ISSPA 90, Signal Processing, Theories, Implementations and Applications.
Gold Coast, Australia, 27-31 August, 1990.
Editors: B. Boashash, P. Boles.

THE AMPLITUDE, PHASE AND FREQUENCY ESTIMATION OF MULTIHARMONIC SIGNALS IN NOISE — AN INVESTIGATION OF THE GENERAL PHASE-FREQUENCY ESTIMATOR[1]

Ben James[2] and Brian D. O. Anderson[2]

[1] Research supported by Australian Telecommunications and Electronics Research Board and the ANU Centre for Information Science Research
[2] Dept. of Systems Eng., RSPhysS, Australian National University, G.P.O. Box 4, Canberra, A.C.T. 2601, AUSTRALIA

ABSTRACT: Previous work, (Parker-Anderson), has applied an Extended Kalman Filter (EKF) to the problem of estimating the fundamental frequency, phases and amplitudes of the harmonic components of a periodic, non-sinusoidal signal contaminated by additive noise. It was shown that the EKF can be decoupled into two separate entities: an amplitude estimator and a phase-frequency estimator. Furthermore, it was shown that the amplitude estimator is explicitly computable. This paper describes the subsequent investigation of the phase-frequency estimator in isolation, where it is assumed that the harmonic amplitudes are known. We derive bounds on the performance of the estimator and show how the phase-frequency estimation problem can be interpreted in terms of multicarrier Frequency Modulation (FM). The paper also presents simulations demonstrating the existence of a 'threshold effect' - familiar to us from single carrier FM demodulation - associated with the general phase-frequency estimation problem.

1 Introduction

Consider an approximately periodic, non-sinusoidal signal contaminated by additive noise. Since the signal is non-sinusoidal, we may think of it as consisting of an infinite number of sinusoidal components, or harmonics. The problem we wish to address is that of tracking the signal's fundamental frequency along with the phase and amplitude of each of its harmonic components.

Parker and Anderson [6,7] have attacked this problem by formulating a state space description of the signal and applying an Extended Kalman Filter (EKF) to the received signal. A complete description of their approach may be found in [6,7]; what follows is a brief recapitulation of the problem formulation given therein. While the work of this paper does not deal with their formulation *per se*, the following summary serves to set the scene for our treatment. We will be consistent with their notation.

We denote the received (or measured) signal by

$$z(t) = y(t) + n(t); \quad t = 0, 1, 2, \ldots \tag{1.1}$$

where $n(t)$ is the corrupting measurement noise caused by passage through a channel, and $y(t)$ the actual periodic, non-sinusoidal signal. (Here, t is a discrete time index.) The periodic signal may be expressed as a Fourier series:

$$y(t) = \sum_{k=1}^{\infty} r_k \sin(k\omega t + \phi_k), \tag{1.2}$$

where we have assumed a zero d.c. component, the fundamental frequency to be ω and the harmonic amplitudes and phases to be r_k and ϕ_k respectively.

Suppose now that $y(t)$ is in fact approximately periodic so that harmonic amplitudes and phases are 'slowly' time varying. Then

$$y(t) = \sum_{k=1}^{\infty} r_k(t) \sin \theta_k(t), \tag{1.3}$$

where

$$\theta_k(t) = \sum_{\tau=0}^{t} k\omega(\tau) + \phi_k(t)\,. \tag{1.4}$$

with $|\omega(t+1) - \omega(t)|/\omega(t)$, $|r_k(t+1) - r_k(t)|/\omega(t)$ and $|\phi_k(t+1) - \phi_k(t)|/\omega(t)$ all small. It is reasonable to assume that only a finite number of harmonics contain the bulk of the signal energy, so that all higher order harmonic contributions may be neglected in the analysis. The signal can then be represented by

$$y(t) = \sum_{k=1}^{m} r_k(t) \sin \theta_k(t)\,, \tag{1.5}$$

where m is the number of significant harmonic components. (We shall henceforth refer to the 'fundamental' component as the 'first harmonic'.)

1.1 State Space Model

It is now straightforward to define a state space model for the signal. The state vector consists of the $2m+1$ signal parameters as follows:

$$x(t) = [r_1(t) \ldots r_m(t)\ \omega(t)\ \theta_1(t) \ldots \theta_m(t)]^T\,. \tag{1.6}$$

The state space model is described as follows.

$$x(t+1) = Fx(t) + v(t) \tag{1.7}$$

$$z(t) = y(t) + n(t) = h[x(t)] + n(t) \tag{1.8}$$

where

$$h[x(t)] = \sum_{k=1}^{m} r_k(t) \sin \theta_k(t) \tag{1.9}$$

and the $(2m+1) \times (2m+1)$ matrix F is given by

$$F = \left(\begin{array}{c|ccccc} I_m & & & & & \\ \hline & 1 & & & & \\ & 1 & 1 & & & \\ & 2 & & 1 & & \\ & \vdots & & & \ddots & \\ & m & & & & 1 \end{array}\right)\,. \tag{1.10}$$

where I_m is the $m \times m$ identity matrix and the blank spaces denote zeros. The process noise term $v(t)$ is a white, Gaussian, zero-mean vector random process of length $2m+1$ with covariance $E[\,v(t)v^T(s)\,] = Q\delta_{ts}$, where Q is the $(2m+1) \times (2m+1)$ diagonal matrix: $Q \triangleq \mathrm{diag}[\,\eta_0\ \eta_1 \ldots \eta_{2m}\,]$.

The measurement noise $n(t)$ is a white, Gaussian, zero-mean scalar random process with variance $E[\,n(t)n(s)\,] = R\delta_{ts}$. We also assume that for all t and s there holds $E[\,n(t)v^T(s)\,] = 0$. The signal model is suited to the application of an EKF due to the nonlinearity of (1.9). The relevant filter equations are as follows, (see [1]).

1.2 Discrete-Time Extended Kalman Filter

Estimation Equations:

$$\hat{x}(t \mid t) = \hat{x}(t \mid t-1) + L(t)\{\,z(t) - h[\,\hat{x}(t \mid t-1)\,]\} \tag{1.11}$$

$$\hat{x}(t+1 \mid t) = F\hat{x}(t \mid t) \tag{1.12}$$

Kalman Gain:

$$L(t) = \Sigma(t)H(t)[\,H^T(t)\Sigma(t)H(t) + R\,]^{-1} \tag{1.13}$$

$$\Sigma(t) = E[\,(x(t) - \hat{x}(t \mid t-1))(x(t) - \hat{x}(t \mid t-1))^T\,] \tag{1.14}$$

Riccati Equation:

$$\Sigma(t+1) = F\{\,\Sigma(t) - \Sigma(t)H(t)[\,H(t)^T\Sigma(t)H(t) + R\,]^{-1}H^T(t)\Sigma(t)\,\}F^T + Q \tag{1.15}$$

Linearized Measurement Vector:

$$H(t) = \left.\frac{\partial h[\,x(t)\,]}{\partial x(t)}\right|_{x(t)=\widehat{x}(t|t-1)} \tag{1.16}$$

Upon application to our signal model, we obtain

$$H(t) = [\,\sin\widehat{\theta}_1 \ldots \sin\widehat{\theta}_m\ 0\ \widehat{r}_1\cos\widehat{\theta}_1 \ldots \widehat{r}_m\cos\widehat{\theta}_m\,]^T \tag{1.17}$$

where for convenience, $\widehat{\theta}_i = \widehat{\theta}_i(t \mid t-1)$ and $\widehat{r}_i = \widehat{r}_i(t \mid t-1)$. The filter is initialized by $\widehat{x}(0 \mid -1) = E[\,x(0)\,] = \overline{x}_0$ and $\Sigma(0) = E[\,(x(0)-\overline{x}_0)(x(0)-\overline{x}_0)^T\,] = P_0$. This completes the formulation of the problem as one of extended Kalman filtering.

EKF Structure: It was shown in [6] that the solution of (1.15) with H given by (1.17) could be closely approximated by the solution of a 'time-averaged' version of (1.15), (see [6] for explanation). This led to the discovery that the EKF could be decoupled into separate amplitude estimation and phase-frequency estimation components whereby amplitude estimates were made as if the phase and frequency estimates were correct and *vice versa*. Furthermore, *the amplitude estimator was found to be explicitly computable.* Our task, therefore, is to attempt an understanding of the phase-frequency estimator.

2 Analysis of Phase-Frequency Estimator

Once again we consider an approximately periodic, non-sinusoidal signal $y(t)$ with m harmonics, although on this occasion we will work in continuous time. Our task is to track the signal's fundamental frequency along with its harmonic phases, given that it has been contaminated by additive noise. We assume perfect knowledge of the harmonic amplitudes, A_k, $1 \le k \le m$, consistent with our wish to analyse the phase-frequency estimator in isolation. Similarly to Section 1, we may express the uncorrupted signal by

$$y(t) = \sum_{k=1}^{m} A_k \sin\theta_k(t)\,. \tag{2.1}$$

The measurements are then given by

$$z(t) = y(t) + n(t) \tag{2.2}$$

where $n(t)$ is the measurement noise. We assume a state space model as follows:

$$\begin{aligned} \dot{x}(t) &= Fx(t) + \underline{\omega}_0 + v(t) && (2.3)\\ z(t) &= \sum_{k=1}^{m} A_k \sin\theta_k(t) + n(t) \\ &= h[x(t)] + n(t) && (2.4) \end{aligned}$$

where $x(t) = [\,\omega(t)\ \theta_1(t)\ldots,\theta_m(t)\,]^T$, $\underline{\omega}_0 = [\,\alpha\omega_0\ 0\ldots 0\,]^T$ and

$$F = \begin{pmatrix} -\alpha & 0 & \ldots & 0 \\ 1 & 0 & \ldots & 0 \\ 2 & \vdots & & \vdots \\ \vdots & & & \\ m & 0 & \ldots & 0 \end{pmatrix} \tag{2.5}$$

Here, F is an $(m+1)\times(m+1)$ matrix and $x(t)$ a vector of length $m+1$. The process noise vector $v(t)$, (also of length $m+1$), has covariance $E[\,v(t)v^T(s)\,] = Q\delta(t-s)$, where $Q \triangleq \mathrm{diag}[\,q_0\ q_1\ldots q_m\,]$. Likewise, the measurement noise scalar $n(t)$ has variance $E[\,n(t)n(s)\,] = R\delta(t-s)$, where $R \triangleq N_0/2$.

The process and measurement noise processes are taken to be independent, just as in the discrete-time formulation. (In fact, the similarity is such that the phase-frequency part of the discrete-time state space model may be recovered by sampling the above continuous-time system and letting α tend to zero.)

Once again, the signal model is nonlinear and we apply a continuous-time EKF to the received signal. The equations are given as follows.

2.1 Continuous-Time Extended Kalman Filter Equations

Estimation Equation:

$$\dot{\widehat{x}}(t) = F\widehat{x}(t) + \underline{\omega}_0 + K(t)[\,z(t) - h(\widehat{x})\,] \tag{2.6}$$

Kalman Gain:

$$K(t) = \Sigma(t)H^TR^{-1} \tag{2.7}$$

Riccati Equation:

$$\begin{aligned} \dot{\Sigma}(t) &= F\Sigma(t) + \Sigma(t)F^T - \Sigma(t)H^TR^{-1}H\Sigma(t) + Q && (2.8)\\ \Sigma(t) &\approx E[\,(x(t)-\widehat{x}(t))(x(t)-\widehat{x}(t))^T\,] && (2.9) \end{aligned}$$

Linearized Observation Vector:

$$H(t) = \left.\frac{\partial h[\,x(t)\,]}{\partial x(t)}\right|_{x(t)=\widehat{x}(t|t-1)} \tag{2.10}$$

The linearized observation vector is given in this case by

$$H = [\,0\ A_1\cos\widehat{\theta}_1\ A_2\cos\widehat{\theta}_2 \ldots A_m\cos\widehat{\theta}_m\,] \tag{2.11}$$

The 'soul' of the EKF is the Kalman Gain, obtained by solving the matrix Riccati equation, (2.8). We next discuss how the solution of (2.8) may be approximated as $t \to \infty$.

2.2 Solution By Time-Averaging

The solution of (2.8) appears to be an intractable problem due to the presence of transcendental functions in the $H(t)$ vector. The task can be considerably simplified with the use of a similar strategy to that employed in the discrete-time situation, namely the replacement of the $H^T(t)R^{-1}H(t)$ *term by its time* average. Theorem 2.1 below then tells us under which conditions the solution of the 'time-averaged' steady-state Riccati equation so obtained can be used to approximate the solution of (2.8) as $t \to \infty$. Before stating the result, we require some definitions.

Let us focus on the $H^T(t)R^{-1}H(t)$ matrix. Apart from the first row and column components which are zero, its $(i+1)(j+1)$th component is given by $2A_iA_j\cos\widehat{\theta}_i\cos\widehat{\theta}_j/N_0$. Time averaging yields

$$\mathrm{ave}[\,2A_iA_j\cos\widehat{\theta}_i\cos\widehat{\theta}_j/N_0\,] = \begin{cases} 0 & i \ne j \\ A_i^2/N_0 & i = j \end{cases}\,. \tag{2.12}$$

We may then define the time-averaged version of $H^TR^{-1}H$ as follows.

$$\overline{H}^T\overline{R^{-1}H} \triangleq \mathrm{ave}[\,H^TR^{-1}H\,] = \mathrm{diag}\frac{1}{N_0}[\,0\ A_1^2\ldots A_m^2\,]\,. \tag{2.13}$$

Here, $\mathrm{ave}(\cdot)$ denotes the operation of averaging each matrix component over one period of its oscillation. The result is the diagonal, time-invariant matrix defined in (2.13), the non-zero diagonal components of which are the individual harmonic signal-to-noise ratios (SNR's).

The 'time-averaged' steady-state version of (2.8) is defined by

$$0 = F\overline{\Sigma} + \overline{\Sigma}F^T - \overline{\Sigma H}^TR^{-1}\overline{H\Sigma} + Q \tag{2.14}$$

The filter system matrix associated with (2.14) is given by $\overline{F} \triangleq (F - \overline{\Sigma H}^TR^{-1}\overline{H})$. We are now ready to state the theorem. (Lack of space precludes us giving a proof.)

Theorem 2.1 *Consider* $\overline{\Sigma}$ *and* $\Sigma(t)$ *as defined in (2.13) and (2.8). Suppose that the eigenvalues of* $\overline{F}$ *are much closer to* 0 *than to the point* $j\omega$ *so that the system* $\dot{\alpha}(t) = \overline{F}\alpha(t) + \beta(t)$ *is a low-pass system with respect to inputs of frequency* ω *and above. Then as* $t \to \infty$, $\overline{\Sigma} - \Sigma(t)$ *is guaranteed to be small.*

(The result implies that for $\overline{\Sigma}$ to be a good approximation to $\Sigma(t)$ as $t \to \infty$, signals of frequency ω must lie outside the 'pass band' of the filter associated with $\overline{\Sigma}$.)

Armed with the result of Theorem 2.1, we turn our attention to the solution of (2.14). Assuming both frequency and phase variation, Q is a full rank diagonal matrix. By decomposing Q into the sum of two singular $(m+1)\times(m+1)$ diagonal matrices

$$Q = Q_1 + Q_2\,. \tag{2.15}$$

where

$$Q_1 \triangleq \text{diag}[\, q_0\ 0\ 0 \ldots 0\,] \tag{2.16}$$

$$Q_2 \triangleq \text{diag}[\, 0\ q_1\ q_2 \ldots q_m\,]\,, \tag{2.17}$$

it is possible to *explicitly* solve the Riccati equations

$$0 = F\overline{\Sigma}_1 + \overline{\Sigma}_1 F^T - \overline{\Sigma}_1 \overline{H}^T R^{-1} \overline{H} \overline{\Sigma}_1 + Q_1 \tag{2.18}$$

$$0 = F\overline{\Sigma}_2 + \overline{\Sigma}_2 F^T - \overline{\Sigma}_2 \overline{H}^T R^{-1} \overline{H} \overline{\Sigma}_2 + Q_2\,, \tag{2.19}$$

for an arbitrary number of harmonics, m. The solutions $\overline{\Sigma}_1$ and $\overline{\Sigma}_2$ can then be used to derive upper and lower bounds on the full solution, (that of (2.14)). (It would seem that (2.14) is not analytically solvable for arbitrary m, hence our indirect approach.)

2.3 Interpretation of $\overline{\Sigma}_1$

The signal model of (2.3) and (2.4) with Q_1 as the process noise covariance matrix, (and $\overline{\Sigma}_1$ the error covariance matrix), corresponds to the case where the fundamental frequency *alone* experiences random variation, with no overlying random phase variation. This can best be conceptualized as frequency modulation (FM) — more specifically, multi-carrier FM. We can think of the random frequency variation as the 'message' modulating a set of harmonically related carrier frequencies. We will later see the importance of this interpretation, particularly in relation to its implications for the EKF's performance at low SNR.

2.4 The Solutions $\overline{\Sigma}_1$ and $\overline{\Sigma}_2$

These are given as follows. (Space constraints preclude us from describing the method of solution.)

$$\overline{\Sigma}_1 = \begin{pmatrix} \xi_\omega & \xi_{\theta\omega} & 2\xi_{\theta\omega} & 3\xi_{\theta\omega} & \cdots & m\xi_{\theta\omega} \\ \xi_{\theta\omega} & \xi_\theta & 2\xi_\theta & 3\xi_\theta & \cdots & m\xi_\theta \\ 2\xi_{\theta\omega} & & 4\xi_\theta & 6\xi_\theta & \cdots & 2m\xi_\theta \\ 3\xi_{\theta\omega} & & & 9\xi_\theta & \cdots & \vdots \\ \vdots & & & & \ddots & m(m-1)\xi_\theta \\ m\xi_{\theta\omega} & & & & & m^2\xi_\theta \end{pmatrix} \tag{2.20}$$

$$\overline{\Sigma}_2 = \begin{pmatrix} 0 & 0 & & \cdots & & 0 \\ 0 & \sqrt{\frac{q_1}{\Lambda_1}} & 0 & \cdots & & 0 \\ 0 & 0 & \sqrt{\frac{q_2}{\Lambda_2}} & 0 & \cdots & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & & \cdots & & & \sqrt{\frac{q_m}{\Lambda_m}} \end{pmatrix} \tag{2.21}$$

where

$$\xi_\theta = \frac{\alpha\overline{\kappa}}{\overline{\Lambda}} \tag{2.22}$$

$$\xi_{\theta\omega} = \frac{\overline{\Lambda}}{2}\xi_\theta^2 \tag{2.23}$$

$$\xi_\omega = \frac{1}{2\alpha}(q_0 - \frac{\alpha^4\overline{\kappa}^4}{4\overline{\Lambda}}) \tag{2.24}$$

$$\overline{\kappa} = -1 + (1 + \frac{2}{\alpha^2} q_0^{1/2}\overline{\Lambda}^{1/2})^{1/2} \tag{2.25}$$

$$\overline{\Lambda} = \Lambda_1 + 4\Lambda_2 + 9\Lambda_3 + \ldots + m^2\Lambda_m \quad \text{('effective' SNR)} \tag{2.26}$$

$$\Lambda_i = \frac{A_i^2}{N_0} \quad 1 \le i \le m \tag{2.27}$$

The physical interpretation of the Q_1 model as multicarrier FM is justified by inspection of the solutions to (2.18). As we will later see, this reveals that the formulae of (2.22) — (2.26) are virtually identical to those for the well known problem of classical (single carrier) FM demodulation. The role of the signal-to-noise ratio (SNR) parameter in the classical case is assumed in the multicarrier case by what we term the 'effective' SNR $\overline{\Lambda}$ — a weighted sum of the individual carrier SNR's.

By comparison, the solution of (2.19), $\overline{\Sigma}_2$, is considerably simpler in form than $\overline{\Sigma}_1$.

Error	$X = X_0$	$X = 0.1 \times X_0$	$X = 0.01 \times X_0$
$(\overline{\Sigma}_1(\gamma) + \beta X^\sharp) - \overline{\Sigma}$	21.9% $\beta = 10^{-6}$	21.8% $\beta = 10^{-7}$	21.8% $\beta = 10^{-8}$
$\overline{\Sigma} - \overline{\Sigma}_1$	21.9%	21.8%	21.8%
$\overline{\Sigma} - \overline{\Sigma}_2$	247%	249%	250%

Table 1: Percentage trace error in approximating $\overline{\Sigma}$ by bounds.

We remark that the solution $\overline{\Sigma}_1$ is *not* a stabilizing solution for (2.18), in the sense that there holds only $\text{Re}\,\lambda_i[F - \overline{\Sigma}_1\overline{H}^T R^{-1}\overline{H}] \le 0$ and not $\text{Re}\,\lambda_i[F - \overline{\Sigma}_1\overline{H}^T R^{-1}\overline{H}] < 0$. In other words, the filter defined by $\overline{\Sigma}_1$ is *not* asymptotically stable. It can be made so by the introduction of a small amount of phase process noise into the signal model used for the purposes of designing the filter, ie. by making the replacement $Q_1 \to \text{diag}[\, q_0\ \epsilon_1 \ldots \epsilon_m\,]$, where $\epsilon_1, \ldots, \epsilon_m$ are small in relation to q_0, but nonzero. With this new filter applied to a signal generated by the signal model assumed in the multicarrier case, this serves to drive the phase estimates towards the true phase values as $t \to \infty$.

2.5 Bounds on $\overline{\Sigma}$

As mentioned earlier in the section, the two decomposition solutions can be used to derive upper and lower bounds on the full solution $\overline{\Sigma}$ of (2.14). The lower bounds are trivially obtained via Nishimura's Theorem [1] to be

$$\max(\overline{\Sigma}_1, \overline{\Sigma}_2) \le \overline{\Sigma}\,. \tag{2.28}$$

An upper bound may be derived, when Q_2 is not large, by considering the solution of (2.18) but with Q_1 replaced by γQ_1 for some constant γ. Let $X \triangleq \overline{H}^T R^{-1}\overline{H} = \text{diag}[\, 0\ \Lambda_1 \ldots \Lambda_m\,]$ and define the pseudo inverse of X by $X^\sharp \triangleq [\, 0\ \Lambda_1^{-1} \ldots \Lambda_m^{-1}\,]$, where $XX^\sharp = X^\sharp X = \text{diag}[\, 0\ 1 \ldots 1\,]$. The upper bound is then given by the following result. (Space limitations prevent us from presenting the proof here.)

Theorem 2.2 *Let $\overline{\Sigma}_1(\gamma)$ be the solution of (2.18) with Q_1 replaced by γQ_1.*

1. *Choose β so that $\beta^2 X^\sharp \ge Q_2$, and assume that the entries of Q_2 are sufficiently small that $\beta \le 2\alpha$.*
2. *Choose $\gamma > 0$ so that*

$$\frac{\gamma - 1}{\gamma} \ge \frac{\beta}{2\alpha}\,. \tag{2.29}$$

Then with $\overline{\Sigma}$ the solution of (2.14) with $Q = Q_1 + Q_2$, there holds

$$\overline{\Sigma} \le \overline{\Sigma}_1(\gamma) + \beta X^\sharp\,. \tag{2.30}$$

For typical values of X and with $\alpha = 0.5$, the percentage trace errors in approximating $\overline{\Sigma}$ by the overbound of Theorem 2.2 along with the underbounds stated earlier in the section, are given in Table 1. There we have defined $X_0 = \text{diag}[\, 1.0\ 1.0\ 1.0\ 1.0\ 1.0\,]$ and $Q = Q_1 + Q_2 = \text{diag}[\, q_0\ q_1 \ldots q_4\,] = 10^{-6} \times X_0$. We chose the minimum β satisfying Theorem 2.2. We also chose γ satisfying equality in (2.29), ie. $\gamma = \frac{1}{1-\beta/2\alpha}$.

3 FM Demodulation and the Threshold Effect

In this section, we aim to describe the similarities between classical single-carrier FM demodulation and the physical situation corresponding to the signal model of (2.3) and (2.4) with Q_1 as the process noise covariance matrix. This, as we have already stated, can be best interpreted as 'multicarrier' FM demodulation. Having achieved this, we intend to demonstrate the existence of a 'threshold effect' in the multicarrier case — a phenomenon that is well known in the special case of a single carrier. We will then present the results of simulations demonstrating the existence of a threshold effect in the case of the original discrete-time EKF as defined in Section 2.

3.1 Single-Carrier FM Demodulation

Our choice among the plethora of treatments of classical FM demodulation, will be the state space formulation of Van Trees [8]. Here, the received frequency-modulated signal is assumed to be of the form

$$r_s(t) = \sqrt{(2P)}\sin(\omega_c t + d_f \int_0^t a(u)du) + n_s(t) \qquad (3.1)$$

where $n_s(t)$ is white, gaussian and zero-mean with variance $\frac{N_0}{2}\delta(t-s)$ and the modulating 'message' process $a(t)$ has the spectral density

$$S_a(\omega) = \frac{q}{\omega^2 + k^2} . \qquad (3.2)$$

In other words, $a(t)$ can be thought of as the output of a first order linear system excited by white noise with spectral density q. We then define a state vector $x_s(t) \triangleq [a(t)\ \theta(t)]^T$ where $\theta(t) \triangleq d_f \int_0^t a(u)du$. (Note that we have changed the order of the state vector components from that originally defined by Van Trees in order to be consistent with our previous notation.) We can rewrite (3.1) as

$$r_s(t) = h_s(x(t)) + n_s(t) \qquad (3.3)$$

where

$$h_s(x(t)) = \sqrt{(2P)}\sin(\omega_c t + \theta(t)) . \qquad (3.4)$$

We obtain the linear 'base-band' signal model below as follows. Apply an EKF to the *nonlinear* signal model represented by (3.5) and (3.1). Then apply Theorem 2.1 to the associated Riccati equation. This gives rise to a 'time-averaged' Riccati equation whose coefficients can be identified with the parameters of the signal model of (3.5) and (3.6) below — the 'base-band' model

$$\dot{x}_s(t) = \begin{pmatrix} -k & 0 \\ d_f & 0 \end{pmatrix} x_s(t) + v_s(t) \qquad (3.5)$$

$$\tilde{r}_s(t) = (0\ \ 1)x_s(t) + \tilde{n}_s(t) . \qquad (3.6)$$

The process noise vector $\tilde{v}(t)$ is white, gaussian, zero-mean and independent of the measurement noise $\tilde{n}(t)$ with $E[\ v_s(t)v_s^T(\tau)\ ] = \mathrm{diag}[\ q\ 0\ ]\delta(t-\tau)$ and $E[\ \tilde{n}_s(t)\tilde{n}_s(\tau)\ ] = \frac{N_0}{2P}\delta(t-\tau)$.

The steady-state mean square Kalman estimation errors can then be found by solving the 2×2 algebraic Riccati equation

$$0 = \begin{pmatrix} -k & 0 \\ d_f & 0 \end{pmatrix}\overline{P} + \overline{P}\begin{pmatrix} -k & d_f \\ 0 & 0 \end{pmatrix} - \overline{P}(0\ 1)^T\frac{2P}{N_0}(0\ 1)\overline{P} + \begin{pmatrix} q & 0 \\ 0 & 0 \end{pmatrix} \qquad (3.7)$$

where

$$\overline{P} = \begin{pmatrix} \xi_\omega & \xi_{\theta\omega} \\ \xi_{\theta\omega} & \xi_\theta \end{pmatrix} . \qquad (3.8)$$

The expressions for the mean square frequency error ξ_ω and the mean square phase error ξ_θ are tabulated in Table 2. Also tabulated are the corresponding expressions for the fundamental frequency and phase errors in the multicarrier case. We can see that the expressions are virtually identical apart from some minor differences caused by the fact that slightly different models have been used. This similarity gives us some confidence in supposing that multicarrier FM demodulation can be understood in terms of its classical analogue.

What are the implications of this clear similarity between the two scenarios? Perhaps the most important implication is one concerning the low SNR performance of the multicarrier demodulator. It is well known that in the single carrier case there exists a phenomenon (due to the problem's inherent nonlinearity) which manifests itself as a sudden worsening in demodulation performance at a sufficiently low SNR. This is aptly termed the 'threshold effect'. At a certain critical value of SNR known as the 'threshold point', the mean square frequency error increases suddenly and rapidly above what is predicted by the linear model. In the single carrier case, this point corresponds to a calculated mean square phase error of about 0.25 rad^2, (see Table 2).

There are two main causes of the phenomenon. The first is that at low SNR's, the linearization assumptions underpinning the EKF become invalid, introducing errors that are not accounted for by the linear model. (Recall that the EKF linearizes the nonlinear signal model about the current state estimate which is assumed close to the true current state value. At low SNR's this assumption fails.)

The second cause, (actually also attributable ultimately to the failure of the linearization assumption), is termed 'cycle slipping' whereby at low SNR's the phase estimate may differ from the true phase by a multiple of 2π. This causes a transient in the frequency estimate thereby increasing the frequency estimation error[1].

In view of the apparent similarity between the multicarrier and classical problems, it is reasonable to ask if a threshold effect occurs in the multicarrier case as well. Indeed, computer simulations show strong evidence for this.

3.2 Simulations

Computer simulations[2] were carried out using the discrete-time EKF defined in Section 2. (Recall our earlier statements about the equivalence between the continuous and discrete-time formulations. In any case, it is the discrete-time EKF that is ultimately of interest.)

Three classes of Monte-Carlo simulations were performed corresponding to the special cases of multicarrier FM demodulation, phase-frequency tracking and phase-frequency-amplitude tracking respectively. Each simulation involved measuring the mean square state estimation errors for SNR[3]'s ranging from 60 dB to −10 dB. The number of Monte-Carlo runs for each SNR was 30 . Each run was 200 time samples long to ensure that all measurements were taken in the steady-state. In the interests of continuity, we have used exactly the same input signal as in [6,7] and previously in [5], namely one with 5 harmonics whose initial amplitudes are given by $r_k(0) = r_1(0)/k$, $k = 2,3,4,5$ and fundamental frequency $\omega = 2\pi 0.08$.

As previously stated, the discrete time EKF of Section 2 was used for *each* class of simulation. The EKF estimates phases, frequency *and* amplitudes, but in the case of both multicarrier FM demodulation and phase-frequency tracking, the amplitudes are assumed constant and known. Strictly speaking therefore, the amplitudes do not need to be estimated. If we set the initial amplitude error variances to zero and the initial amplitude estimates to the true amplitude values, the EKF will estimate the frequency and phases on the basis that the amplitudes are known (and this happens in the first two classes of simulation below). This is because of the fact that the amplitude estimator and the phase-frequency estimator are decoupled, which firstly implies that the phase-frequency estimator treats the amplitude estimates as if they are correct, and secondly that uncertainty in the phase and frequency estimates has no effect on the amplitude 'estimates'.

The following initializations and parameter values were chosen for each class of simulation.

1. Multicarrier FM Demodulator: Here, the amplitudes are assumed constant and known so that we set $\hat{r}_k(0 \mid -1) = r_k(0)$, $1 \leq k \leq m$. The initial phases were also set *equal* , ie. $\hat{\theta}_k(0 \mid -1) = \theta(0) = 0$, $1 \leq k \leq m$. Similarly to [6,7], the initial frequency estimate was taken to be $\hat{\omega}(0 \mid -1) = 5/8\omega(0) = 0.05 \times 2\pi$ and the measurement noise variance to be $R = 1$. The initial state error covariance was taken to be $\Sigma(0) = \mathrm{diag}[\ 0\ 0\ 0\ 0\ 0\ (0.06\pi)^2\ 0.02\ 0.02\ 0.02\ 0.02\ 0.02\ ]$. Note that the first 5 diagonal entries are zero, consistent with our (assumed) perfect knowledge of the harmonic amplitudes. Also note that the last five entries are non-zero which simply reflects what would in practice be our poor initial knowledge of the phase — (we have used the same values as in [6,7] here). In this and the following classes of simulation, we use *two* values of the process noise covariance matrix. The first is the actual signal model process covariance used in the simulation program to generate the signal to be estimated. In this case the signal has no amplitude or phase variation. The covariance is then given by $Q_s = \mathrm{diag}[\ 0\ 0\ 0\ 0\ 0\ a_2\ 0\ 0\ 0\ 0\ 0\ ]$ where $a_2 = 3 \times 10^{-7}$. The *other covariance*, $\tilde{Q}_s = \mathit{diag}[\ 0\ 0\ 0\ 0\ 0\ a_2\ a_1\ a_1\ a_1\ a_1\ a_1\ ]$ where $a_1 = 1 \times 10^{-4}$, is used in the on-line solution of the Riccati difference equation. We note that the phase entries of $\tilde{Q}_s$ are nonzero which appears inconsistent with the multicarrier FM demodulation assumption that there be no random phase. This choice is deliberately made to avoid the phenomenon of data saturation as described in [6]. We also note that the amplitude entries are zero because we *know* the amplitudes: they do not need to be estimated.

[1] This is sometimes referred to in practical FM systems as 'click' noise.

[2] MatLab was used.

[3] We use the same definition of SNR as do Parker and Anderson, ie. $\mathrm{SNR(dB)} = 10\log_{10}(\sum_{k=1}^{m} r_k^2/2)$.

	Mean Square Frequency error (fundamental)	Mean Square Phase Error (fundamental)	Threshold point (phase error)
Single Carrier FM Demodulation (Van Trees)	$\xi_\omega = \frac{1}{2\delta}(q - \frac{k^4\kappa^4}{4d_f^2\Lambda})$ $\kappa = -1 + (1 + \frac{2}{k^2}d_f q^{1/2}\Lambda^{1/2})^{1/2}$ $\Lambda = \frac{2P}{N_0}$	$\xi_\theta = \frac{k\kappa}{\Lambda}$	$\approx 0.25\text{rad}^2$
Multicarrier FM Demodulation (Q_1)	$\xi_\omega = \frac{1}{2\delta}(q_0 - \frac{\alpha^4\overline{\kappa}^4}{4\overline{\Lambda}})$ $\overline{\kappa} = -1 + (1 + \frac{2}{\alpha^2}q_0^{1/2}\overline{\Lambda}^{1/2})^{1/2}$ $\overline{\Lambda} = \frac{A_1^2}{N_0} + \frac{4A_3^2}{N_0} + \ldots + \frac{m^2A_m^2}{N_0}$	$\xi_\theta = \frac{\alpha\overline{\kappa}}{\overline{\Lambda}}$	$\approx 0.25\text{rad}^2$
Phase-Frequency Tracker (Q)	?	?	?

Table 2: Comparison between Multicarrier and Classical FM

2. Phase-Frequency Tracker: This situation corresponds to the signal model of (4.8) and (4.9) with Q as the process noise covariance matrix. In other words, we are permitting the frequency *and* the phase to have random variation. However, the harmonic amplitudes are still assumed constant and known. Our signal model process covariance is therefore taken to be $Q_s = \text{diag}[\,0\;0\;0\;0\;0\;a_2\;a_3\;a_3\;a_3\;a_3\;a_3\,]$ where $a_3 = 1 \times 10^{-3}$, while the covariance used in the solution of the Riccati difference equation is $\widetilde{Q}_s = \text{diag}[\,0\;0\;0\;0\;0\;a_2\;a_3\;a_3\;a_3\;a_3\;a_3\,]$. Otherwise, all initializations and parameter values are the same as for the multicarrier FM demodulator.

3. Phase-Frequency-Amplitude Tracker: We now permit frequency, phases and amplitudes to vary randomly so that the signal model process covariance becomes $Q_s = \text{diag}[\,a_3 \ldots a_3\;\; a_2\;\; a_3 \ldots a_3\,]$. (This time we are estimating amplitudes as well as frequency and phases.) The possibility of data saturation is no longer of concern, since all the states in the signal model are excited. In other words, we set $\widetilde{Q}_s = Q_s$. Apart from initializing the fundamental amplitude estimate to the square root of 3/4 of the peak signal energy and the remaining harmonic amplitude estimates to zero, all initializations and parameter values are the same as in the multicarrier FM demodulator.

Results: Plots of the inverse steady-state frequency error variance versus SNR, both actual (asterisks) and theoretical (unbroken line), for each of the three classes are given in Figures 1 - 3. (It should be noted that *filtered*, not *predicted* estimates are used.) The existence of a threshold effect is clear. At an SNR of approximately 12 dB the inverse of the actual frequency error variance drops suddenly below the corresponding (linear) theoretical value. We obtained the theoretical value of error variance for each SNR by averaging the on-line solution of the Riccati difference equation after 200 time samples over the 30 Monte Carlo runs. (An averaging approach was necessitated by the measurement dependent nature of the H vector in the Riccati equation.)

We also present plots of the steady-state phase error variance versus SNR, both actual and theoretical, for the first harmonic (fundamental) in Figures 4 - 6. The onset of threshold is clearly exhibited. (The presence of cycle slipping is not deducible from these plots, as we have used mod π values of the phase error in calculating the variance.) Of interest is the difference between the theoretical and actual phase error variance in the linear region, (high SNR), a difference not observed in the frequency plots. A possible broad explanation may be the fact that application of an EKF involves a linearization about the *predicted* state estimate, thereby introducing extra uncertainty into the state estimates above that associated with a linearization about the *filtered* state estimate. How this might explain the difference between the phase error variance plots and the inverse frequency error variance plots is not known. The phenomenon is still under investigation.

4 Conclusions

This paper has continued an investigation into the properties of a frequency estimation algorithm first mooted in [6]. Some new insights into the structure of the algorithm were discovered, primarily that the phase-frequency estimation problem is interpretable in terms of multicarrier FM demodulation. The multicarrier FM problem was found to be very similar to the well known problem of classical single carrier FM. This recognition led to the discovery of a threshold effect in the case of the full amplitude, phase and frequency tracking problem.

This in turn has spawned a variety of directions for further investigation. First among these is gaining an understanding of how the threshold effect occurs. The question here is: are there a number of *separate* threshold points each associated in some way with individual harmonics? Can this be established analytically?

Another is characterization of the threshold point. Is this possible with a single value of phase error for example, as in the case of single carrier FM, or do harmonics introduce unexpected complications?

Finally turning from the analytical to the practical: how can we lower the threshold point, ie. how can we get better performance from the EKF at low SNR's?

References

[1] B.D.O. Anderson, and J.B. Moore, *Optimal Filtering*, Prentice Hall, Englewood Cliffs, NJ, 1979.

[2] A. Gelb, *Applied Optimal Estimation*, MIT Press, Cambridge Ma, 1974.

[3] A.H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, Inc., New York, N.Y., 1970.

[4] F.L. Lewis, *Optimal Estimation with an introduction to stochastic control theory*, John Wiley, 1986.

[5] A.Nehorai and B. Porat, "Adaptive Comb Filtering for Harmonic Signal Enhancement", *IEEE Trans. Acoustics, Speech and Signal Processing*, Vol. SSP-34, pp. 1124-1138, Oct. 1986.

[6] P.P. Parker and B.D.O. Anderson, "Frequency Tracking of Periodic Signals in Noise", *Signal Processing*, to appear.

[7] P.P. Parker and B.D.O. Anderson, "Frequency Tracking of Periodic Signals in Noise", *Proc. ASSPA-89*, Adelaide, pp. 263-267, April 1989.

[8] H.L. Van Trees, *Detection, Estimation and Modulation Theory*, Vol. I and II, John Wiley, 1971.

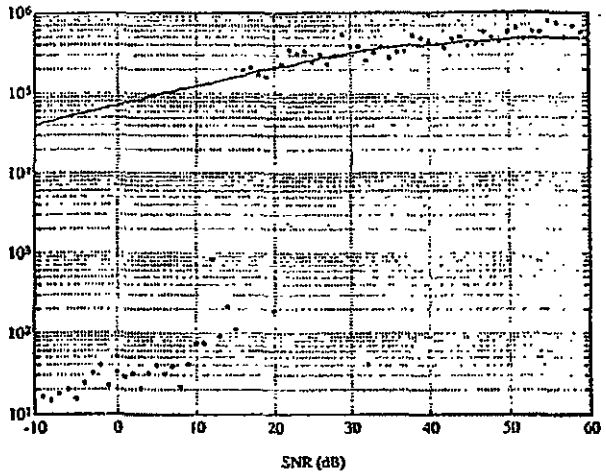

Figure 1: Inverse frequency error variance vs SNR - Multicarrier FM

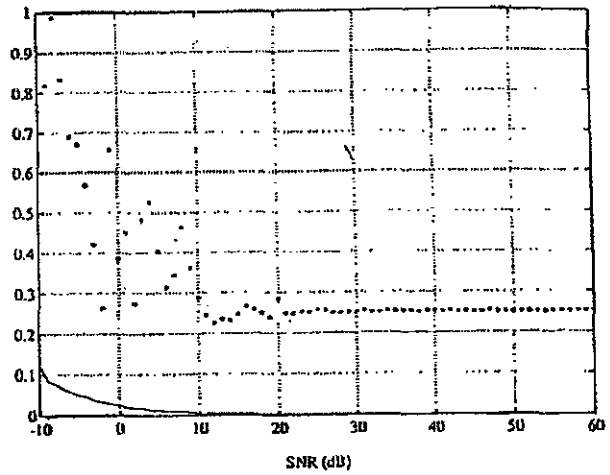

Figure 4: Phase error variance vs SNR - Multicarrier FM

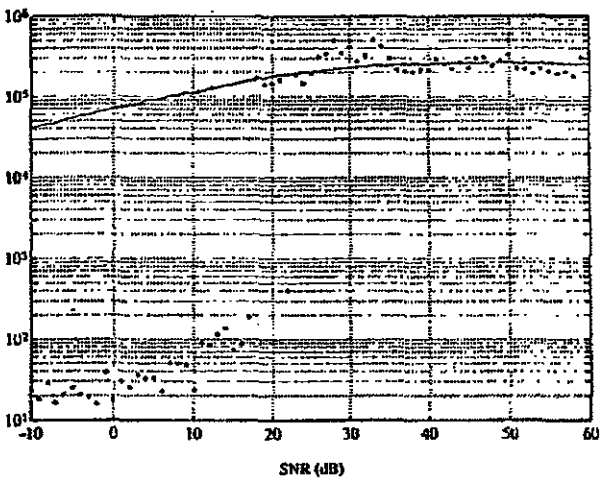

Figure 2: Inverse frequency error variance vs SNR - Phase-Frequency Tracker

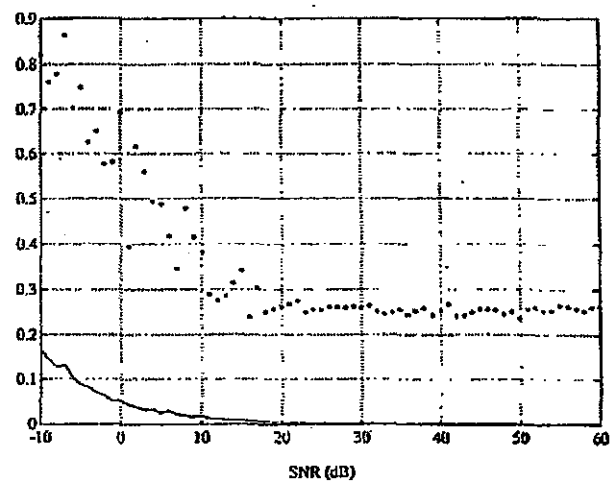

Figure 5: Phase error variance vs SNR - Phase-Frequency Tracker

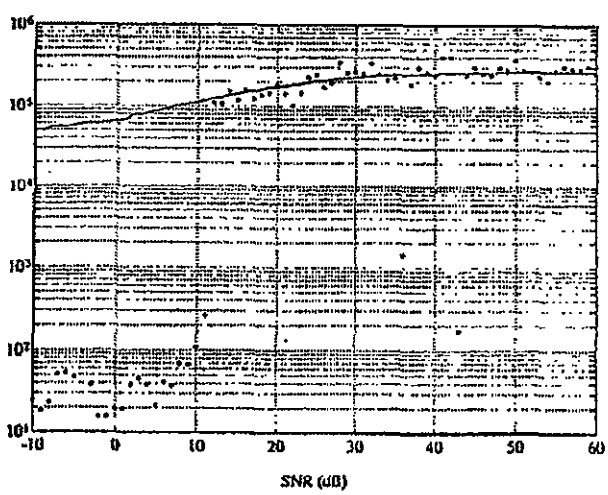

Figure 3: Inverse frequency error variance vs SNR - Phase-Frequency-Amplitude Tracker

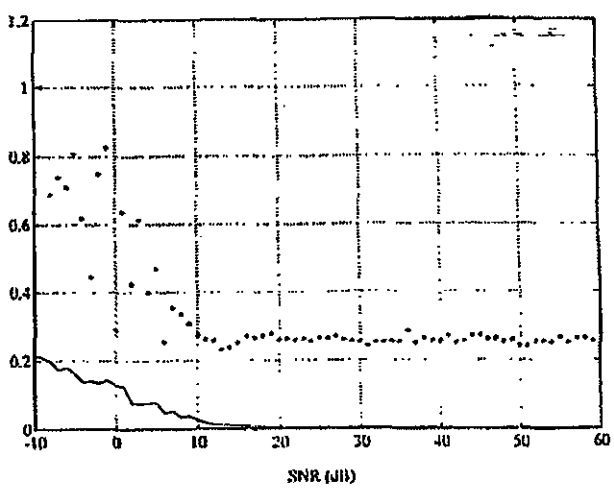

Figure 6: Phase error variance vs SNR - Phase-Frequency-Amplitude Tracker