
$$

The instantaneous frequencies at time n are calculated as

$$
\hat{\omega}_{kn} = \mathrm{Arg}\ \hat{z}_{kn} \qquad k = 1, \ldots, p \tag{33}
$$

The computational complexity of one update in this procedure is $O(p^2)$, because this order of operations is required for calculating the polynomial coefficients in (29) using (13) where $e^{i\omega_{kn}}$ is replaced with z_{kn}.
Equilibrium state: $\bar{z}_{kn} = e^{i\omega_k}$,

$$\bar{y}^f_{kn} = \frac{A_{kn}}{1-\lambda} \prod_{\substack{j=1 \\ j\neq k}}^{p} \left(\frac{1 - e^{i(\omega_j - \omega_k)}}{1 - \lambda e^{i(\omega_j-\omega_k)}} \right) \approx \frac{A_{kn}}{1-\lambda}. \tag{34}$$

The approximation in (34) is good if λ close to one and the frequencies are well separated. (The proof follows easily from the definition.)

D. Hyperstable Adaptive Line Enhancer

Design variables: $g \in (0, 2/p)$ and $\mu \in (0, \infty)$. (Usually the option $g = 1/p$ is used, cf. [7].)
State-space parameters: vector of the polynomial coefficients $\theta_n = (a_{1n}, \ldots, a_{pn})^T$, the prediction error ε_n and its preceding $p-1$ values, $\varepsilon_{n-1}, \ldots, \varepsilon_{n-p+1}$.
Update Law:

$$\varepsilon_n = \frac{y_n - \theta^T_{n-1} X_n}{1 + \mu X^H_n X_n} \tag{35}$$

$$\theta_n = \theta_{n-1} + \mu X^*_n \varepsilon_n \tag{36}$$

where

$$X_n = \begin{bmatrix} -y_{n-1} + (1-g)\varepsilon_{n-1} \\ -y_{n-2} + (1-2g)\varepsilon_{n-2} \\ \vdots \\ -y_{n-p} + (1-pg)\varepsilon_{n-p} \end{bmatrix} \qquad (p \times 1) \tag{37}$$

The instantaneous frequencies at time n are calculated as angular positions of the roots of the polynomial equation

$$1 + a_{1n} q^{-1} + \ldots + a_{pn} q^{-p} = 0. \tag{38}$$

Equilibrium state: $\bar{\varepsilon}_n = 0$ and $\bar{\theta}_n$ is the vector of polynomial coefficients corresponding to the roots $e^{i\omega_k}$, $k = 1, \ldots, p$.

III. LOCAL BEHAVIOR OF THE ALGORITHMS, $p = 1$

The notation is simplified, here, by skipping the index k, so that the remaining indices are related to time.

A. Review of Known Results for MFT and ANF

Let $\{\delta\omega_n\}$ be the sequence of the frequency estimation errors, $\delta\omega_n = \hat{\omega}_n - \omega_n$, and

$$g_n = \operatorname{Im}\left\{ \frac{v_n}{A_n} \right\} \qquad n = 1, 2, \ldots. \tag{39}$$

be a "normalized" noise and

$$e_n = \omega_n - \omega_{n-1} \tag{40}$$

be the sequence of the frequency increments, for $n = 0, 1, 2, \ldots$. Then, the following holds [4].

Proposition 1 **(Linear filter approximation)** Assume that the frequency increments $\{e_n\}$, the relative increments of amplitudes $\{h_n\}$ and the normalized noise $\{g_n\}$ are uniformly small so that we can neglect higher than first-order moments of their elements. Then, the MFT and ANF applied to the signal (1) can be described by the identical linear filtering, written by aid of the z-transform as

$$\mathcal{Z}[\delta\omega_n] = \Phi_1(\lambda, \rho; z)\, \mathcal{Z}[g_n] + \Phi_2(\lambda, \rho; z)\, \mathcal{Z}[e_n] \tag{41}$$

where $\mathcal{Z}[\,\cdot\,]$ denotes the image of the sequence in the square brackets. The transfer functions in (41) are given by

$$\Phi_1(\lambda, \rho; z) = \frac{(1-\rho)(1-\lambda)(1-z^{-1})}{1-(2\lambda+\rho-\rho\lambda)z^{-1}+\lambda z^{-2}} \tag{42}$$

$$\Phi_2(\lambda, \rho; z) = -\frac{\lambda+\rho-\rho\lambda-\lambda z^{-1}}{1-(2\lambda+\rho-\rho\lambda)z^{-1}+\lambda z^{-2}}. \tag{43}$$

In particular, the frequency estimation errors do not depend on $\{h_n\}$ in the used approximation. ■

The local tracking ability of the algorithms is derived form the stability of the approximating linear filter in (41). Note that the poles of the ALF are complex-valued and mutually conjugate for $\rho < \rho_0$, where

$$\rho_0 = \frac{2\sqrt{\lambda}}{1+\sqrt{\lambda}}. \tag{44}$$

These poles are real-valued distinct for $\rho > \rho_0$, and complex-valued, mutually conjugated for $\rho < \rho_0$.

Also, it can be shown that for all λ and ρ in (0,1), both poles of the ALF lie inside the unit circle. In particular, if β denotes the pole with the maximum module, then

$$\beta = \begin{cases} \sqrt{\lambda} & \text{for} \quad \rho \le \rho_0 \\ \frac{1}{2}\left[2\lambda + \rho(1-\lambda) + \sqrt{(2\lambda+\rho-\rho\lambda)^2 - 4\lambda}\right] & \text{for} \quad \rho_0 < \rho < 1 \end{cases} \tag{45}$$

We call this β *local convergence factor*. It has the property that for a stationary noiseless cisoid, i.e. for the ALF with zero input and large n, the sequence $\{|\delta\omega_n|\}$ decays exponentially as $O(\beta^n)$ if the poles are distinct, and as $O(n\beta^n)$ in the case of the double pole). Based on this property, the so called *tracking time* is defined as $T_r = (-\log\beta)^{-1} \approx (1-\beta)^{-1}$, [4]. (This approximation is good for β close to 1.) The tracking time T_r characterizes the time constant related to transfer of state-space parameters of procedures between two equilibrium states after a small abrupt change of signal parameters. The same convergence factor and tracking time will be studied in the next section for $p > 1$.

For practical usage of the algorithms it is convenient to tune only one design variable of the procedures instead of two. In [4] it is shown, using the ALF approximation, that the optimum relation between ρ and λ in the sense of maximum noise rejection subject to a given relaxing time T_r, or equivalently subject to a given local convergence factor $\beta = \sqrt{\lambda}$, is $\rho = \rho_0(\lambda)$ in (44). In Section IV we show that this is approximately true also for $p > 1$.

Next, assume that the sequence $\{e_n\}$ in (40) is white noise with variance $\gamma^2 \ll 1$, statistically independent of the additive noise $\{v_n\}$. Then, it has been computed that

$$\mathrm{E}\{[\hat{\omega}_n - \omega_n]^2\} = \frac{\sigma^2}{|A_n|^2} \frac{(1-\lambda)(1-\rho)^2}{1+3\lambda+\rho(1-\lambda)} + \gamma^2 \frac{\rho^2(1-\lambda)^2 + 2\lambda(\rho+\lambda-2\rho\lambda)}{[1+3\lambda+\rho(1-\lambda)](1-\lambda)(1-\rho)}. \tag{46}$$

For a given SNR$=|A_n|^2/\sigma^2$ and γ^2, the expression (46) can be minimized analytically with respect to the design variables ρ and λ, respectively. The result is that the minimum is achieved for

$$\lambda = \lambda_\Gamma \triangleq 1 - \frac{\xi}{2} \approx 1 - \sqrt[4]{8\Gamma} \tag{47}$$

$$\rho = \rho_\Gamma \triangleq 1 - \frac{\xi}{4-\xi} \approx 1 - \sqrt[4]{\tfrac{1}{2}\Gamma} \tag{48}$$

and has the value

$$\mathrm{E}\{[\hat{\omega}_n - \omega_n]^2|_{\lambda=\lambda_\Gamma,\rho=\rho_\Gamma}\} = \gamma^2 \frac{\xi}{w} \approx \gamma^2 \sqrt[4]{2\Gamma^{-1}} \tag{49}$$

where

$$\xi = -w + \sqrt{w^2+4w} \tag{50}$$

$$w = \Gamma + \sqrt{\Gamma^2 + 8\Gamma} \tag{51}$$

$$\Gamma = \gamma^2\,\mathrm{SNR}. \tag{52}$$

The approximations in (47), (48) and (49) are valid for Γ close to zero.

B. Linear Filter Approximation for AIS-2 and HALE

We show that the performance of the AIS-2 and HALE algorithms for a single cisoid can be approximated by the same linear filter as in Proposition 1.

Proposition 2. The ALF for the AIS-2 algorithm with variables λ and μ_1 is

$$\mathcal{Z}[\delta\omega_n] = \Phi_1(\lambda, \rho_{\mathrm{AIS-2}}; z)\,\mathcal{Z}[g_n] + \Phi_2(\lambda, \rho_{\mathrm{AIS-2}}; z)\,\mathcal{Z}[e_n] \tag{53}$$

where $\Phi_1(.,.;z)$ and $\Phi_2(.,.;z)$ are defined in (42) and (43), respectively, and

$$\rho_{\mathrm{AIS-2}} = 1 - \frac{\mu_1|A_n|^2}{(1-\lambda)^2} \tag{54}$$

Proof: See the Appendix. ■

Proposition 3. The ALF for the HALE algorithm with variables g and μ is

$$\mathcal{Z}[\delta\omega_n] = \Phi_1(\lambda_{\text{HALE}}, \rho_{\text{HALE}}; z)\, \mathcal{Z}[g_n] + \Phi_2(\lambda_{\text{HALE}}, \rho_{\text{HALE}}; z)\, \mathcal{Z}[e_n] \tag{55}$$

where

$$\lambda_{\text{HALE}} = (1-\kappa)(1-g) \qquad \rho_{\text{HALE}} = \frac{g(1-\kappa)}{\kappa + g(1-\kappa)} \tag{56}$$

$$\kappa = \frac{\mu|A_n|^2}{1+\mu|A_n|^2} \tag{57}$$

Proof: See the Appendix. ■

The inversion formulas for κ and g that correspond to given λ and ρ read

$$\kappa = (1-\lambda)(1-\rho) \qquad g = \frac{\rho(1-\lambda)}{\lambda + \rho(1-\lambda)}. \tag{58}$$

We conclude that for $p = 1$ the AIS-2 and HALE algorithms have essentially the same local properties as the MFT and the ANF.

C. (λ, ρ) parameterization of AIS-2 and HALE

In this subsection we suggest a parameterization of AIS-2 and HALE in terms of λ and ρ for arbitrary p that has the property that the four methods, ANF, MFT, AIS-2 and HALE with the same λ and ρ have approximately the same tracking properties. The idea is to define design variables $\mu_1, \ldots, \mu_p$ for AIS-2, and g and μ for HALE, respectively, as functions of λ, ρ, and (possibly) of estimated signal parameters, so that for the single cisoid case the methods have the same ALF.

The solution of this task for AIS-2 is $\mu_k = (1-\rho)(1-\lambda)^2|A_{kn}|^{-2}$. If the sinusoidal magnitude is not known, we put

$$\mu_k(n) = \frac{1-\rho}{Y_k(n)} \qquad k = 1, \ldots, n \tag{59}$$

where $Y_k(n)$ approximates the term $|\bar{y}^f_{kn}|^2 \approx |A_{kn}|^2/(1-\lambda)^2$ and is calculated recursively as

$$Y_k(n) = \lambda\, Y_k(n-1) + (1-\lambda)\, |y^f_{kn}|^2 \qquad k = 1, \ldots, n \tag{60}$$

For HALE we define g and κ according to (58), and

$$\mu(n) = \frac{\kappa}{1-\kappa} \frac{p}{Y_0(n)} \tag{61}$$

where $Y_0(n) \approx |\bar{y}_n|^2 = |\sum_{k=1}^{p} A_{kn}|^2$ and is calculated recursively as

$$Y_0(n) = \lambda\, Y_0(n-1) + (1-\lambda)|y_n|^2 \tag{62}$$

Note that the parameter g calculated from (58) need not necessarily fall in the range $(0, 2/p)$. If this happens, the HALE algorithm cannot be easily compared with other methods.

In order to assure global convergence of the algorithm for $p > 1$ it is proposed to take $g = 1/p$, see [6]. This requirement rules out using the formulas (58): If g is an ad-hoc constant, it is in general impossible to fit the performance of the HALE algorithm to the ALF of algorithm MFT by tuning single variable μ. In this case we propose to choose μ, which possibly depend on the sinusoidal magnitudes, so that both algorithms will have the same local convergence factor, β. An easy calculation using (45) and (56) shows that this is fulfilled for $\mu(n)$ given by (61) with

$$\kappa = \begin{cases} 1 - \frac{\beta^2}{1-g} & \text{for} \quad \beta \leq \frac{2(1-g)}{2-g} \\ \frac{(1-\beta)(\beta+g-1)}{\beta(2-g)+g-1} & \text{otherwise} \end{cases} \tag{63}$$

where β can be calculated from (45). In the sequel we consider exclusively the latter $(\lambda, \rho)-$ parameterization of HALE.

IV. LOCAL BEHAVIOR OF THE ALGORITHMS, $p \geq 2$

In this section we extend the analysis of the algorithms to multiple cisoids.

Let $\{x_n\}$ be the noiseless signal, $x_n = y_n - v_n$, see (1). The equilibrium state of the parameters of the procedures for nonstationary signal obeys the equation

$$\bar{\Theta}_n = \mathcal{F}(\bar{\Theta}_{n-1}, x_n) + \Delta\bar{\Theta}_n \tag{64}$$

where $\Delta\bar{\Theta}_n$ is due to one step change of the signal parameters. Assume that the mapping $\mathcal{F}$ can be expanded in the Taylor series expansion around the point $(\bar{\Theta}_{n-1}, x_n)$,

$$\begin{aligned} \Theta_n = \mathcal{F}(\Theta_{n-1}, y_n) \;\; &= \;\; \mathcal{F}(\bar{\Theta}_{n-1}, x_n) + \frac{\partial \mathcal{F}}{\partial \Theta_{n-1}^T} (\Theta_{n-1} - \bar{\Theta}_{n-1}) + \frac{\partial \mathcal{F}}{\partial y_n} v_n \\ &+ \text{higher-order terms.} \end{aligned} \tag{65}$$

Subtracting (64) from (65) and neglecting the higher-than-first-order terms we obtain

$$\delta\Theta_n \stackrel{\triangle}{=} \Theta_n - \bar{\Theta}_n = F_n \, \delta\Theta_{n-1} + E_n \tag{66}$$

where $F_n = \partial\mathcal{F}/\partial\Theta_{n-1}^T$ is the matrix of partial derivatives and the term $E_n = \partial\mathcal{F}/\partial y_{n-1} \, v_n - \Delta\bar{\Theta}_n$ represents the influence of the additive noise and nonstationarity of the signal.

Tracking properties of the algorithm in a neighborhood of the equilibrium state, local stability of the system and the convergence factor β will be derived from the matrix sequence $\{F_n\}$. Note that these matrices depend on the design variables of the procedure, on the amplitudes, frequencies and initial phases of the cisoids.

The partial derivatives of the mapping $\mathcal{F}$ contained in F_n can be obtained for particular tracking algorithms by calculating the first-order terms in the Taylor series expansion of the corresponding recursions, as it is done in the derivation of the ALF for $p = 1$ in [4].

It may occur that the state-space parameters in Θ_n can be divided into two groups, say $\Theta_n = [\Theta_{n1}^T, \Theta_{n2}^T]^T$, so that Θ_{n1} contains the unknown frequencies, and has the property that the perturbation $\delta\Theta_{n1}$ is a (linear) function of $\delta\Theta_{n-1,1}$ only, independent of $\delta\Theta_{n-1,2}$. This means that the matrix F_n in (66) is block lower triangular.

For example, recall that in the ALF approximation ($p = 1$) the frequency estimation errors $\{\delta\omega_n\}$ do not depend on the magnitude estimation errors (in the considered first-order approximation).

Since our objective is to study tracking of the frequencies, we restrict the meaning of the term "tracking speed" to estimating of Θ_{n1}. We ignore the fact that the speed of tracking of Θ_{n2} can be different in general. For simplicity, in the sequel we shall use slightly inconsistent notation $\delta\Theta_n$ instead of $\delta\Theta_{n1}$, in spite of that it has less components than Θ_n. Similarly, F_n will mean the left upper corner submatrix of the "true" F_n with dimension corresponding to Θ_{n1}.

For example, note that the ALF in (41) can be written in the form (66) with

$$\delta\Theta_n^{\text{ALF}} = [\delta\omega_n, \delta\omega_{n-1}]^T \tag{67}$$

$$E_n^{\text{ALF}} = [(1-\lambda)(1-\rho)(g_n - g_{n-1}) + (\lambda+\rho-\rho\lambda)e_n - \lambda e_{n-1}, 0]^T \tag{68}$$

$$F_n^{\text{ALF}} = \begin{bmatrix} 2\lambda+\rho-\rho\lambda & -\lambda \\ 1 & 0 \end{bmatrix} \tag{69}$$

We can see that the matrix F_n^{ALF} in (69) does not depend explicitly neither on n nor on the signal parameters.

For $p=2$ we introduce the following notation: let ω and ϕ be the sinusoidal frequencies (constants), and $\delta\omega_n$ and $\delta\phi_n$ be the estimation errors at time n. Next, let A_n and B_n be the true values of the cisoids at time instant n. For the MFT, ANF and AIS-2 we put

$$\delta\Theta_n^{\text{MFT}} = \left[\frac{\delta A_n}{A_n}, \frac{\delta A_n^*}{A_n^*}, \frac{\delta B_n}{B_n}, \frac{\delta B_n^*}{B_n^*}, \delta\omega_n, \delta\phi_n\right]^T \tag{70}$$

$$\delta\Theta_n^{\text{ANF}} = \left[\frac{\varepsilon_n}{A_n}, \frac{\varepsilon_{n-1}}{A_{n-1}}, \frac{\varepsilon_n^*}{A_n^*}, \frac{\varepsilon_{n-1}^*}{A_{n-1}^*}, \delta\omega_n, \delta\phi_n\right]^T \tag{71}$$

$$\delta\Theta_n^{\text{AIS}-2} = \left[\frac{\varepsilon_n}{A_n}, \frac{\varepsilon_{n-1}}{A_{n-1}}, \frac{\delta z_{1n}}{e^{i\omega}}, \frac{\delta z_{2n}}{e^{i\psi}}\right]^T \tag{72}$$

$$\delta\Theta_n^{\text{HALE}} = \left[\frac{\varepsilon_n}{A_n}, \frac{\varepsilon_{n-1}}{A_{n-1}}, \frac{\delta a_{1n}}{e^{i\omega}}, \frac{\delta a_{2n}}{e^{2i\omega}}\right]^T \tag{73}$$

where, for AIS-2, $\delta z_{in} = z_{in} - \bar{z}_{in}$, $i=1,2$, and for HALE, $\delta a_{in} = a_{in} - \bar{a}_{in}$, $i=1,2$.

The explicit expressions of the corresponding matrices F_n^{MFT}, F_n^{ANF}, $F_n^{\text{AIS}-2}$ and F_n^{HALE} are derived in the appendix.

Considering the formula (66), the local convergence factor β can be defined as

$$\beta = \limsup_{n\to\infty} [\sigma_{\max}(F_n F_{n-1} \ldots F_1)]^{1/n} \tag{74}$$

where $\sigma_{\max}(M)$ denotes the largest singular value of the matrix M or, in other words, the *spectral norm* of the matrix. Note that $\sigma_{\max}(M)$ is equal to maximum of $\|Mx\|$ over all vectors x of the appropriate dimension with $\|x\|=1$, where $\|\cdot\|$ denotes the Euclidean norm; It can be computed as the square root of the maximum eigenvalue of $M^H M$.

It can be easily seen that β according to this definition has the following desirable property: for $\beta < 1$ the system is stable, and the deviations of the state-space parameters from their equilibrium values go to zero as $\mathrm{O}(\beta^n)$ for $n\to\infty$ or, more exactly, as $\mathrm{o}(\tilde{\beta}^n)$ for all $\tilde{\beta} > \beta$.

Obviously, for $p=1$, F_n does not depend on n, cf. (69), $F_n = F$, and β can be calculated according to the following lemma.

Lemma 1 For any square matrix F it holds that

$$\lim_{n\to\infty} [\sigma_{\max}(F^n)]^{1/n} = |\lambda_{\max}(F)| \tag{75}$$

where $\lambda_{\max}(F)$ denotes the eigenvalue of F with the largest magnitude.
Proof: See Theorem 10.13 in [22]. ■

It can be easily shown that the magnitude of the maximum eigenvalue of F_n^{ALF} in (69) exactly coincides with β given in (45), as desired.

Next, note that for $p = 2$, the matrices F_n depend on n only through of the terms $e^{in(\omega-\phi)}$. It follows that if the frequency separation is a rational factor of 2π, say $\omega - \phi = \frac{\ell}{k} 2\pi$, the matrix sequence $\{F_n\}$ is periodic with period k, i.e., $F_{n+k} = F_n$ for all n. If this happens, then β in (74) can be computed as

$$\begin{aligned}
\beta &= \limsup_{n\to\infty}[\sigma_{\max}(F_nF_{n-1}\dots F_1)]^{1/n} = \limsup_{\substack{n=km\\ m\to\infty}}[\sigma_{\max}(F_nF_{n-1}\dots F_1)]^{1/n} \\
&= \limsup_{m\to\infty}[\sigma_{\max}((F_kF_{k-1}\dots F_1)^m)]^{1/(mk)} = \left\{\lim_{m\to\infty}[\sigma_{\max}(F_kF_{k-1}\dots F_1)^m]^{1/m}\right\}^{1/k} \\
&= |\lambda_{\max}(F_kF_{k-1}\dots F_1)|^{1/k}
\end{aligned} \tag{76}$$

The second equality in (76) is proved in the Appendix.

Proposition 4 For the MFT, ANF and AIS-2, the local convergence factor for $p = 2$ does not depend on the magnitudes of the cisoids. It is a function of only the design variables λ, ρ of the procedures, the frequency separation and the initial phase separation of the cisoids.
Proof: See appendix.

Our computer simulations show that β is also independent of the initial phase difference, but so far we have no analytical proof of this conjecture.

In contrast to Proposition 4, the local convergence factor for HALE depends on the sinusoidal magnitudes, also: it appears that if the ratio of the two magnitudes goes to zero or infinity, than β goes to 1, which means that the tracking speed of the algorithm goes to 0.

We conclude this section by numerical results.

A. Optimality of ρ_0

In the case of $p = 1$ it was shown that the relationship $\rho = \rho_0(\lambda)$ in (44) between the design variables of the algorithms leads to maximum noise rejection subject to a given local convergence factor β. In this subsection we show by numerical examples, that the relation $\rho = \rho_0(\lambda)$ is a good choice also for $p = 2$.

Consider a signal model with two well separated frequencies, $\omega - \phi = \pi/3$, fixed λ ($\lambda = 0.7$ and $\lambda = 0.9$), and varying ρ. The corresponding β is plotted in Figure 1. It follows from the figures that in these cases β behaves similarly to the case of $p = 1$ in (45): for $\rho \leq \rho_0$, β is approximately constant (very close to $\sqrt{\lambda}$ for the MFT), and for $\rho > \rho_0$ it rapidly increases with growing ρ. For constant β and λ, the largest value of ρ is preferable because it leads to the best noise rejection, as stated.

From a theoretical point of view, it is worth noting that for very small values of ρ, all of the three algorithms, MFT, ANF and AIS-2, may become locally unstable, i.e. $\beta > 1$.

B. Convergence factor versus frequency separation

In this example we take constant design variables $\lambda = 0.6$, $\lambda = 0.9$ and the corresponding $\rho = \rho_0(\lambda)$, and plot β versus varying frequency separation of the cisoids. The results are shown in Figure 2.

The figure shows that for well separate cisoids, β is approximately constant; the minimum value is achieved by MFT, where it corresponds to the value predicted for $p = 1$, $\beta = \sqrt{\lambda}$. When the frequency separation converges to 0, the only method where β is well isolated from 1 is the ANF. The other three methods, MFT, AIS-2 and HALE either fail in resolving cisoids with closely spaced frequencies ($\beta > 1$), or the convergence is slow ($\beta \approx 1$).

C. Minimum achievable β

In this example we consider cisoids with a constant angle separation, relatively large, $\pi/3$, and a relatively small one, $\pi/33$. The variables λ and ρ are varying, satisfying the relation $\rho = \rho_0(\lambda)$. The results are shown on Figure 3.

The figure shows that for the MFT and AIS-2 there exists a minimum rate of convergence for each frequency separation, that cannot be overcome by any choice of the design variables. The other two algorithms, ANF and HALE, have theoretically unlimited tracking speed for $\lambda \to 0$. It is, however, worth noting that β characterizes the *asymptotic* tracking speed for $n \to \infty$, only. The empirical tracking delay is usually longer than $(-\log \beta)^{-1}$ for moderate frequency changes, especially for small β. The additional tracking delay is caused by the nonlinearity of the algorithms, and in particular it is long in the case of the ANF, see [4] and examples in the next section.

V. GLOBAL PROPERTIES OF THE ALGORITHMS

In this section, all four algorithms are tested by aid of computer simulations for the properties that cannot be derived from the linear approximation given in the previous section.

A. Global convergence

Consider a model with two stationary noiseless cisoids with constant unit magnitudes and well separated frequencies, $\omega = -\pi/3$, $\phi = \pi/3$. The algorithms are initialized with these frequencies: the first one is correct, equal to ω, and the other one is wrong and varying in the range $(-\pi, \pi)$. The initial frequencies and first 5 date samples of the signal are used to estimate the amplitudes and phases of the cisoids, and consequently, the equilibrium state of the procedures. Starting from these ES, the time in which both the estimated frequencies fall into the ± 0.1 tolerance interval around the true frequencies is measured. The result for $\lambda = 0.8$ and $\rho = \rho_0$ is shown on Figure 4.

The Figure shows poor global convergence of AIS-2, which sometimes tends to converge to false frequency estimates. HALE exhibits the best uniform rate of convergence. ANF converges somewhat slower, to the true frequencies, also. MFT converges as rapidly as HALE for a wide range of the initial frequencies among all methods, but it has some numerical problems, if the initial estimates lie close to the other frequency. The later frequency tends to follow the former one. If some careful implementation of MFT can avoid this problem, the algorithm (together with HALE) becomes the best globally converging method.

B. Noise rejection

In this example we consider a model with two cisoids of constant unit amplitude, one frequency is constant and the second one is slowly varying as shown in Figure 5.

The signal is embedded in additive noise with SNR varying in wide range. The procedures are initialized using true equilibrium states, and the design parameters are chosen (a) constant, $\lambda = 0.9$, $\rho = \rho_0(\lambda)$, and (b) varying with respect to the additive noise. In the later example we have used the formulas (47)-(48), where we inserted the true SNR, and $\gamma^2 = (1/200)^2$, that was estimated from the time-varying frequency; the resultant λ decreases from the value 0.9304 for SNR=-20 dB to the value 0.8218 for SNR=20 dB. For each SNR we measured the mean-square frequency estimation error (over both frequencies). The results are shown on Figure 6.

We note that the mean-square estimation error in the example (a) with constant λ and ρ does not go to zero as SNR$\rightarrow \infty$, because the methods have nonzero tracking delay even for noiseless signal. If the SNR is high, the tracking delay is minimum for the HALE algorithm, and maximum for the ANF. However, at high SNR $\lambda = 0.9$ is too big: for the optimum λ in Example 5B, the mean-square tracking error is minimum for MFT and maximum for HALE. The ANF algorithm exhibits the largest (worst) SNR threshold between all four algorithms. Note that the last observation apparently contradicts the simulation result in [4]. The difference is that in [4] the authors considered tracking of constant frequencies only, and the worse tracking ability of ANF was not taken into account.

VI. CONCLUSIONS

The linear filter approximation technique for analyzing algorithms for tracking of a single slowly varying cisoid in noise was extended to the multiple cisoids case. Based on the results for the single cisoid case, the four algorithms, MFT, ANF, AIS-2 and HALE were all parameterized (reparameterized) in terms of the identical design variables λ and ρ, which made the comparison easier. The comparative study showed that:

- For $p = 1$ all four algorithms have the same approximating linear filter.
- The local convergence factor calculated for a single cisoid applies approximately to the multiple cisoids case well, if the sinusoidal frequencies are well separated. The agreement is especially good for MFT. The choice $\rho = \rho_0(\lambda)$ is judicious for all four methods in the multiple case as well.
- Global convergence of the algorithms is 100% for ANF and HALE, the convergence of HALE is about twice faster than the convergence of ANF. The convergence of MFT is as rapid as the convergence of HALE except for the cases where the initial frequencies lie close each other. AIS-2 converges eventually to false frequency estimates.
- Among the four algorithms, ANF has the largest number of the internal parameters, what follows the longest time delay in tracking of rapidly changing frequencies.
- The only algorithm that resolves well closely-spaced frequencies is ANF.
- For $p = 2$ the local tracking speed of the MFT, ANF and AIS-2 does not depend on the sinusoidal magnitudes. For HALE the tracking speed is highest if both cisoids have the same magnitude.

Acknowledgement

The authors wish to acknowledge Dr. Mukund Padmanabhan for his assistance in the formulation of the HALE algorithm.

APPENDIX

Proof of the second equality in (68)

Put

$$s_n = [\sigma_{\max}(F_n F_{n-1} \ldots F_1)]^{1/n} \tag{77}$$

It is sufficient to show that

$$\limsup_{\substack{n=km+\ell \\ m\to\infty}} s_n = \limsup_{\substack{n=km \\ m\to\infty}} s_n \tag{78}$$

for all $\ell = 1, 2, \ldots, k-1$. We use the property of the spectral matrix norm that $\sigma_{\max}(AB) \le \sigma_{\max}(A)\,\sigma_{\max}(B)$ for arbitrary matrices A, B. It holds that

$$\begin{aligned} s_{km+\ell} &= [\sigma_{\max}(F_{km+\ell}F_{km+\ell-1}\ldots F_1)]^{1/(km+\ell)} \\ &\le [\sigma_{\max}(F_\ell F_{\ell-1}\ldots F_1)\,\sigma_{\max}(F_{km}F_{km-1}\ldots F_1)]^{1/(km+\ell)} \\ &= [\sigma_{\max}(F_\ell F_{\ell-1}\ldots F_1)]^{1/(km+\ell)}\, s_{km}^{km/(km+\ell)} \end{aligned} \tag{79}$$

Taking the limes superior from both sides of (79) for $m \to \infty$ we get

$$\limsup_{\substack{n=km+\ell \\ m\to\infty}} s_n \le \limsup_{\substack{n=km \\ m\to\infty}} s_n \tag{80}$$

Similarly we get the opposite inequality. The statement follows.

Proof of Proposition 2

For $p = 1$ we have $a_1(\hat{Z}_{n-1}) = -\hat{z}_{n-1}$, and the algorithm AIS-2 can be written as

$$\varepsilon_n = y_n - \hat{z}_{n-1}[y_{n-1} - \lambda\varepsilon_{n-1}] \tag{81}$$

$$\hat{z}_n = \hat{z}_{n-1} + \mu_1 (y^f_{n-1})^* \varepsilon_n \tag{82}$$

$$\hat{\omega}_n = \mathrm{Im}\,\log \hat{z}_n \tag{83}$$

For simplicity we assume that the signal is stationary. Considering the relations $y_n = A_n + v_n$, $\hat{\omega}_n = \omega + \delta\omega_n$ and $\hat{z}_n = e^{i\omega} + \delta z_n$, and neglecting higher-than-first-order small terms we get

$$\varepsilon_n = v_n - \delta z_{n-1}A_{n-1} - e^{i\omega}[v_{n-1} - \lambda\varepsilon_{n-1}] \tag{84}$$

$$\delta z_n = \delta z_{n-1} + \mu_1(\bar{y}^f_{n-1})^*\varepsilon_n = \delta z_{n-1} + \mu_1 A^*_{n-1}(1-\lambda)^{-1}\varepsilon_n \tag{85}$$

$$\delta\omega_n = \mathrm{Im}\,(\delta z_n e^{-i\omega}) \tag{86}$$

Using the notation (39) we easily get

$$\vartheta_n \triangleq \mathrm{Im}(\varepsilon_n / A_n) = g_n - \delta\omega_{n-1} - g_{n-1} + \lambda\vartheta_{n-1} \tag{87}$$

$$\delta\omega_n = \delta\omega_{n-1} + \mu_1 |A_n|^2 (1-\lambda)^{-1}\vartheta_n \tag{88}$$

Applying the $z-$transform and excluding $\mathcal{Z}[\vartheta_n]$ we obtain the filtering

$$\mathcal{Z}[\delta\omega_n] = \Phi_1(\lambda, \rho_{\text{AIS}-2}; z)\,\mathcal{Z}[g_n] \tag{89}$$

where $\Phi_1(\lambda, \rho_{\text{AIS}-2}; z)$ and $\rho_{\text{AIS}-2}$ are defined in (42) and (54), respectively. The transfer function $\Phi_2(\lambda, \rho_{\text{AIS}-2}; z)$ can be verified similarly.

Proof of Proposition 3

For $p=1$ we have $\theta_n = a_{1n} = -\hat{z}_n$, hence the algorithm HALE can be written as

$$\begin{aligned} \varepsilon_n &= \frac{y_n + \hat{z}_{n-1}^T X_n}{1+\mu|X_n|^2} && (90)\\ \hat{z}_n &= \hat{z}_{n-1} - \mu X_n^* \varepsilon_n && (91)\\ \hat{\omega}_n &= \operatorname{Im}\log \hat{z}_n && (92)\end{aligned}$$

where

$$X_n = -y_{n-1} + (1-g)\varepsilon_{n-1}. \tag{93}$$

Again assume first that the signal is stationary. In the first-order approximation we get

$$\begin{aligned} \varepsilon_n &= \frac{v_n - \delta z_{n-1}^T A_{n-1} + e^{i\omega}[-v_{n-1} + (1-g)\varepsilon_{n-1}]}{1+\mu|A_n|^2} && (94)\\ \delta z_n &= \delta z_{n-1} + \mu A_{n-1}^* \varepsilon_n && (95)\\ \delta\omega_n &= \operatorname{Im}\left(\delta z_n e^{-i\omega}\right) && (96)\end{aligned}$$

Using the notation (39) we easily get

$$\begin{aligned} \vartheta_n &\triangleq \operatorname{Im}(\varepsilon_n/A_n) = \frac{g_n - \delta\omega_{n-1} - g_{n-1} + (1-g)\vartheta_{n-1}}{1+\mu|A_n|^2} && (97)\\ \delta\omega_n &= \delta\omega_{n-1} + \mu|A_n|^2\vartheta_n && (98)\end{aligned}$$

Applying the $z-$transform and excluding $\mathcal{Z}[\vartheta_n]$ we obtain the filtering

$$\mathcal{Z}[\delta\omega_n] = \frac{\kappa(1-z^{-1})}{1 - z^{-1}(2-2\kappa-g+\kappa g) + z^{-2}(1-\kappa)(1-g)}\,\mathcal{Z}[g_n] \tag{99}$$

Comparing the coefficients in the transfer function with those in (42) we get the relations (56). The correctness of the transfer function $\Phi_2(\lambda_{\mathrm{HALE}}, \rho_{\mathrm{HALE}}; z)$ can be verified similarly.

Analysis of the methods for $p = 2$

The signal with two cisoids has the form

$$y_n = A_n + B_n + v_n$$

with

$$\bar{A}_n = A_0 e^{in\omega} \qquad \bar{B}_n = B_0 e^{in\phi}.$$

Put

$$\begin{aligned} \tilde{\rho} &= 1-\rho && (100)\\ \chi &= B_0/A_0 \qquad z = e^{i(\omega-\phi)} && (101)\\ \alpha_n &= \frac{\delta A_n}{\bar{A}_n} \qquad \beta_n = \frac{\delta B_n}{\bar{B}_n} && (102)\end{aligned}$$

Thus we have

(103)

$$\frac{\bar{B}_n}{\bar{A}_n} = \chi z^{-n} \tag{104}$$

Analysis of MFT

The signal decomposition update has the form

$$\begin{aligned}\hat{A}_n &= e^{i\hat{\omega}_{n-1}}\hat{A}_{n-1} + \eta[y_n - e^{i\hat{\omega}_{n-1}}\hat{A}_{n-1} - e^{i\hat{\phi}_{n-1}}\hat{B}_{n-1}] \\ \hat{B}_n &= e^{i\hat{\phi}_{n-1}}\hat{B}_{n-1} + \eta^*[y_n - e^{i\hat{\omega}_{n-1}}\hat{A}_{n-1} - e^{i\hat{\phi}_{n-1}}\hat{B}_{n-1}]\end{aligned}$$

where η is the first element of $\bar{G}_n^{-1}J$,

$$\bar{G}_n = G = \begin{bmatrix} \frac{1}{1-\lambda} & \frac{1}{1-\lambda z} \\ \frac{1}{1-\lambda z^*} & \frac{1}{1-\lambda} \end{bmatrix} \qquad G^{-1}J = \begin{bmatrix} \eta \\ \eta^* \end{bmatrix}$$

$$\eta = \frac{(1-\lambda)(1-z)(1-\lambda z^{-1})}{2(1-\operatorname{Re} z)}$$

In the first-order approximation it holds

$$\begin{aligned}e^{i\hat{\omega}_{n-1}} &= e^{i\omega} + ie^{i\omega}(\delta\omega_{n-1} - e_n) \\ e^{i\hat{\phi}_{n-1}} &= e^{i\phi} + ie^{i\phi}(\delta\phi_{n-1} - f_n)\end{aligned}$$

Thus we get

$$\begin{aligned}\delta A_n &= e^{i\omega}\delta A_{n-1} + i\bar{A}_n(\delta\omega_{n-1} - e_n) + \eta[v_n - e^{i\omega}\delta A_{n-1} - i\bar{A}_n(\delta\omega_{n-1} - e_n) \\ &\qquad - e^{i\phi}\delta B_{n-1} - i\bar{B}_n(\delta\phi_{n-1} - f_n)] \\ \delta B_n &= e^{i\phi}\delta B_{n-1} + i\bar{B}_n(\delta\phi_{n-1} - f_n) + \eta^*[v_n - e^{i\omega}\delta A_{n-1} - i\bar{A}_n(\delta\omega_{n-1} - e_n) \\ &\qquad - e^{i\phi}\delta B_{n-1} - i\bar{B}_n(\delta\phi_{n-1} - f_n)]\end{aligned}$$

Hence

$$\begin{aligned}\alpha_n &= (1-\eta)\alpha_{n-1} - \eta\chi z^{-n}\beta_{n-1} + i(1-\eta)(\delta\omega_{n-1} - e_n) - i\eta\chi z^{-n}(\delta\phi_{n-1} - f_n) + \eta v_n/\bar{A}_n \\ \alpha_n^* &= (1-\eta^*)\alpha_{n-1}^* - \eta^*\chi^* z^n\beta_{n-1}^* - i(1-\eta^*)(\delta\omega_{n-1} - e_n) + i\eta^*\chi^* z^n(\delta\phi_{n-1} - f_n) + (\eta v_n/\bar{A}_n)^* \\ \beta_n &= (1-\eta^*)\beta_{n-1} - \eta^*\chi^{-1}z^n\alpha_{n-1} + i(1-\eta^*)(\delta\phi_{n-1} - f_n) - i\eta^*\chi^{-1}z^n(\delta\omega_{n-1} - e_n) + \eta^* v_n/\bar{B}_n \\ \beta_n^* &= (1-\eta)\beta_{n-1}^* - \eta\chi^{-*}z^{-n}\alpha_{n-1}^* - i(1-\eta)(\delta\phi_{n-1} - f_n) + i\eta\chi^{-*}z^{-n}(\delta\omega_{n-1} - e_n) + (\eta^* v_n/\bar{B}_n)^* \\ \delta\omega_n &= (1-\rho)\left[\operatorname{Im}\frac{\delta A_n}{\bar{A}_n} - \operatorname{Im}\frac{\delta A_{n-1}}{\bar{A}_{n-1}}\right] + \rho(\delta\omega_{n-1} - e_n) \\ &= \frac{\tilde{\rho}}{2i}[\alpha_n - \alpha_n^* - \alpha_{n-1} + \alpha_{n-1}^*] + \rho(\delta\omega_{n-1} - e_n) \\ \delta\phi_n &= (1-\rho)\left[\operatorname{Im}\frac{\delta B_n}{\bar{B}_n} - \operatorname{Im}\frac{\delta B_{n-1}}{\bar{B}_{n-1}}\right] + \rho(\delta\phi_{n-1} - f_n) \\ &= \frac{\tilde{\rho}}{2i}[\beta_n - \beta_n^* - \beta_{n-1} + \beta_{n-1}^*] + \rho(\delta\phi_{n-1} - f_n)\end{aligned}$$

In the matrix form we can write $\Theta_n = F_n\Theta_{n-1} + \mathcal{E}_n$ with

$$\begin{aligned}\Theta_n &= [\alpha_n, \alpha_n^*, \delta\omega_n, \beta_n, \beta_n^*, \delta\phi_n]^T \\ F_n^{\mathrm{MFT}} &= \begin{bmatrix} 1-\eta & 0 & i(1-\eta) & -\eta\chi z^{-n} & 0 & -i\eta\chi z^{-n} \\ 0 & 1-\eta^* & -i(1-\eta^*) & 0 & -\eta^*\chi^* z^n & i\eta^*\chi^* z^n \\ \frac{i}{2}\tilde{\rho}\eta & -\frac{i}{2}\tilde{\rho}\eta^* & 1-\tilde{\rho}\operatorname{Re}\eta & -\frac{i}{2}\tilde{\rho}\eta\chi z^{-n} & -\frac{i}{2}\tilde{\rho}\eta^*\chi^* z^n & -\tilde{\rho}\operatorname{Re}(\eta\chi z^{-n}) \\ -\eta^*\chi^{-1}z^n & 0 & -i\eta^*\chi^{-1}z^n & 1-\eta^* & 0 & i(1-\eta^*) \\ 0 & -\eta\chi^{-*}z^n & i\eta\chi^{-*}z^n & 0 & 1-\eta & -i(1-\eta) \\ -\frac{i}{2}\tilde{\rho}\eta^*\chi^{-1}z^n & -\frac{i}{2}\tilde{\rho}\eta\chi^{-*}z^{-n} & -\tilde{\rho}\operatorname{Re}(\eta^*\chi^{-1}z^n) & \frac{i}{2}\tilde{\rho}\eta^* & -\frac{i}{2}\tilde{\rho}\eta & 1-\tilde{\rho}\operatorname{Re}\eta \end{bmatrix}\end{aligned}$$

$$
\mathcal{E}_n = \begin{bmatrix} \eta v_n/\bar{A}_n \\ (\eta v_n/\bar{A}_n)^* \\ \tilde{\rho}\mathrm{Im}(\eta v_n/\bar{A}_n) \\ \eta^* v_n/\bar{B}_n \\ (\eta^* v_n/\bar{B}_n)^* \\ \tilde{\rho}\mathrm{Im}(\eta^* v_n/\bar{B}_n) \end{bmatrix} + \begin{bmatrix} -i(1-\eta) \\ i(1-\eta^*) \\ -1+\tilde{\rho}\mathrm{Re}\,\eta \\ i\eta^*\chi^{-1}z^n \\ -i\eta\chi^{-*}z^{-n} \\ \tilde{\rho}\mathrm{Re}(\eta^*\chi^{-1}z^n) \end{bmatrix} e_n + \begin{bmatrix} i\eta\chi z^{-n} \\ -i\eta^*\chi^* z^n \\ \tilde{\rho}\mathrm{Re}(\eta\chi z^{-n}) \\ -i(1-\eta^*) \\ i(1-\eta) \\ -1+\tilde{\rho}\mathrm{Re}\,\eta \end{bmatrix} f_n
$$

Analysis of ANF

The algorithm is given by the recursions

$$
\begin{aligned}
\hat{\theta}_n &= \hat{\theta}_{n-1} + (\mathrm{Re}\,\{R_n\})^{-1}\,\mathrm{Re}\,\{\psi_n^* \varepsilon_n\} \qquad \theta_n = [\omega_n, \phi_n]^T \\
\varepsilon_n &= y_n + \sum_{k=1}^{p} a_k(\hat{\theta}_{n-1}) \left[y_{n-k} - \lambda^k \varepsilon_{n-k} \right]
\end{aligned}
$$

where

$$
a_1(\theta) = -e^{i\omega} - e^{i\phi} \qquad a_2(\theta) = e^{i\omega}e^{i\phi} \qquad \text{for} \quad \theta = (\omega, \phi).
$$

In the first-order approximation we can replace R_n and ψ_n by their equilibrium values $\bar{R}_n$ and $\bar{\psi}_n$. From (21) we find

$$
\begin{aligned}
q_1 &= \frac{-a_1 e^{-i\omega} - 2a_2 e^{-2i\omega}}{1 + a_1\lambda e^{-i\omega} + a_2\lambda^2 e^{-2i\omega}} = \frac{1 - z^{-1}}{(1-\lambda)(1-\lambda z^{-1})} \triangleq q^* \qquad q_2 = q \\
\bar{\psi}_n &= \begin{bmatrix} iA_n q^* \\ iB_n q \end{bmatrix} = iA_n \begin{bmatrix} q^* \\ \chi q z^{-n} \end{bmatrix} \qquad R = \begin{bmatrix} \frac{|q|^2}{\tilde{\rho}} & \frac{q^{*2}}{\tilde{\rho}z^{-1}} \\ \frac{q^2}{\tilde{\rho}z} & \frac{|q|^2}{\tilde{\rho}} \end{bmatrix} \\
\bar{R}_n &= \mathrm{diag}(A_n, B_n)\, R\, \mathrm{diag}(A_n^*, B_n^*) = |A_n|^2 \begin{bmatrix} \frac{|q|^2}{\tilde{\rho}} & \frac{q^{*2}\chi^* z^n}{\tilde{\rho}z^{-1}} \\ \frac{q^2\chi z^{-n}}{\tilde{\rho}z} & \frac{|\chi|^2|q|^2}{\tilde{\rho}} \end{bmatrix} \\
\left\{\mathrm{Re}\,\bar{R}_n\right\}^{-1} &= \frac{1}{|A_n|^2} \begin{bmatrix} |\chi|^2 r_n & s_n \\ s_n & r_n \end{bmatrix} \\
r_n &= \frac{|q|^2}{\tilde{\rho}} \left[\frac{|\chi|^2|q|^4}{\tilde{\rho}^2} - \left(\mathrm{Re}\,\frac{q^2\chi z^{-n}}{\tilde{\rho}z} \right)^2 \right]^{-1} \\
s_n &= -\mathrm{Re}\,\frac{q^2\chi z^{-n}}{\tilde{\rho}z} \left[\frac{|\chi|^2|q|^4}{\tilde{\rho}^2} - \left(\mathrm{Re}\,\frac{q^2\chi z^{-n}}{\tilde{\rho}z} \right)^2 \right]^{-1}
\end{aligned}
$$

Now, we can write

$$
\begin{aligned}
\hat{\theta}_n &= \hat{\theta}_{n-1} + \vartheta_n \varepsilon_n^A + \vartheta_n^* \varepsilon_n^{A*} \qquad \text{for} \quad \varepsilon_n^A = \varepsilon_n / A_n \\
\vartheta_n &= \frac{1}{2}(\mathrm{Re}\,\bar{R}_n)^{-1}\psi_n^* A_n = -\frac{i}{2} \begin{bmatrix} |\chi|^2 q r_n + \chi^* q^* z^n s_n \\ q s_n + \chi^* q^* z^n r_n \end{bmatrix} \triangleq \begin{bmatrix} u_n \\ w_n \end{bmatrix}
\end{aligned}
$$

Next,

$$
\varepsilon_n = y_n - \left(e^{i\hat{\omega}_{n-1}} + e^{i\hat{\phi}_{n-1}} \right) [y_{n-1} - \lambda\varepsilon_{n-1}] + e^{i\hat{\omega}_{n-1}} e^{i\hat{\phi}_{n-1}} [y_{n-2} - \lambda^2 \varepsilon_{n-2}] =
$$

$$
\begin{aligned}
&= \lambda\left(e^{i\omega}+e^{i\phi}\right)\varepsilon_{n-1} - \lambda^2 e^{i\omega}e^{i\phi}\varepsilon_{n-2} - iA_n(1-z^{-1})\delta\omega_{n-1} - iB_n(1-z)\delta\phi_{n-1} \\
&\quad + iA_n e_n + iB_n f_n + v_n - \left(e^{i\omega}+e^{i\phi}\right)v_{n-1} + e^{i\omega}e^{i\phi}v_{n-2} \\
\varepsilon_n^A &= \lambda(1+z^{-1})\varepsilon_{n-1}^A - \lambda^2 z^{-1}\varepsilon_{n-2}^A - i(1-z^{-1})\delta\omega_{n-1} - i\chi z^{-n}(1-z)\delta\phi_{n-1} \\
&\quad + ie_n + i\chi z^{-n} f_n + v_n^A - (1-z^{-1})v_{n-1}^A + z^{-1}v_{n-2}^A
\end{aligned}
$$

In the matrix form we can write $\Theta_n = F_n\Theta_{n-1} + \mathcal{E}_n$ with

$$
\begin{aligned}
\Theta_n &= [\varepsilon_n^A, \varepsilon_{n-1}^A, \varepsilon_n^{A*}, \varepsilon_{n-1}^{A*}, \delta\omega_n, \delta\phi_n]^T \\
F_n &=
\end{aligned}
$$

$$
\begin{bmatrix}
\lambda(1+z^{-1}) & -\lambda^2 z^{-1} & 0 & 0 & -i(1-z^{-1}) & -i\chi z^{-n}(1-z) \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \lambda(1+z) & -\lambda^2 z & i(1-z) & i\chi^* z^n(1-z^{-1}) \\
0 & 0 & 1 & 0 & 0 & 0 \\
\lambda(1+z^{-1})u_n & -\lambda^2 z^{-1}u_n & \lambda(1+z)u_n^* & -\lambda^2 z u_n^* & 1+2\mathrm{Im}[(1-z^{-1})u_n] & 2\mathrm{Im}[\chi z^{-n}(1-z)u_n] \\
\lambda(1+z^{-1})w_n & -\lambda^2 z^{-1}w_n & \lambda(1+z)w_n^* & -\lambda^2 z w_n^* & 2\mathrm{Im}[(1-z^{-1})w_n] & 1+2\mathrm{Im}[\chi z^{-n}(1-z)w_n]
\end{bmatrix}
$$

Analysis of AIS-2

$$
\begin{aligned}
\varepsilon_n &= y_n - (\hat{z}_{n-1}^{(1)} + \hat{z}_{n-1}^{(2)})[y_{n-1} - \lambda\varepsilon_{n-1}] + \hat{z}_{n-1}^{(1)}\hat{z}_{n-1}^{(1)}[y_{n-2} - \lambda^2\varepsilon_{n-2}] \\
Y_n &= \varepsilon_n J + \mathrm{diag}(\hat{z}_{n-1}^{(1)}, \hat{z}_{n-1}^{(2)})\,Y_{n-1} \qquad \bar{Y}_n = [q^* A_n, qB_n]^T \\
Y_n^{(1)} &= \frac{1-\bar{z}^{(2)}z^{-1}}{(1-\lambda\bar{z}^{(1)}z^{-1})(1-\lambda\bar{z}^{(2)})}\,y_n = \frac{1-e^{i\phi}z^{-1}}{(1-\lambda e^{i\omega}z^{-1})(1-\lambda e^{i\phi}z^{-1})}\,A_n = q^* A_n \\
\hat{Z}_n &= \hat{Z}_{n-1} + \mathrm{diag}(\mu_1,\mu_2)\varepsilon_n Y_{n-1}^* \\
\mu_1 &= (1-\lambda)^2(1-\rho)/|A_n|^2 \qquad \mu_2 = (1-\lambda)^2(1-\rho)/|B_n|^2 \\
\varepsilon_n &= \lambda\left(e^{i\omega}+e^{i\phi}\right)\varepsilon_{n-1} - \lambda^2 e^{i\omega}e^{i\phi}\varepsilon_{n-2} - A_{n-1}(1-z^{-1})\delta z_{n-1}^{(1)} - B_{n-1}(1-z)\delta z_{n-1}^{(2)} \\
&\quad + iA_n e_n + iB_n f_n + v_n - \left(e^{i\omega}+e^{i\phi}\right)v_{n-1} + e^{i\omega}e^{i\phi}v_{n-2} = A_n\varepsilon_n^A \\
\varepsilon_n^A &= \lambda(1+z^{-1})\varepsilon_{n-1}^A - \lambda^2 z^{-1}\varepsilon_{n-2} - (1-z^{-1})\xi_{n-1} - \chi z^{-n}(1-z)\vartheta_{n-1} \\
&\quad + ie_n + i\chi z^{-n} f_n + v_n^A - (1-z^{-1})v_{n-1}^A + z^{-1}v_{n-2}^A \\
\xi_n &= \frac{\delta z_n^{(1)}}{e^{i\omega}} = \xi_{n-1} + \mu_1\varepsilon_n q A_n^* = \xi_{n-1} + (1-\lambda)^2(1-\rho)q\varepsilon_n^A \\
\vartheta_n &= \frac{\delta z_n^{(2)}}{e^{i\phi}} = \vartheta_{n-1} + \mu_2 q^*\varepsilon_n^A A_n B_n^* = \vartheta_{n-1} + (1-\lambda)^2(1-\rho)q^*\chi^{-1}z^n\varepsilon_n^A \\
\Theta_n &= [\varepsilon_n^A, \varepsilon_{n-1}^A, \xi_n, \vartheta_n]^T \qquad \Theta_n = F_n\Theta_{n-1} + \mathcal{E}_n \qquad \mu = (1-\lambda)^2(1-\rho)
\end{aligned}
$$

$$
F_n = \begin{bmatrix}
\lambda(1+z^{-1}) & -\lambda^2 z^{-1} & -1+z^{-1} & -\chi z^{-n}(1-z) \\
1 & 0 & 0 & 0 \\
\mu q\lambda(1+z^{-1}) & -\mu q\lambda^2 z^{-1} & 1-\mu q(1-z^{-1}) & -\mu q\chi z^{-n}(1-z) \\
\mu q^*\chi^{-1}z^n\lambda(1+z^{-1}) & -\mu q^*\chi^{-1}z^n\lambda^2 z^{-1} & -\mu q^*\chi^{-1}z^n(1-z^{-1}) & 1-\mu q^*(1-z)
\end{bmatrix}
$$

Analysis of HALE

$$
\begin{aligned}
\varepsilon_n &= \frac{y_n - \theta_{n-1}^T X_n}{1+\mu X_n^H X_n} \approx (v_n - \delta\theta_{n-1}^T X_n - \theta_{n-1}^T \delta X_n)/h_{n5} \\
&= \{v_n + \delta a_{1,n-1}(A_{n-1}+B_{n-1}) + (e^{i\omega}+e^{i\phi})[-v_{n-1}+(1-g)\varepsilon_{n-1}] \\
&\quad +\delta a_{2,n-1}(A_{n-2}+B_{n-2}) - e^{i(\omega+\phi)}[-v_{n-2}+(1-2g)\varepsilon_{n-2}]\}/h_{n5} \\
h_{n5} &= 1+\mu \bar{X}_n^H \bar{X}_n = 1+\mu[|A_{n-1}+B_{n-1}|^2 + |A_{n-2}+B_{n-2}|^2] \\
&= 1+2\mu|A_n|^2[1+|\chi|^2 + \mathrm{Re}(\chi z^{-n+1}) + \mathrm{Re}(\chi z^{-n+2})] \\
\varepsilon_n^A &= \varepsilon_n/A_n = \{\frac{v_n}{A_n} + (1-g)(1+z^{-1})\varepsilon_{n-1}^A - z^{-1}(1-2g)\varepsilon_{n-2}^A \\
&\quad +(1+\chi z^{-n+1})\,\frac{\delta a_{1,n-1}}{e^{i\omega}} + (1+\chi z^{-n+2})\,\frac{\delta a_{2,n-1}}{e^{2i\omega}}\}/h_{n5} \\
\theta_n &= \theta_{n-1} + \mu X_n^* \varepsilon_n \\
\delta a_{1n} &= \delta a_{1,n-1} - \mu(A_{n-1}+B_{n-1})^*\varepsilon_n = \delta a_{1,n-1} - \mu|A_n|^2(1+\chi z^{-n+1})e^{i\omega}\varepsilon_n^A \\
\delta a_{2n} &= \delta a_{2,n-1} - \mu(A_{n-2}+B_{n-2})^*\varepsilon_n^A = \delta a_{2,n-1} - \mu|A_n|^2(1+\chi z^{-n+2})e^{2i\omega}\varepsilon_n^A
\end{aligned}
$$

Thus we can write $\Theta_n = F_n\Theta_{n-1} + \mathcal{E}_n$ for

$$
\begin{aligned}
\Theta_n &= [\varepsilon_n^A, \varepsilon_{n-1}^A, \delta a_{1n}e^{-i\omega}, \delta a_{2n}e^{-2i\omega}]^T \\
F_n^{\mathrm{HALE}} &= \frac{1}{h_{n5}}\begin{bmatrix} h_{n1} & h_{n2} & h_{n3} & h_{n4} \\ h_{n5} & 0 & 0 & 0 \\ -\mu|A_n|^2 h_{n1}h_{n3}^* & \mu|A_n|^2 h_{n2}h_{n3}^* & h_{n5}-\mu|A_n|^2|h_{n3}|^2 & -\mu|A_n|^2 h_{n4}h_{n3}^* \\ -\mu|A_n|^2 h_{n1}h_{n4}^* & \mu|A_n|^2 h_{n2}h_{n4}^* & -\mu|A_n|^2 h_{n3}h_{n4}^* & h_{n5}-\mu|A_n|^2|h_{n4}|^2 \end{bmatrix}
\end{aligned}
$$

where

$$
\begin{aligned}
h_{n1} &= (1+z^{-1})(1-g) \qquad h_{n2} = -z^{-1}(1-2g) \\
h_{n3} &= 1+\chi z^{-n+1} \qquad h_{n4} = 1+\chi z^{-n+2}.
\end{aligned}
$$

Proof of Proposition 4

The proof is based on the observation that the matrices F_n^{MFT}, F_n^{ANF} and $F_n^{\mathrm{AIS-2}}$ can be written in the form

$$F_n = P^{-1}\tilde{F}_n P$$

where the matrices $\tilde{F}_n$ may possibly depend on the phase of the quantity χ, but they are independent of the ratio of the sinusoidal magnitudes, $|\chi|$. In particular, P's for the above algorithms equal

$$
P = \begin{cases} \mathrm{diag}(1,1,1,|\chi|,|\chi|,|\chi|) & \text{for MFT} \\ \mathrm{diag}(1,1,1,1,1,|\chi|) & \text{for ANF} \\ \mathrm{diag}(1,1,1,|\chi|) & \text{for AIS-2} \end{cases}
$$

Now, it is easy to show that

$$
\begin{aligned}
\beta &= \limsup_{n\to\infty}[\sigma_{\max}(F_nF_{n-1}\ldots F_1)]^{1/n} = \limsup_{n\to\infty}[\sigma_{\max}(P^{-1}\tilde{F}_n\tilde{F}_{n-1}\ldots\tilde{F}_1 P)]^{1/n} \\
&= \limsup_{n\to\infty}[\sigma_{\max}(\tilde{F}_n\tilde{F}_{n-1}\ldots\tilde{F}_1)]^{1/n},
\end{aligned}
$$

independent of $|\chi|$, as stated.

References

[1] A. Nehorai, "A minimal parameter adaptive notch filter with constrained poles and zeros", *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol. ASSP-33, pp. 983–996, August 1985.

[2] B.S. Chen, T.Y. Yang and B. H. Lin, "Adaptive notch filter by direct frequency estimation", *Signal Processing*, Vol. 27, No. 2, pp. 161-167, May 1992.

[3] P.Tichavský and P. Händel, "Efficient tracking of multiple sinusoids with slowly varying parameters", *Proc. 1993 IEEE International Conference Acoustics, Speech and Signal Processing*, Minneapolis, MN, Pt. III, April 1993, pp. 368-371.

[4] P.Tichavský and P. Händel, "Two Algorithms for Adaptive Retrieval of Slowly Time-Varying Multiple Cisoids in Noise", *IEEE Trans. on Signal Processing*, vol. 43, no.5, pp. 1116–1127, May 1995.

[5] C.C. Ko and C.P. Li, "An adaptive IIR structure for the separation, enhancement, and tracking of multiple sinusoids", *IEEE Trans. on Signal Processing*, Vol.42, No.10, pp. 2832-2834, October 1994.

[6] M. Padmanabhan and K. Martin, "A second-order hyperstable adaptive filter for frequency estimation", *IEEE Tr. on Circuits and Systems–II: Analog and Digital Signal Processing*, Vol. 40, No. 6, pp. 398–403, June 1993.

[7] M. Padmanabhan "A hyperstable adaptive line enhancer for frequency estimation", *Proc. of the 1995 IEEE Int. Symp. on Circuits and Systems* in Seattle, WA, May 1995, pp. 1600–1603. Full version of the paper was submitted for publication.

[8] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*, Cambridge, MA, M.I.T. Press, 1983.

[9] N.I. Cho, C.H. Choi, and S.U. Lee, "Adaptive line enhancement by using an IIR lattice notch filter", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 585-589, April 1989.

[10] V.M. Popov, *Hyperstability of Control Systems.* New York: Springer-Verlag, 1973.

[11] C. Richard Johnson Jr., "A convergence proof for a hyperstable adaptive recursive filter", *IEEE Tr. Info. Theory*, vol. IT-25, pp. 745-749, Nov. 1979.

[12] P.Tichavský, "Posterior Cramér–Rao bound for adaptive harmonic retrieval", *IEEE Trans. on Signal Processing*, vol. 43, no.5, pp. 1299-1302, May 1995.

[13] H.L. van Trees, *Detection, Estimation and Modulation Theory*, Part I, John Wiley & Sons, New York, 1968.

[14] B. Rao and R. Peng, "Tracking characteristics of the constrained IIR adaptive notch filter", *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. ASSP-36, pp. 1466–1479, September 1988.

[15] P. Händel and A. Nehorai, "Tracking analysis of an adaptive notch filter with constrained poles and zeros ", *IEEE Transaction on Signal Processing*, Vol. 42, No.2, pp. 281-291, February 1994.

[16] L. Ljung and S. Gunnarsson, "Adaptation and tracking in system identification – a survey", *Automatica*, Vol. 26, No. 1, pp. 7-21, January 1990.

[17] N. Cho and S. Lee, "On the tracking properties of the adaptive lattice notch filter", *Proc. 1993 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Minneapolis, MN, Pt. III, April 1993, pp. 396-399.

[18] A. Nehorai and D. Starer, "Adaptive pole estimation", *IEEE Tr. on Acoust., Speech and Signal Processing*, Vol. ASSP-38, No.5, pp. 825–838, May 1990.

[19] P.A. Regalia, "An improved lattice-based adaptive IIR notch filter", *IEEE Trans. Signal Processing*, vol. 39, no. 9, pp. 2124–2128, September 1991.

[20] N. Cho and S. Lee, "On the tracking properties of the adaptive lattice notch filter", *Proc. 1993 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Minneapolis, MN, Pt. III, pp. 396-399, April 1993.

[21] M.V. Dragošević and S.S. Stanković, "Time-varying frequency tracker with adaptive pole contraction", *Proc. 1992 IEEE International Conference Acoustics, Speech and Signal Processing*, San Francisco, CA, Pt. IV, April 1992, pp.181-184.

[22] W. Rudin, *Functional Analysis*, McGraw-Hill, 1973.

[23] P. Tichavský and A. Nehorai, "Comparative study of four adaptive frequency trackers", Research report no. 1845, Institute of Information Theory and Automation, Prague, Czech Republic, August 1995.

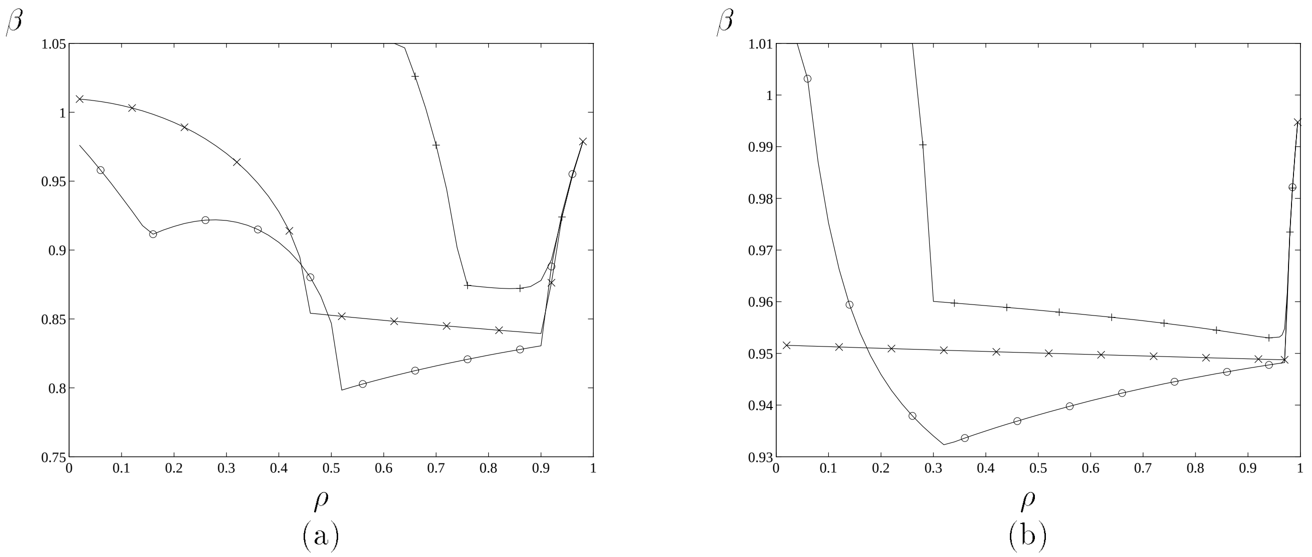


Figure 1: Local convergence factor as a function of ρ for frequency separation $\pi/3$ and $\lambda = 0.7$ in the figure (a) and $\lambda = 0.9$ in the figure (b), "x" for MFT, "o" for ANF and "+" for AIS-2.

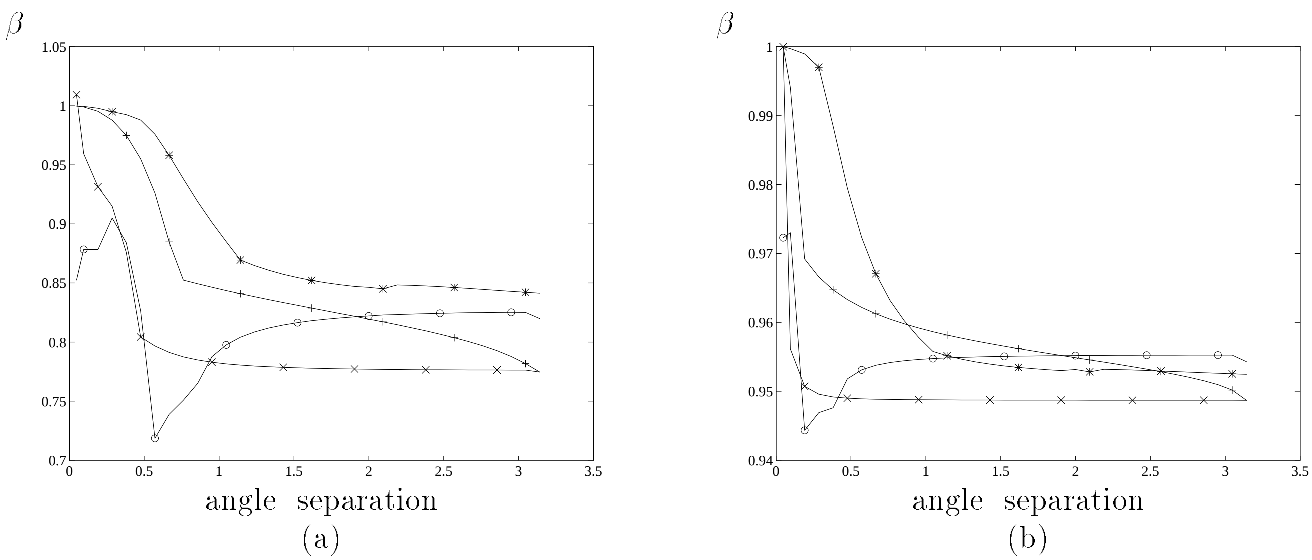
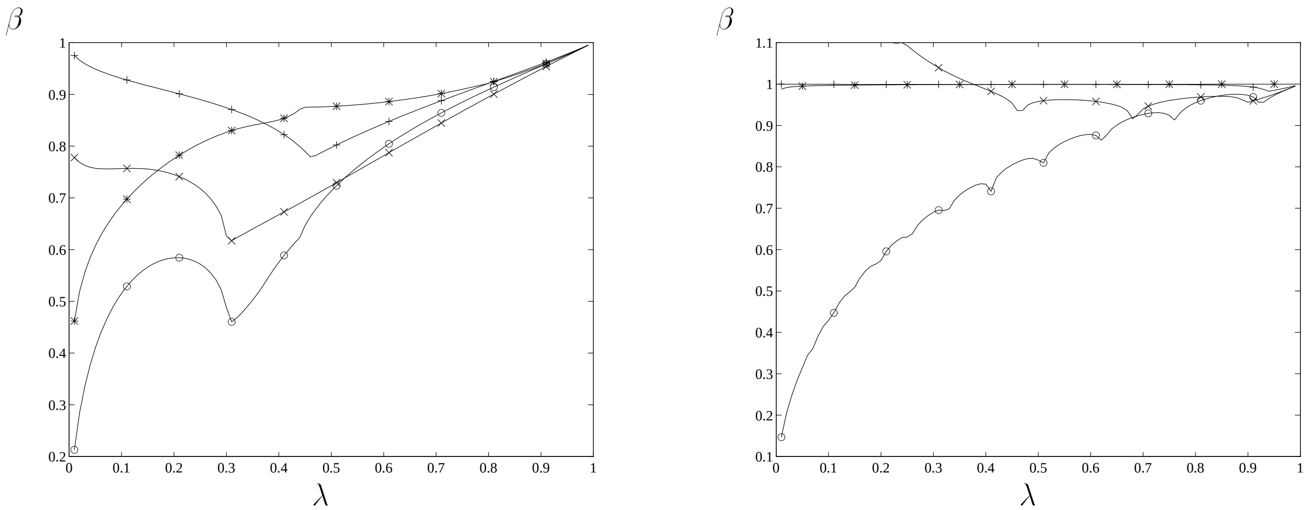


Figure 2: Local convergence factor as a function of frequency separation for $\lambda = 0.6$ in the figure (a) and $\lambda = 0.9$ in the figure (b), $\rho = \rho_0(\lambda)$, "x" for MFT, "o" for ANF, "+" for AIS-2 and "$*$" for HALE.



(a) (b)

Figure 3: Local convergence factor as a function of λ for $\rho = \rho_0(\lambda)$ and large and small frequency separation ($\pi/3$ in figure (a) and $\pi/33$ in figure (b)); "x" for MFT, "o" for ANF, "+" for AIS-2 and "$*$ " for HALE.

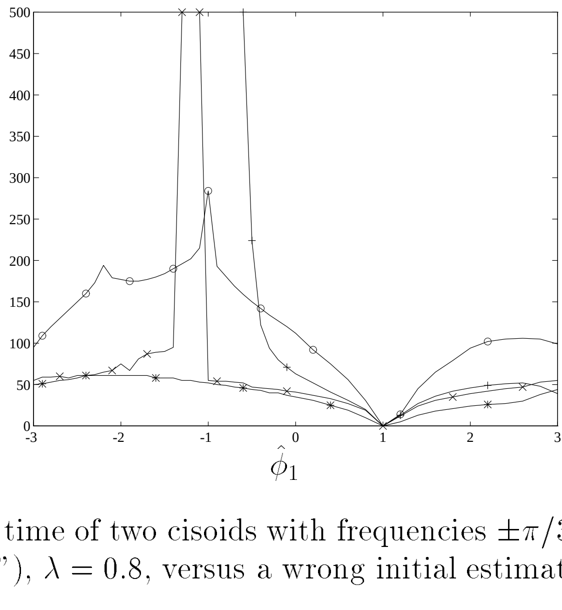


Figure 4: Empirical tracking time of two cisoids with frequencies $\pm\pi/3$, for MFT ("x"), ANF ("o"), AIS-2 ("+") and HALE ("$*$ "), $\lambda = 0.8$, versus a wrong initial estimate of the second frequency.

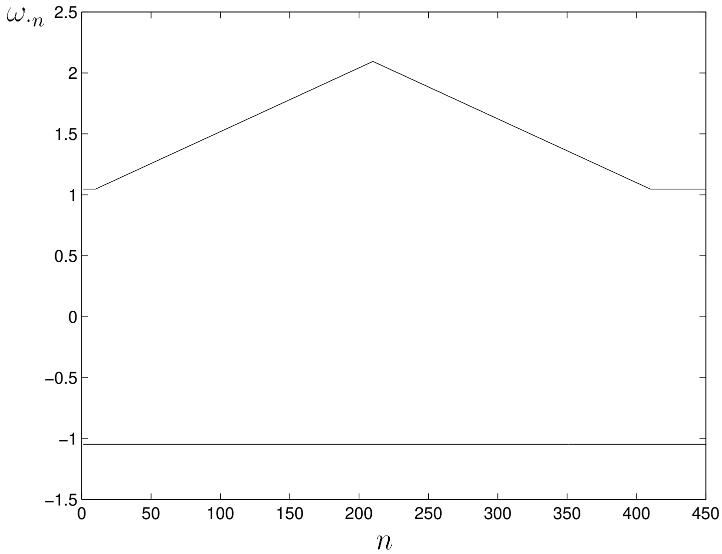


Figure 5: True sinusoidal frequencies in the model considered in subsection 5B.

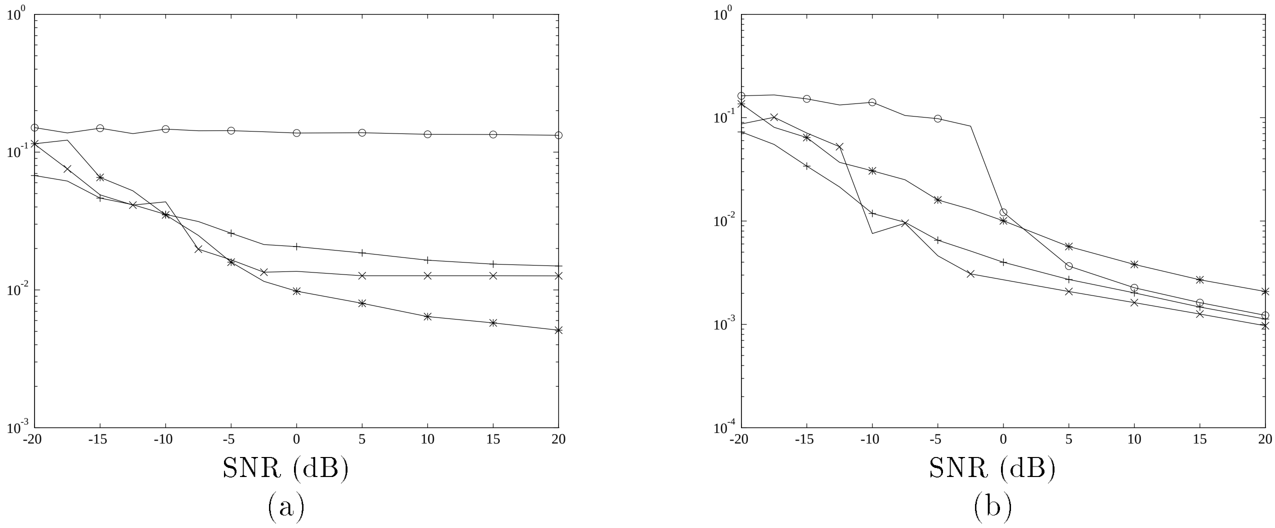


Figure 6: Empirical mean-square frequency estimation error versus SNR for MFT ("x"), ANF ("o"), AIS-2 ("+") and HALE ("$*$ "), in the example 5B. In simulation (a) λ and ρ were constant, in simulation (b) they were SNR-dependent.