Signal Processing 20 (1990) 127-152
Elsevier

FREQUENCY TRACKING OF NONSINUSOIDAL PERIODIC SIGNALS IN NOISE

Philip J. PARKER†

Department of Electrical Engineering II, Kyoto University, Honmachi Yosida, Sakyo-Ku, Kyoto 606, Japan

Brian D.O. ANDERSON

Department of Systems Engineering, Research School of Physical Sciences, Australian National University, Canberra, ACT 2601, Australia

Received 1 May 1989
Revised 1 December 1989

Abstract. For a periodic signal measured in noise, this paper applies extended Kalman filtering to the problem of estimating the signal's frequency and the amplitudes and phases of the signal's first m harmonic components. The resultant estimator will also track the signal's frequency and its amplitudes and phases should these change over time. In this respect, it is unique among approaches to this problem. A partial theoretical analysis of the estimator appears in the paper. This analysis shows that there is some measure of decoupling in the estimator: the amplitudes are estimated as if the phase and frequency estimates are correct; the phases and frequency are estimated as if the amplitude estimates are correct. For the special case that the signal is a sinusoid and has known amplitude, the estimator becomes the well-known phase-locked loop. The paper also contains extensive simulations demonstrating both the tracking and the asymptotic behaviour of the estimator. The asymptotic behaviour is compared with results for another known estimator, and the relative strengths of each method are examined.

Zusammenfassung. Im folgenden wird eine erweiterte Kalman-Filterung angewandt auf das Problem, für ein periodisches, in Rauschen eingebettetes Signal die Frequenz sowie die Amplituden und Phasen der ersten m harmonischen Komponenten zu schätzen. Der sich ergebende Schätzer ist auch fähig, die Signalfrequenz sowie die Amplituden und Phasen zu verfolgen, wenn sie sich mit der Zeit ändern. In dieser Hinsicht handelt es sich um den bislang einzigen Ansatz zur Lösung dieses Problems. Der Schätzer wird zum Teil theoretisch analysiert. Diese Analyse zeigt, daß es im Schätzer ein Maß der Entkopplung gibt: die Amplituden werden geschätzt, als ob die Schätzwerte der Frequenz und der Phasen korrekt wären, die Phasen und die Frequenz werden bestimmt, als ob die geschätzten Amplituden fehlerfrei vorlägen. Für den Sonderfall, daß es sich um ein Sinus-signal bekannter Amplitude handelt, wird der Schätzer zur wohlbekannten Phase-Locked-Loop. Enthalten sind weiterhin umfangreiche Simulationen, welche sowohl die Trackingfähigkeit als auch das asymptotische Verhalten des Schätzers zeigen. Letzteres wird verglichen mit den Ergebnissen für einen anderen bekannten Schätzer, und die jeweiligen Vorzüge beider Verfahren werden untersucht.

Résumé. Cette étude concerne l'application du filtrage de Kalman étendu (extended Kalman filter) à un signal périodique bruité pour estimer les fréquences du signal ainsi que les amplitudes et les phases des m premiers composantes harmoniques de celui-ci. L'estimateur résultant pourra également suivre la fréquence du signal et ses amplitudes et phases au cas où elles changeraient dans le temps. Cette analyse montre que dans une certaine mesure l'estimateur présente un découplage: on estime les amplitudes comme si les fréquences estimées étaient correctes; ensuite on estime les phases et les fréquences comme si les amplitudes estimées étaient correctes. Dans le cas spécial d'un signal sinusoidal dont on connaît l'amplitude, l'estimateur prend la forme bien étudiée de la boucle à verrouillage de phase (phase-locked loop). L'étude et des simulations approfondies montrent à la fois la capacité de poursuite et le comportement asymptotique de l'estimateur. On fait la comparaison entre ce dernier et les résultats correspondants à un estimateur connu et on examine les avantages relatifs de chaque méthode.

Keywords. Frequency estimation, extended Kalman filter, phase-locked loop, harmonic signal analysis, periodic signal analysis.

† Formerly at Department of Systems Engineering, Research School of Physical Sciences, Australian National University, Canberra, ACT 2601, Australia.

1. Introduction

Consider a periodic signal which is not, in general, sinusoidal. Three sets of parameters characterize the signal: the frequency of the fundamental component, the amplitude of each harmonic component and the phase of each harmonic component.

Suppose now that the signal is not exactly periodic but rather has frequency, amplitudes and phases that change slowly over time. Here, 'slowly' means slowly compared to the period of the signal, so that over a few cycles the signal appears periodic. Suppose also that the signal is measured in the presence of additive noise. From such measurements we wish to estimate and track the time-varying frequency and the time-varying amplitudes and phases of the harmonic components.

The problem has applications in many fields. For example, in sonar systems, hydrophones detect the engine noise of a vessel. Having been generated by a rotating engine, the sound is approximately periodic; but if the engine speed changes, so does the frequency. The amplitudes and phases may also change over time. Knowing the frequency, amplitudes and phases may be sufficient for identifying the class of vessel generating the sound, hence the importance of the problem. Sonar is one of a more general class of problems where rotating machinery generates the signal frequency. From the sound detected, the frequency tells the speed of rotation; the relative strength of amplitudes shows the amount of vibration in the machine.

We now define some notation. Consider first a periodic signal $y(t)$ with zero d.c. component. A Fourier series representation is

$$y(t)=\sum_{k=1}^{\infty} r_k \sin(k\omega t+\phi_k). \tag{1.1}$$

Here, and throughout this paper, we will work in discrete time: $t=0,1,2,\ldots$ There is no real loss of generality in doing this. Now suppose that $y(t)$ is not exactly periodic, but has slowly time-varying frequency ω, amplitudes r_k and phases ϕ_k:

$$\omega=\omega(t), \tag{1.2}$$

$$r_k=r_k(t), \tag{1.3}$$

$$\phi_k=\phi_k(t), \tag{1.4}$$

where $\omega(t)$, $r_k(t)$ and $\phi_k(t)$ are nearly constant over several cycles of $y(t)$. Analytically, $|\omega(t+1)-\omega(t)|/\omega(t)$, $|r_k(t+1)-r_k(t)|/\omega(t)$ and $|\phi_k(t+1)-\phi_k(t)|/\omega(t)$ are all small. Now instead of $k\omega t+\phi_k$ that appears in (1.1) we have

$$\theta_k(t)=\sum_{\tau=0}^{t} k\omega(\tau)+\phi_k(t), \tag{1.5}$$

and thus

$$y(t)=\sum_{k=1}^{\infty} r_k(t)\sin\theta_k(t). \tag{1.6}$$

In this model of $y(t)$, the quantities $\omega(t)$, $r_k(t)$ and $\theta_k(t)$, are the instantaneous frequency, amplitudes and phases, where $\theta_k(t)$ includes the phase due to the integration or summing of the frequency together with the additional phase shift $\phi_k(t)$. Henceforth in the paper, we shall use the term total phase to denote $\theta_k(t)$ and relative phase to denote $\phi_k(t)$. It is the relative phase which is slowly varying. Later in this paper we will give specific models for the time variation of $\omega(t)$, $r_k(t)$ and $\theta_k(t)$.

An alternative, equivalent parametrization for $y(t)$ replaces amplitudes and phases with sine and cosine components. That is,

$$y(t)=a_1(t)\sin\theta(t)+\sum_{k=2}^{\infty}[a_k(t)\sin k\theta(t)+b_k(t)\cos k\theta(t)], \tag{1.7}$$

where

$$\theta(t)=\sum_{\tau=0}^{t}\omega(\tau)+\theta(-1). \tag{1.8}$$

Here the parameters are $a_1(t)$, $a_2(t),\ldots,$ $b_2(t),\ldots,$ $\theta(t)$ and $\omega(t)$. Note that there is no cosinusoidal component at the fundamental frequency. This is because such a cosinusoidal component would represent a phase shift in the fundamental, which can equivalently be represented by a change in $\theta(t)$. Thus to include a

$b_1(t)\cos\theta(t)$ component would be an overparametrization. Therefore, we keep to the parametrization of (1.7).

We will call the first model, that of (1.5) and (1.6), the polar model and the second model, that of (1.7) and (1.8), the rectangular model. Most of the analysis and all of the simulations in this paper will be for the polar model; however, throughout the paper, we will point to places where the two models give similar results and to places where they give differing results.

Regardless of which model we work with, we assume that the signal $y(t)$ is corrupted by noise. Denote the noise by $n(t)$ and let it be white and gaussian. Then we assume we have measurements

$$z(t)=y(t)+n(t) \tag{1.9}$$

available. The task, restated, is to take measurements $\{z(\tau)|\tau=0,1,\ldots,t\}$ and from these measurements, estimate $r_1(t),\ldots,r_m(t)$, $\theta_1(t),\ldots,\theta_m(t)$, $\omega(t)$ for the polar model, or $a_1(t),\ldots,a_m(t)$, $b_2(t),\ldots,b_m(t)$, $\theta(t)$, $\omega(t)$ for the rectangular model. Furthermore, we wish to update these estimates as each new measurement $z(t)$ becomes available, thus tracking changes in the parameters. Note carefully that we only estimate parameters up to the mth harmonic: the higher harmonics are assumed to be negligible, perhaps having been eliminated by an anti-aliasing filter. In short, we approach the problem as a $2m+1$ parameter estimation and tracking problem.

Let us note some qualitative issues surrounding the problem. In estimating the signal's frequency, not only the fundamental, but also the higher harmonics contain information. Indeed, in the extreme case that no fundamental exists, with only the second, third and higher harmonics present, the fundamental obviously contains no information about the frequency $\omega(t)$. Thus, we seek an estimator that uses not only the energy in the fundamental, but also the energy in the higher harmonics to estimate the signal's frequency.

Moreover, the information about the frequency contained in any harmonic will depend upon the energy in that harmonic, or more precisely, the signal-to-noise ratio of that harmonic. Because if the second harmonic is weak, but the third is strong, then an estimator of the frequency should give more weight to the information available in the third harmonic and less weight to the information available in the second harmonic. But we do not assume a priori knowledge of the different harmonics' amplitudes; indeed, estimating these harmonic amplitudes is part of our problem. Given that a knowledge of the harmonic amplitudes helps in finding the frequency, we would expect that the estimates of the harmonic amplitudes would also help in calculating the frequency. In actual fact, the estimator we propound here estimates the frequency $\omega(t)$ (and the phases) using estimates of the harmonic amplitudes as though they are the true, correct amplitudes. This is an intuitively appealing result.

The converse argument also applies. That is, a knowledge of the frequency and phases in the polar model helps in calculating the amplitudes—or, for the rectangular model, the sine and cosine amplitudes. Indeed, if the frequency is known exactly, then estimating the sine and cosine components in the rectangular model is a linear estimation problem. Since we do not assume knowledge of the frequency, the best we can use is an estimate of the frequency. It is a hallmark of our estimator that it uses the frequency estimate as though it is the true frequency when estimating the sine and cosine components or the amplitudes.

The preceding argument has shown two things: first, in estimating the frequency it is also advantageous to estimate the amplitudes (or sine and cosine components); second, the estimation of frequency (and phases) on the one hand, and the estimation of amplitudes on the other hand, are inter-related. The estimator of this paper combines these two parts of the problem in a reasonable, intuitive way, something that other estimators do not always accomplish.

A short survey of other frequency estimators is now necessary. Frequency estimation is a mature problem with many contributions from many authors. Naturally we cannot include an exhaus-

tive survey. To highlight the salient aspects of other schemes, we divide them into four classes, depending upon the assumed characteristics of the incoming signal. These classes are (1) constant frequency, sinusoidal signal (i.e., no higher harmonics); (2) constant frequency, higher harmonics present; (3) time-varying frequency, sinusoidal signal; (4) time-varying frequency, higher harmonics present. The problem we address here falls into category (4), and is, to our knowledge, the only estimator yet to cover this category. The other categories have a number of candidate estimators as we now show.

(1) Constant frequency, sinusoidal signal

The most common estimator in this category is the discrete Fourier transform (DFT). While its simplicity and computational ease are well acknowledged, it suffers from several drawbacks when applied to our problem. The signal must have constant frequency over the length of the DFT, which may be many periods of the signal—hence its designation here as a 'constant frequency' method. The frequency resolution is poor; improving the frequency resolution means lengthening the data sequence of the DFT and thus increasing the computational complexity. It also performs unnecessary calculations: the signal may only be present in one frequency bin, but it calculates more than one frequency bin's result. It does not combine the higher harmonics and the fundamental to give an improved frequency estimate, hence the designation as being for sinusoidal signals. Lastly, it is not recursive: the previous frequency estimate is not used in finding the present estimate; calculation begins from scratch each time.

By introducing another level of algorithm that takes the output from the DFT and passes it through a hidden Markov model estimator, Streit and Barrett [18] have overcome some of these difficulties. Their prime achievement is to arrange that the hidden Markov model estimator is recursive, making theirs an estimator which tracks changes in the signal's frequency, although the frequency must still be constant over the length of the front end DFT processor. Their algorithm also works at impressively low signal-to-noise ratio levels.

Amongst the many other frequency estimators for constant frequency sinusoidal signals, see [3, 6, 10–13, 15, 16, 21], we consider only one here, namely adaptive line enhancement (ALE) which is discussed in most of these references. Here the received signal passes through a notch filter, parametrized by the frequency of the 'notch'. The estimation procedure places the notch frequency so that the output of the filter has minimum energy. This should be achieved by placing the notch at the signal's frequency, thus nulling the signal as it passes through the filter.

Like the DFT, the ALE method is for constant frequency signals. That is, over the data length, the frequency should not change, or at least it should change by no more than the width of the notch, otherwise poor filtering will result. No doubt, including a forgetting factor in the adaptive algorithm would allow tracking of slow changes in the frequency.

In its original form, the ALE method only copes with sinusoidal signals. Higher harmonics would not be nulled by the notch filter. However, as we shall mention next, the method has recently been extended to cope with higher harmonics.

(2) Constant frequency, higher harmonics present

In [14] Nehorai and Porat extended the ALE method to cope with higher harmonics. Their extension involved two advances. The first was to replace the notch filter by a comb filter, with m notches placed at the estimated frequency $\hat{\omega}$ and at $2\hat{\omega}$, $3\hat{\omega}, \ldots, m\hat{\omega}$, respectively. By correctly estimating the fundamental frequency, the comb filter would null the first m harmonics of the signal. The second advance was to take the frequency estimate $\hat{\omega}$, and use this to estimate the sinusoidal and cosinusoidal amplitudes of each harmonic. This second step is a linear estimation problem

provided one has estimated the frequency correctly. We will denote this method adaptive comb filtering (ACF).

This ACF method exhibits a desirable property. It uses the information in all of the first m harmonics in finding the frequency. That is, if the frequency estimate is slightly in error, not only the fundamental, but also the higher harmonics leak through the notch filter, encouraging a correction in the frequency estimate. However, the method does not take into account the relative strengths of the different harmonics in determining the frequency. For example, suppose the third harmonic is not present; then it would not be necessary to place a notch at $3\hat{\omega}$. Indeed, placing a notch there makes the filter sub-optimal in that it places as much weight upon filtering out the noise at $3\hat{\omega}$ as it does in filtering out the signal at $2\hat{\omega}$. In other words, the depth of the notch at each multiple of $\hat{\omega}$ should depend upon the strength of the signal at that harmonic, but in Nehorai and Porat's method, the notches are of equal depth.

The second part of Nehorai and Porat's algorithm, estimating the amplitudes of sinusoidal and cosinusoidal components, is done by treating the frequency estimate as the true frequency, and then finding the amplitudes by a least squares algorithm, an intuitively reasonable approach. Indeed, it works reasonably well for constant frequency, but separating the amplitude and frequency estimation runs into difficulties when the frequency is time varying. Should the frequency vary, then the amplitude estimation algorithm must be run anew with each new estimate of the frequency, a cumbersome process. Therefore, the ACF method is not well suited to tracking a time-varying frequency.

However, the ACF method is significant in that it looks at the frequency estimation problem when higher harmonics are present. Furthermore, [14] also gives some theoretical analysis, deriving Cramer-Rao bounds for the variance of estimates of frequency, amplitudes and phases. The results show that using higher harmonics results in significant improvements in the Cramer-Rao bounds for the frequency and phase estimates, justifying the use of such higher harmonic information.

The same Cramer-Rao bounds were derived independently by Barrett and McMahon, see [2]. This latter paper is similar to [14] in that it gives a method for estimating a fundamental frequency using the information contained in the first m harmonics. Similarly to [14], Barrett and McMahon's method does not readily adapt to tracking time-varying frequencies and amplitudes; unlike Nehorai and Porat's method, Barrett and McMahon weight the different harmonics according to their amplitude estimates when calculating the frequency estimates. In this respect, Barrett and McMahon improve upon Nehorai and Porat.

(3) Time-varying frequency, sinusoidal signal

To this point, we have considered estimators for constant frequency signals; yet there does exist an estimator for a time-varying frequency well known both through its theory and its application: it is the phase locked loop (PLL). Users of the PLL always assume that the signal is sinusoidal (or at least of known shape, such as a square wave). They also implicitly assume that the signal has known amplitude, since the signal's amplitude usually affects the PLL's dynamics, and so the PLL is designed for a certain signal level, and this signal level is usually ensured in practice by automatic gain control.

Now the PLL is known to be an approximation of the extended Kalman filter (EKF), applied to the frequency estimation problem. See [1] for a general treatment of the EKF. Snyder [17] gives the equivalence between the PLL and EKF in continuous time, and Kelly and Gupta in [7] give the discrete time case. But the EKF is a much more general tool than just the PLL, so it would seem possible to take the EKF as applied to the frequency estimation problem with known signal amplitude—which reduces to the PLL—and to generalize this EKF, allowing for higher harmonics with time-varying, unknown amplitudes. It is just this path we follow in this paper.

Indeed, Snyder in [17] takes the first step along this path. In considering a Rayleigh communication channel, he generalizes the estimator so that the signal may have a time-varying, unknown amplitude. The resulting estimator is a generalized phase-locked loop: in addition to estimating the frequency and phase, it estimates the amplitude.

The next section of this paper goes further than Snyder, first developing a state space model for the multi-harmonic signal, then giving the extended Kalman filter for the system, covering both the rectangular and polar models. In Section 3, we present theoretical results relating to our estimator based on the extended Kalman filter. Section 4 contains simulation results. The simulations are of two kinds: asymptotic results and tracking results. With the asymptotic results, we evaluate the performance of our scheme in estimating the frequency, amplitudes and phases of a periodic signal, that is a signal where the frequency, amplitudes and phases are unchanging. We get comparable results to those of [14]. In the tracking results, we demonstrate graphically how our estimator can follow changes in the amplitudes, phases and frequency. Other estimators, for example those of [14], do not have this property. Finally, Section 5, with some directions for future research, concludes the paper.

At the time of writing, we are unaware of any other papers that tackle the problem of time-varying frequency with nonsinusoidal signal. A reviewer has suggested that [9] may be relevant, although it is possible that the same remarks apply in respect of [9] as apply in respect of [14].

2. The estimator

As already stated, we apply an extended Kalman filter to estimating frequency, phases and amplitudes. To do this, we need a state space model for the quasi-periodic signal defined for the polar case of (1.5) and (1.6), or alternatively for the rectangular case of (1.7) and (1.8).

Taking the polar case first, and using the notation of (1.5) and (1.6), the state space model is

$$x(t+1) = Fx(t) + v(t), \tag{2.1}$$

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

$$= h(x(t)) + n(t), \tag{2.3}$$

where

$$x(t) = [r_1(t), r_2(t), \ldots, r_m(t), \omega(t), \theta_1(t), \theta_2(t), \ldots, \theta_m(t)]^{\mathrm{T}}, \tag{2.4}$$

$$F = \begin{bmatrix} I_m & & & & 0 \\ & 1 & & & \\ & 1 & 1 & & \\ 0 & 2 & & 1 & \\ & \vdots & & & \ddots \\ & m & & & 1 \end{bmatrix}, \tag{2.5}$$

$$h(x(t)) = \sum_{k=1}^{m} r_k(t) \sin \theta_k(t) \tag{2.6}$$

and $v(t)$ is white gaussian noise, with zero mean and variance

$$\mathrm{E}[v(t)v(t)^{\mathrm{T}}] = Q, \tag{2.7}$$

while $n(t)$ is white gaussian noise, of zero mean and

$$\mathrm{E}(n(t)v(t)) = 0, \tag{2.8}$$

$$\mathrm{E}(n^2(t)) = R. \tag{2.9}$$

The most natural Q will be one which is diagonal. However, this property is not strictly required in this section. A block diagonality property will be assumed in the next session.

In respect of the first m entries of x, viz. the harmonic amplitudes, (2.1) and (2.5) imply that they evolve as a random walk (and if Q is diagonal, the evolutions are all independent of one another). Also, $\omega(t)$ evolves in this way. Examination of the last m entries shows that each phase θ_k evolves by integrating (or summing) $k\omega(t)$, with an additional random walk. With diagonal Q, the different entries dictate the rate of random change that can occur. A zero Q would correspond to constant

amplitude, frequency and phase (the latter after subtraction of $k\omega t$).

To this signal model we apply an extended Kalman filter, which takes the measurement $z(t)$ and produces estimates $\hat{x}(t|t)$ or $\hat{x}(t|t-1)$ of $x(t)$. Here $\hat{x}(t|t)$ denotes an estimate of $x(t)$ given measurements $z(\tau)$ up to and including time t; $\hat{x}(t|t-1)$ is a one-step prediction of $x(t)$, that is an estimate of $x(t)$ given $z(\tau)$ up to time $t-1$. The extended Kalman filter, see [1, p. 195], is

$$\hat{x}(t|t)=\hat{x}(t|t-1)+L(t)[z(t)-h(\hat{x}(t|t-1))], \tag{2.10}$$

$$\hat{x}(t+1|t)=F\hat{x}(t|t), \tag{2.11}$$

$$L(t)=\Sigma(t)H(t)(H(t)^{\mathrm{T}}\Sigma(t)H(t)+R)^{-1}, \tag{2.12}$$

$$\begin{aligned}\Sigma(t+1)=F[&\Sigma(t)-\Sigma(t)H(t)\\&\times[H(t)^{\mathrm{T}}\Sigma(t)H(t)+R]^{-1}\\&\times H^{\mathrm{T}}(t)\Sigma(t)]F^{\mathrm{T}}+Q,\end{aligned} \tag{2.13}$$

where

$$H(t)=\frac{\partial h(\hat{x}(t|t-1))}{\partial \hat{x}(t|t-1)} \tag{2.14}$$

$$=\begin{bmatrix}\sin\hat{\theta}_1(t|t-1)\\ \sin\hat{\theta}_2(t|t-1)\\ \vdots\\ \sin\hat{\theta}_m(t|t-1)\\ 0\\ \hat{r}_1(t|t-1)\cos\hat{\theta}_1(t|t-1)\\ \hat{r}_2(t|t-1)\cos\hat{\theta}_2(t|t-1)\\ \vdots\end{bmatrix}, \tag{2.15}$$

and the filter is initialized by

$$\hat{x}(0|-1)=\mathrm{E}[x(0)]=\bar{x}_0, \tag{2.16}$$

$$\Sigma(0)=\mathrm{E}[(x(0)-\bar{x}_0)(x(0)-\bar{x}_0)^{\mathrm{T}}], \tag{2.17}$$

$x(0)$ having been assumed a random variable. When quantities in (2.16) or (2.17) are unknown, the algorithm must be initialized with some reasonable values of $\hat{x}(0|-1)$ and $\Sigma(0)$. Thus, to implement our estimator requires initialization, plus the update at each time instant of the four equations (2.10)-(2.13) and (2.15). This compares well (at least conceptually) with more than twenty equations updated at every time instant in Nehorai and Porat's algorithm. Of course, when our matrix equations are written out on a term-by-term basis, there are far more than five.

The application to the rectangular parametrization is almost the same as for the polar case. Equations (2.1)-(2.3) remain unchanged, while (2.4)-(2.6) become, respectively,

$$\begin{aligned}x(t)=[&a_1(t), a_2(t), \ldots, a_m(t),\\&b_2(t), b_3(t), \ldots, b_m(t), \omega(t), \theta(t)],\end{aligned} \tag{2.18}$$

$$F=\left[\begin{array}{c|c|cc}\mathrm{I}_m & 0 & \multicolumn{2}{c}{0}\\ \hline 0 & \mathrm{I}_{m-1} & \multicolumn{2}{c}{0}\\ \hline & & 1 & \\ 0 & 0 & 1 & 1\end{array}\right], \tag{2.19}$$

$$\begin{aligned}h(x(t))=&a_1(t)\sin\theta(t)\\&+\sum_{k=2}^{m}[a_k(t)\sin k\theta(t)\\&\quad+b_k(t)\cos k\theta(t)],\end{aligned} \tag{2.20}$$

here the notation is consistent with that in (1.7), (1.8). This defines the state space signal model for the rectangular case. The estimator for the rectangular case needs no change to the estimator for the polar case, save that (2.15) becomes

$$H(t)=\begin{bmatrix}\sin\hat{\theta}(t|t-1)\\ \sin 2\hat{\theta}(t|t-1)\\ \vdots\\ \sin m\hat{\theta}(t|t-1)\\ \cos 2\hat{\theta}(t|t-1)\\ \cos 3\hat{\theta}(t|t-1)\\ \vdots\\ \cos m\hat{\theta}(t|t-1)\\ 0\\ \hat{\gamma}(t|t-1)\end{bmatrix}, \tag{2.21}$$

where

$$\hat{y}(t|t-1) = \sum_{k=1}^{m} k\hat{a}_k(t|t-1)\cos k\hat{\theta}(t|t-1) - \sum_{k=2}^{m} k\hat{b}_k(t|t-1)\sin k\hat{\theta}(t|t-1). \tag{2.22}$$

Having defined the state space signal model and extended Kalman filter for both the rectangular and polar cases, it is now worthwhile to examine these more closely. First, note that other models could legitimately be advanced. For example, in our model, each of the amplitudes and the sine and cosine components is modelled by a random walk process, that is a single state, but it is quite conceivable to introduce more complicated models of these variations. Similarly, the frequency is modelled by a random walk, and the phases as summations of present and past frequency values plus a random disturbance. A typical variation is one where the frequency has a nominal value, a random walk perturbs away from this value, and a restoring term towards the nominal value is included in the model. All of these models could also be made more complicated by introducing extra states.

As the model stands, it has $2m+1$ states. The price of increasing the number of harmonics m is to increase the state size and thus the computational burden and complexity. It should be noted, however, that the only quantity inverted in the extended Kalman filter, $H^{\mathrm{T}}(t)\Sigma(t)H(t)+R$, is a scalar. Also, for both rectangular and polar cases, the F matrix is quite sparse, thus causing a lower than expected computational load.

Whether the rectangular model and estimator is better or the polar model and estimator is better is a moot point. Since the models are non-linear, it is quite likely that the extended Kalman filter does perform better for one model compared to the other. However, the presence of the non-linearity would make any theoretical resolution very difficult. Perhaps simulations would resolve the matter; we have not performed such simulations here. However, it is common knowledge[1] that for radar tracking systems, extended Kalman filters using rectangular spatial coordinates outperform those using polar spatial coordinates. It is also true that the rectangular model is more linear. However, if one actually seeks the amplitudes and phases, it is easier to estimate them directly with the polar estimator than to use the rectangular estimator and then make a rectangular-polar conversion. Therefore, each model has its advantages.

In both models there are some potential pitfalls of implementation, largely caused by the non-linearity. The first of these is that $\hat{\omega}$ can lock onto a fraction or multiple of the true frequency ω. Here for simplicity write $\hat{\omega}$ for $\hat{\omega}(t|t-1)$ and ω for $\omega(t)$. For suppose $\hat{\omega}$ locks onto $\frac{1}{2}\omega$. Then, in the polar model, $\hat{r}_1$ is around zero; $\hat{r}_2$ tracks r_1, $\hat{r}_3$ is near zero, $\hat{r}_4$ tracks r_2 and so on. A similar case applies if $\hat{\omega}$ locks onto $\frac{1}{3}\omega$. An easy remedy is to check if $\hat{r}_1 \cong 0$, and, if so, re-initialize with $\hat{\omega}$ twice (or three times) its previous value.

Similarly, if $\hat{\omega}$ locks onto 2ω then no state tracks r_1, $\hat{r}_1$ tracks r_2, nothing tracks r_3, $\hat{r}_2$ tracks r_4 and so on. In this case a remedy is also possible. Since the first, third and other odd harmonics are not being modelled, the 'innovations process' of the filter,

$$z(t) - h(\hat{x}(t|t-1)),$$

will have greater energy than would be the case if $\hat{\omega} \cong \omega$. Thus if, either as a matter of course or through lack of confidence in $\hat{\omega}$, one initializes a second filter with half the frequency estimate of the first, monitoring the 'innovations process' of the two filters as well as the amplitude will tell one to reset the estimate $\hat{\omega}$ to half its previous value.

In different simulations, we observed $\hat{\omega}$ to lock onto 2ω and $\frac{1}{2}\omega$. The difficulty is inherent in the problem and could equally occur in Nehorai and Porat's estimator. Essentially it is a problem caused by poorly initializing the state estimate, in particular the frequency estimate. It is well-known

[1] We are indebted to colleagues for drawing our attention to this point.

that a phase locked loop may not capture a signal if its centre frequency—i.e., the initial frequency estimate—is too distant from the signal's actual frequency. Therefore, it is no surprise that good initialization is also important in our more general problem.

Another potential difficulty is cycle slipping, viz., $\theta_k(t)$ and $\hat{\theta}_k(t)$, for some k and t differing (approximately) by a multiple of 2π. In simulations, such cycle slipping was only observed at low SNR levels, and then only in the transient phase. Generally, slipping a cycle is probably of little consequence and can probably be ignored. Indeed, as the signal model is written, all the phases are cumulative summations of the instantaneous frequency $\omega(t)$; therefore, the phases and their estimates will be roughly ramp functions. To prevent numerical overflow, it is probably wise to check that the phase estimates do not exceed 2π and if they do, subtract 2π from the overflowing estimates.

A more subtle form of cycle slipping can occur only in the polar model. In this case $\hat{r}_k$ could become negative for some k, implying tracking of $-r_k$, and $\hat{\theta}_k$ would then differ from θ_k by π. The remedy is easy. If $\hat{r}_k$ is negative, change its sign and add π to $\hat{\theta}_k$. This type of cycle slipping was also observed in simulations.

This section has derived the state space models for both the rectangular and polar cases, along with the extended Kalman filter estimators. The estimators are considerably simpler than the estimator of Nehorai and Porat. We have also surveyed how the estimators may fail to work, and have given a remedy for each such difficulty. In other words, we have dealt with some practical difficulties of implementation. By contrast, the next section covers theoretical issues.

3. A theoretical understanding of the estimator

Since the signal model is nonlinear, it is very difficult to analyse our estimator. The best we can hope for is an intuitive and partial theoretical understanding. This section attempts such an analysis. To avoid repetition, we will consider only the polar model; however, most of the results should extend to the rectangular model. For simplicity, we sometimes will consider the case of only two harmonics, $m = 2$; however, all the results generalize to the case of higher harmonics present.

We begin with a result concerning solutions of the Riccati equation (2.13) for this problem. While the result may not be of intrinsic interest, it is useful in establishing later results. We need to introduce some notation. Recall the vector $H(t)$ of (2.15). The time variation in $H(t)$ is due to that in $\hat{\theta}_k$ and that in $\hat{r}_k$ for $k = 1, \ldots, m$. The former variation is fast while the latter is slow from our earlier assumptions. These two types of variation will similarly be found in expressions such as $H(t)R^{-1}H(t)^{\mathrm{T}}$. Let us define

$$N = \text{ave}\, H(t)R^{-1}H(t)^{\mathrm{T}} \tag{3.1}$$

to denote the effect of averaging over the fast variation, or equivalently, the application of low pass filtering to $H(t)R^{-1}(t)H(t)^{\mathrm{T}}$, where the frequency ω lies outside the filter bandwidth. This point requires some clarification. Consider first a quantity such as $\sin\theta_k(t)$. Since $\theta_k(t)$ is the total phase, the time variation $\theta_k(t)$ is due to the integration or summing of $\omega(\cdot)$ together with slow variation in the relative phase. Over a few cycles [defined by $\omega(\cdot)$], $\sin^2\theta_k(t)$ then takes the average value of $\frac{1}{2}$. Now we can extend this idea to quantities such as $\sin^2\hat{\theta}_i(t/t-1)$ which is approximately $\sin^2\theta_i(t)$, and $\hat{r}_j(t/t-1)\cos\hat{\theta}_j(t/t-1)$. The only fast variation is attributable to ω, as we have postulated relative phases and harmonic amplitudes—and thus presumably their estimates—are slowly varying. Evidently, from (2.15) it follows that

$$N = R^{-1}\,\text{diag}\,[\tfrac{1}{2}, \tfrac{1}{2}, \ldots, \tfrac{1}{2}, 0, \tfrac{1}{2}\hat{r}_1^2, \ldots, \tfrac{1}{2}\hat{r}_{m^2}], \tag{3.2}$$

where $\hat{r}_1$ is short for $\hat{r}_1(t|t-1)$. Clearly N is a slowly varying function of time and constant if the amplitudes are known and constant (or estimated to be constant).

The slow variation in N is attributable to the slow variation in the $\hat{r}_i$, or ultimately the r_i; this is not averaged out over the fast variation averaging (which occurs over a few cycles), or one can say that it will not be averaged out by a low pass filtering operation, provided the bandwidth of the filter exceeds the bandwidth of the r_i variations (while of course being significantly less than ω).

The next theorem concerns the Riccati equation for $\Sigma(t)$, given in (2.13). It is well-known that there are certain time constants associated with many Riccati equations, see [1], primarily because a linear equation (Hamiltonian equation) can be closely coupled with a Riccati equation. Snyder, see [17], argued that when the coefficients of a Riccati equation varied with a time constant much faster than that of the Riccati equation itself, then the coefficients could be replaced by their average values, without significantly affecting the solution of the equation, because of some inbuilt filtering action associated with the equation. It is this sort of result that we shall use here.

In the following theorem, we shall make use of the quantity

$$\bar{H} = \operatorname{diag}\left(\frac{1}{\sqrt{2}}\ \frac{1}{\sqrt{2}} \cdots \frac{1}{\sqrt{2}}\ 0\ \frac{\hat{r}_1}{\sqrt{2}}\ \frac{\hat{r}_2}{\sqrt{2}} \cdots \frac{\hat{r}_m}{\sqrt{2}}\right). \tag{3.3}$$

Notice that

$$N = \bar{H}(R\mathrm{I})^{-1}\bar{H}^{\mathrm{T}}. \tag{3.4}$$

Replacing $H(t)$ with $\bar{H}$ in the Riccati equation (2.13) gives a new Riccati equation with solution $\bar{\Sigma}$:

$$\bar{\Sigma} = F(\bar{\Sigma} - \bar{\Sigma}\bar{H}[\bar{H}^{\mathrm{T}}\bar{\Sigma}\bar{H} + R\mathrm{I}]^{-1}\bar{H}^{\mathrm{T}}\bar{\Sigma})F^{\mathrm{T}} + Q, \quad \bar{\Sigma} \geqslant 0. \tag{3.5}$$

Note: if N is constant, (3.5) is just a steady state Riccati equation, with the constraint $\bar{\Sigma} \geqslant 0$ picking out the solution of interest in filtering theory. With N slowly time-varying, we must reevaluate the (steady state) solution of (3.5) for each t.

The time constants of the filter defined by $\bar{\Sigma}$ are determined by the eigenvalues of the filter system matrix, which is

$$\bar{F} = F(\mathrm{I} - \bar{\Sigma}\bar{H}[\bar{H}^{\mathrm{T}}\bar{\Sigma}H + R\mathrm{I}]^{-1}\bar{H}^{\mathrm{T}}). \tag{3.6}$$

Now we can state the following key result, which generalizes an idea probably originally presented in [17].

THEOREM 3.1. Consider the definition of $\Sigma(t)$ in (2.13) *and $\bar{\Sigma}$ above. Suppose that the eigenvalues of $\bar{F}$ are much closer to the point* $+1$ *than the point* $\exp(\mathrm{j}\omega)$, *so that the system $\alpha_{k+1} = \bar{F}\alpha_k + \beta_k$ (with β_k as input, α_k as response) would be a low pass system in relation to signals of frequency ω. Then as $t \to \infty$, $\Sigma(t) - \bar{\Sigma}$ is guaranteed to be small. Further, if Q is block diagonal of the form [diagonal $m \times m$ matrix] + [$(m+1) \times (m+1)$ matrix], a form also possessed by F and N, then $\bar{\Sigma}$ has this form.*

PROOF. See Appendix.

NOTE. For the rectangular model, the block diagonality above will take a different structure, but apart from that the result does not change.

The matrix $\bar{\Sigma}$ defined above is approximately a constant, since all elements in its definition are constant except N, which varies only slowly with time. The block diagonality of F is clear, that of Q is reasonable to assume and that of N is checkable. Thus, the matrix $\Sigma(t)$ is nearly constant and block diagonal except for some small, possibly high frequency terms. In what follows, we will repeatedly use the fact that $\Sigma(t) \cong \bar{\Sigma}$. In the replacement of Σ by $\bar{\Sigma}$ to constitute a good approximation, it is important that the frequency ω lies outside the pass band of the filter defined by $\bar{\Sigma}$. This is generally found in applications, but has to be verified.

There are several immediate consequences to note.

(1) The block diagonal character of $\bar{\Sigma}$ and indeed the diagonal character of its top left $m \times m$ block means that the first m diagonal entries of $\bar{\Sigma}$ are analytically computable. Suppose that $Q =$

diag $[q_1, q_2, \ldots, q_m, \ldots]$. Then the i-i entry of $\bar{\Sigma}$ satisfies

$$\sigma_i = [\sigma_i - \sqrt{\tfrac{1}{2}}\sigma_i(\tfrac{1}{2}\sigma_i + R)^{-1}\sqrt{\tfrac{1}{2}}\sigma_i] + q_i,$$

whence

$$\sigma_i = \tfrac{1}{2}q_i[1 + \sqrt{(1 + 8R/q_i)}].$$

(2) Suppose that all phases $\theta_k(t)$ in the model (1.6) were known, while the r_k were unknown. The state variable model would be a truncated version of (2.1)-(2.5) involving the first m entries of $x(t)$ only, and an $m \times m$ identity F matrix. Also, $h(x(t))$ would become linear in the state. It is not hard to check that, with no approximation at all, the steady state covariance matrix associated with one-step predictions of $r_1(t), \ldots, r_m(t)$ would be precisely the top left corner of $\bar{\Sigma}$, which with diagonal Q is analytically computable as explained above. Thus the true performance of the EKF in estimating amplitudes, of which the top left $m \times m$ corner of $\Sigma(t)$ is an indication, is approximated by the performance of an amplitude estimator when all phases are known, viz. diag $[\sigma_1, \sigma_2, \ldots, \sigma_m]$.

(3) Suppose now that all the amplitudes r_k are known. In this case, there is an EKF, but not exact linear Kalman filter, for $\omega(t)$, $\theta_1(t), \ldots, \theta_m(t)$. (It is derivable by simplifying the derivation given earlier, to delete $r_k(t)$ variation and estimation thereof.) An averaging argument, like that behind Theorem 3.1, shows that the solution of the Riccati equation associated with the EKF can be approximated by the solution of a second Riccati equation (with averaging of fast changing quantities). This second Riccati equation is precisely the same as the last $(m+1)$ rows and $(m+1)$ columns of the equation for $\bar{\Sigma}$, with r_i replacing $\hat{r}_i$. In other words, the equation for $\bar{\Sigma}$ presents approximate information concerning the frequency and phase estimator errors, approximate because the amplitude estimates are treated as exact.

(4) The following special cases can be identified: (i) $m = 1$ and r_1 is known, this being angle modulation, a continuous-time version of which is treated in [17] (using the Riccati equation) and in [20] (using MAP estimator theory); (ii) $m = 1$ and r_1 is unknown, this being angle modulation over a Rayleigh channel, and this problem is treated in the same references.

(5) Focus on the bottom right $(m+1) \times (m+1)$ corner of $\bar{\Sigma}$. From (3.5), it is clear that if we scale R upwards by α^2 and $\bar{H}$ upwards by α, there is no change in $\bar{\Sigma}$. Considering how $\bar{H}$ is constructed (see (3.2)), this says that the accuracy in estimating ω and the θ_k is effectively determined by signal-to-noise ratios (the collections r_i^2/R or, approximately, $\hat{r}_i^2/R$). This is a phenomenon familiar from frequency modulation and consistent with remarks made earlier in the paper. Obviously, if the signal-to-noise ratio is low enough, we can expect large errors in our estimate, and consequential errors in the linearization assumption underpinning the EKF. (This assumption says that a linearization of the signal model using estimated values is virtually the same as a linearization using true but unknown values.)

We now turn to examining the estimator itself. The first major point we aim to show is that the estimator we have derived is like a generalized phase locked loop.

We warn the reader that the purpose of this analysis (even more so than with the covariance analysis) is to give understanding and guidance, rather than exact calculations. Consider the gain $L(t)$ used in the estimator:

$$L(t) = \Sigma(t)H(t)[H(t)^{\mathrm{T}}\Sigma(t)H(t) + R]^{-1}. \tag{3.7}$$

If signals are sampled fast enough, so that the discrete-time filter approximates closely a continuous-time filter, it is known that $H(t)^{\mathrm{T}}\Sigma(t)H(t) + R \to R$. This suggests that

$$L(t) \cong \Sigma(t)H(t)R^{-1} \cong \bar{\Sigma}H(t)R^{-1}. \tag{3.8a}$$

Alternatively, we have (after minor manipulation)

$$L(t) = [\Sigma^{-1}(t) + H(t)R^{-1}H^{T}(t)]^{-1}H(t)R^{-1}$$
$$\cong [\bar{\Sigma}^{-1} + N]^{-1}H(t)R^{-1}. \quad (3.8b)$$

In either case, we are suggesting that

$$L(t) \cong KH(t), \quad (3.8c)$$

for some matrix K in which any time variation is slow.

Now the filter is driven by $L(t)[z(t) - h(\hat{x}(t-1))]$. We shall study the frequency content of this signal. For the sake of simplicity, suppose that two harmonics are present. Observe that

$$L(t)z(t) = K \begin{bmatrix} \sin \hat{\theta}_1(t|t-1) \\ \sin \hat{\theta}_2(t|t-1) \\ 0 \\ \hat{r}_1(t|t-1)\cos \hat{\theta}_1(t|t-1) \\ \hat{r}_2(t|t-1)\cos \hat{\theta}_2(t|t-1) \end{bmatrix}$$
$$\times [r_1(t)\sin \theta_1(t) + r_2(t)\sin \theta_2(t) + n(t)]. \quad (3.9)$$

Recall that r_1, r_2, and therefore $\hat{r_1}$, $\hat{r}_2$, are presumed slowly varying while the total phases θ_1, θ_2, and therefore $\hat{\theta}_1$, $\hat{\theta}_2$, have an underlying variation of ωt and $2\omega t$, together with a slow variation on top of that from the relative phase. This means that $\theta_1 - \hat{\theta}_1$, $\theta_2 - \hat{\theta}_2$ are slowly varying. Putting this together means

$$L(t)z(t) = \tfrac{1}{2}K \begin{bmatrix} r_1 \cos(\theta_1 - \hat{\theta}_1) \\ r_2 \cos(\theta_2 - \hat{\theta}_2) \\ 0 \\ r_1\hat{r}_1 \sin(\theta_1 - \hat{\theta}_1) \\ r_2\hat{r}_2 \sin(\theta_2 - \hat{\theta}_2) \end{bmatrix}$$
$$+ \text{high frequency terms} + \text{noise} \quad (3.10)$$

('high frequency' means frequency ω or higher). Similarly,

$$L(t)h(\hat{x}(t|t-1)) = \tfrac{1}{2}K \begin{bmatrix} \hat{r}_1 \\ \hat{r}_2 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$
$$+ \text{high frequency terms.} \quad (3.11)$$

The amplitude update and the frequency phase updates separate. Assume (as is reasonable) that $K = K_a + K_p$. From (2.1) and (2.11) together with the expressions in (3.10) and (3.11), we obtain

$$\begin{bmatrix} \hat{r}_1(t|t) \\ \hat{r}_2(t|t) \end{bmatrix} = \begin{bmatrix} \hat{r}_1(t|t-1) \\ \hat{r}_2(t|t-1) \end{bmatrix}$$
$$+ \tfrac{1}{2}K_a \begin{bmatrix} r_1 \cos[\theta_1(t) - \theta_1(t|t-1)] - \hat{r}_1(t|t-1) \\ r_2 \cos[\theta_2(t) - \hat{\theta}_2(t|t-1)] - \hat{r}_2(t|t-1) \end{bmatrix}$$
$$+ \text{noise}, \quad (3.12a)$$

$$\begin{bmatrix} \hat{r}_1(t+1|t) \\ \hat{r}_2(t+1|t) \end{bmatrix} = \begin{bmatrix} \hat{r}_1(t|t) \\ \hat{r}_2(t|t) \end{bmatrix} \quad (3.12b)$$

and

$$\begin{bmatrix} \hat{\omega}(t|t) \\ \hat{\theta}_1(t|t) \\ \hat{\theta}_2(t|t) \end{bmatrix} = \begin{bmatrix} \hat{\omega}(t|t-1) \\ \hat{\theta}_1(t|t-1) \\ \hat{\theta}_2(t|t-1) \end{bmatrix}$$
$$+ \tfrac{1}{2}K_p \begin{bmatrix} 0 \\ r_1\hat{r}_1 \sin[\theta_1(t) - \hat{\theta}_1(t|t-1)] \\ r_2\hat{r}_2 \sin[\theta_2(t) - \hat{\theta}_2(t|t-1)] \end{bmatrix} + \text{noise}, \quad (3.13a)$$

$$\begin{bmatrix} \hat{\omega}(t+1|t) \\ \hat{\theta}_1(t+1|t) \\ \hat{\theta}_2(t+1|t) \end{bmatrix} = \begin{bmatrix} \hat{\omega}(t|t) \\ \hat{\theta}_1(t|t) + \hat{\omega}(t|t) \\ \hat{\theta}_2(t|t) + 2\hat{\omega}(t|t) \end{bmatrix}. \quad (3.13b)$$

These equations show in effect what the EKF is doing. There is clearly some decoupling of amplitude estimation from frequency and phase estimation. In case the error $\theta_1 - \hat{\theta}_1$ is small (so that $\cos(\theta_1 - \hat{\theta}_1) \cong 1$), the update equations for $\hat{r}_1$ are standard Kalman filter equations, identical with those applicable when the phases are exactly known (and the decoupling is then complete). Notice that $\cos^{-1} 0.9 \cong 26°$; so quite sizeable phase errors can presumably be tolerated before amplitude estimation becomes significantly different because of phase error.

The frequency-phase equations (3.13) in effect define two coupled phase-locked loops. Since continuous-time phase-locked loops are perhaps better understood by most readers, let us examine a

continuous-time equivalent of (3.13) to grasp the structure better:

$$\begin{bmatrix} \dot{\hat{\omega}} \\ \dot{\hat{\theta}}_1 \\ \dot{\hat{\theta}}_2 \end{bmatrix} = \tfrac{1}{2}K_p \begin{bmatrix} 0 \\ r_1\hat{r}_1 \sin(\theta_1 - \hat{\theta}_1) + \hat{\omega} \\ r_2\hat{r}_2 \sin(\theta_2 - \hat{\theta}_2) + 2\hat{\omega} \end{bmatrix} + \text{noise}.$$

This is diagrammatically represented in Fig. 1.

Actually, the low pass filters can be removed, since the dynamics to their right in Fig. 1 provide low pass filtering. Figure 2 is a model for the phase estimation of Fig. 1, of a type often used in considering a single phase-locked loop (noise is neglected). When $\theta_1 - \hat{\theta}_1$ and $\theta_2 - \hat{\theta}_2$ are small, the sine nonlinearity can be removed and linear analysis can be used to obtain quantities such as time constants.

If the lower low-pass filter and $\hat{\theta}_2$ generator in Fig. 1 are neglected for a moment, one can see that the upper structure corresponds to a standard second order phase-locked loop, see [9], which is capable of capturing a constant frequency signal with zero steady state offset. Obviously the figure depicts two interconnected phase-locked loops. That there should be interconnection (through the integrator with output $\hat{\omega}$) is hardly any surprise, given the harmonic relationship between the different components in the signal.

Note also that, due to lack of knowledge of r_1, r_2, the estimates $\hat{r}_1$, $\hat{r}_2$ are used, where otherwise true values would be used. These estimates are present both directly (as inputs to multipliers) and indirectly (because K_p depends in part on the $\hat{r}_i$).

The latter material of this section, dealing with the rough equivalence of the EKF with coupled phase-locked loops, while not intended to explain how the estimate should be constructed, nevertheless shows how it could be constructed. The earlier material of this section is important no matter how the estimator is constructed, in that it describes how the estimator performance can be approximately evaluated. All in all, this section has shown

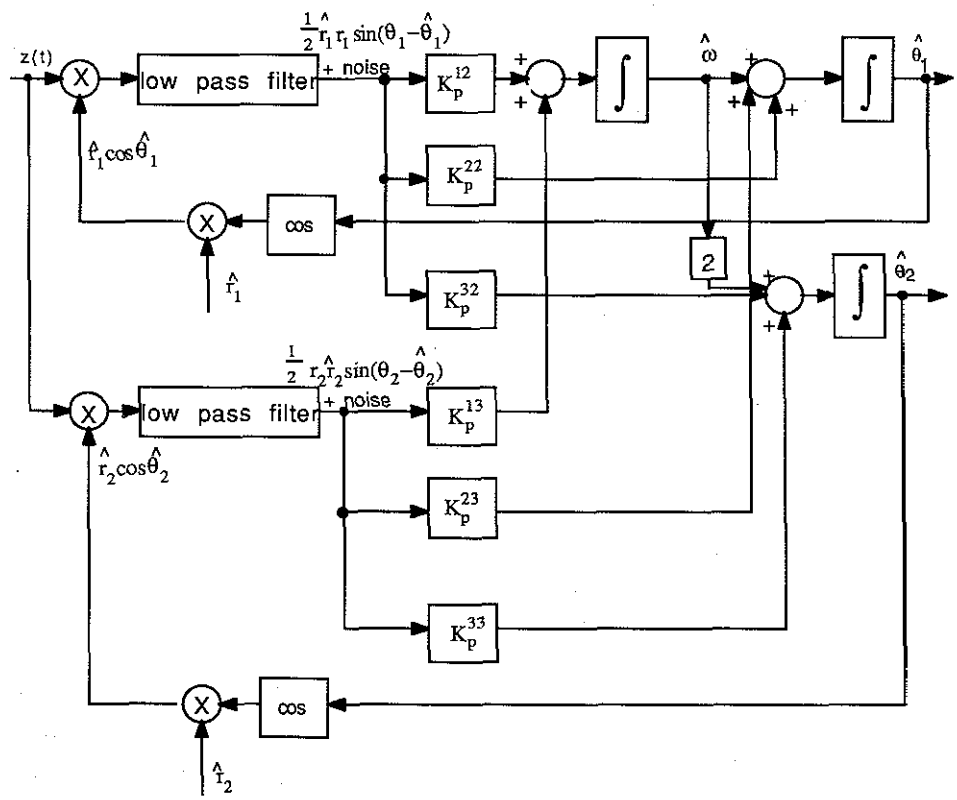

Fig. 1. Frequency and phase estimators via coupled phase-locked loops.

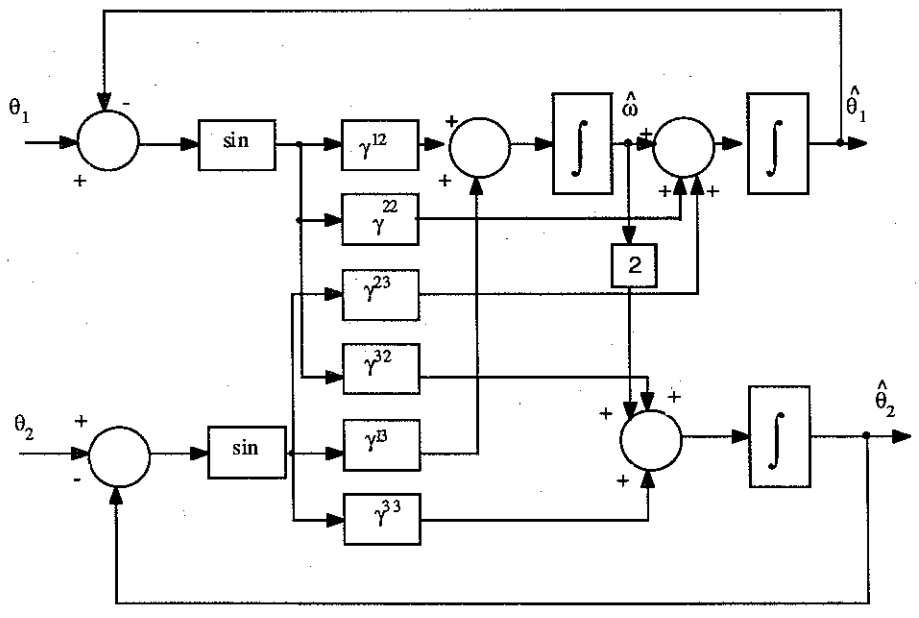

Fig. 2. Equivalent model for phase and frequency estimators.

both conceptually and practically useful properties of our estimator. Although the tasks of producing the amplitude estimates and the phase and frequency estimates are inter-linked, the estimator has a certain separation of these estimates. That is, the phase and frequency estimates are calculated as if the amplitude estimates were correct, while the amplitude estimates are calculated as if the phase and frequency estimates were correct. Though these properties are pleasing, they are hardly a complete analysis of the estimator, so for further insight we must turn to simulations.

4. Simulations

All the simulations presented here are for the polar model. The first set of simulations demonstrate asymptotic behaviour; the second set of simulations shows transient and tracking behaviour. Because the estimator we have propounded is highly non-linear—a necessary consequence of the problem—it is impossible to analyse it completely. Therefore, the simulations are an important part of evaluating the estimator's performance.

To make this evaluation easier, the asymptotic simulations are exactly the same as those performed by Nehorai and Porat in [14]. We present both our results and those of Nehorai and Porat. The input signal to the estimator was, as in [14],

$$z(t) = \sum_{k=1}^{5} r_k \sin 2\pi 0.08kt + n(t). \tag{4.1}$$

That is, the number of harmonics $m = 5$, while $n(t)$ was the unit variance, zero mean white gaussian noise. The signal's fundamental frequency, $\omega = 2\pi 0.08$, and the amplitudes r_k were constant over time. The amplitudes were

$$r_k = r_1/k, \quad k = 2, 3, 4, 5, \tag{4.2}$$

where r_1 was chosen such that the signal-to-noise ratio, given by

$$\mathrm{SNR(dB)} = 10 \log_{10}\left(\sum_{k=1}^{5} \tfrac{1}{2} r_k^2 \right), \tag{4.3}$$

was at 0 dB, 8 dB or 16 dB depending on the trial. For each of the SNR levels of 0, 8 and 16 dB, one hundred independent Monte Carlo experiments

were run, each over 200 time samples. At the final time instant, the error in the state estimate was calculated and then the statistics over the hundred Monte Carlo runs were computed. Exactly the same experiments were repeated for 500 time samples on each Monte Carlo run.

When the algorithm was initialized at the start of each Monte Carlo run, the following initial conditions were used:

$$\hat{r}_1(0|-1) = (0.75\ \mathrm{E}\{z(t)^2\})^{1/2}, \tag{4.4}$$

$$\hat{r}_2(0|-1) = \hat{r}_3(0|-1) = \hat{r}_4(0|-1) = \hat{r}_5(0|-1) = 0. \tag{4.5}$$

The value of $\hat{r}(0|-1)$, was chosen in (4.4) as a way of easily estimating the fundamental component's strength from only the statistics of the incoming signal. Setting the estimates of higher harmonics' amplitudes to zero in (4.5) reflects a lack of knowledge about the signal's harmonic structure as may well be encountered in practice.

Additional initial conditions were as follows:

$$\hat{\theta}_k(0|-1) = 0, \quad k = 1, \ldots, 5 \tag{4.6}$$

and

$$\hat{\omega}(0|-1) = 2\pi 0.05. \tag{4.7}$$

Obviously, (4.6) is a natural initial estimate. Use of (4.7) provides a test for an algorithm, in particular of its capture region, but also allows some possibility that $\hat{\omega}$ approaches $\frac{1}{2}\omega$, rather than ω, since $\hat{\omega}(0|-1)$ is in fact nearer to $\frac{1}{2}\omega$ than ω. Equations (4.4)-(4.7) describe the initial, one step ahead, state estimate. The initial Σ matrix was

$$\Sigma(0) = \mathrm{diag}\{\sigma_1, \tfrac{1}{4}\sigma_1, \tfrac{1}{9}\sigma_1, \tfrac{1}{16}\sigma_1, \tfrac{1}{25}\sigma_1, (0.06\pi)^2, 0.02, \ldots, 0.02\}, \tag{4.8}$$

where

$$\sigma_1 = \tfrac{1}{4}(\hat{r}_1(0|-1))^2. \tag{4.9}$$

The relationship between the first five levels reflects (4.2), while the choice of σ_1 is a matter of judgement. The same is true for the other entries of $\Sigma(0)$.

Consistent with the signal model (4.1) we chose $R = 1$, while

$$Q = \mathrm{diag}\{q_1, q_1, \ldots, q_1, q_2, q_1, \ldots, q_1\}, \tag{4.10}$$

where $q_1 = 1 \times 10^{-4}$ and $q_2 = 1 \times 10^{-8}$, the latter being the element of Q corresponding to the frequency ω. These were the values of Q and R used to run the estimator. In the case of R, the value used was, in fact, the variance of the measurement noise $n(t)$; in the case of Q, the value used did not correspond to the variance in the amplitudes and phases, since these parameters, as already stated, were constant. It is of course quite common in Kalman filtering to use values of Q and R for the filter design which are not identical with those of the signal model, see [1]. Indeed, when the state being estimated is known to be constant, it is regarded as bad practice to take $Q = 0$ in designing the estimator, because of a phenomenon called data saturation.

The entries of Q and R roughly determine the bandwidth of the estimator. This in turn affects the capture properties, tracking properties and steady state error. The capture properties can also be influenced by the choice of $\Sigma(0)$; if this is very different to a typical steady state value of Σ, there can be a transient phase in which the capture bandwidth is, for example, larger.

The values assumed here were largely obtained by trial and error.

The results of the simulations appear in Tables 1-3. In each of these tables, the numbers in parentheses are the results of Nehorai and Porat in [14], while the numbers without parentheses are our results. In Table 1, the third column, labelled 'outliers', is an appellation given by Nehorai and Porat to those Monte Carlo experiments where the error in the estimated frequency was greater than an arbitrary threshold of 0.003π radians per time sample. Since this error threshold is only 1.875% of the true frequency, it is rather a tight definition of an outlier; but, in the interests of comparison with Nehorai and Porat's results, we have kept the same definition. For the same reason, we excluded

Table 1
Statistical results for the fundamental frequency estimate via EKF (Nehorai and Porat results shown in parentheses)

			ω	
N (samples)	SNR (dB)	Outliers	Bias $\times 10^{-6}$	Standard deviation $\times 10^{-5}$
200	0	39 (7)	79.6 (−98.7)	40 (42.3)
	8	4 (3)	11.8 (−26.0)	10 (14.3)
	16	0 (0)	−2.59 (1.20)	10 (5.05)
500	0	34 (6)	4.76 (−22.6)	20 (14.4)
	8	4 (3)	15.9 (0.40)	10 (2.03)
	16	0 (0)	6.52 (−0.99)	10 (0.78)

the outlier trials when compiling the statistics of the frequency, amplitude and phase estimates.

Examining Table 1 shows that while the number of outliers our scheme incurred was approximately the same as Nehorai and Porat at SNR of 8 dB and 16 dB, at the lower SNR of 0 dB our scheme performed much worse. Indeed, at 0 dB our method appears to work very poorly, scoring 39 outliers versus 7 for Nehorai and Porat when the number of time samples $N = 200$, and 34 outliers versus 6 for $N = 500$.

Why our method performed so poorly at 0 dB is open to some conjecture. Certainly some of the outliers would have been due to $\hat{\omega}$ locking mistakenly onto 2ω or $\frac{1}{2}\omega$ as discussed in Section 2. In our experience of running simulations, locking onto a multiple, or a fraction, of the true frequency is a problem with the transient response: once the estimator has locked onto a frequency, its local stability properties seem to prevent it from readily losing that frequency. This is rather like the disparity between capture and lock regions in the regular phase-locked loop.

If the high number of outliers is a result of the transient response, this suggests a possible remedy. In initializing the estimator, the values of $\Sigma(0)$, Q and R may be chosen so that the transient response is more sluggish, less likely to be greatly influenced by a single noisy measurement. In other words, one could substitute fast transient response for good noise rejection. (With R constant, reduction of $\Sigma(0)$ will lower the bandwidth of the estimator during its transient phase, while reduction of Q will lower the steady state bandwidth of the estimator.) In our simulations we used the same $\Sigma(0)$, Q and R for all trials. By contrast, Nehorai and Porat used different parameter values in their algorithm for the SNR = 0 dB case, and this may account for their much better performance at that low value of the signal-to-noise ratio.

Looking now at the 8 dB and 16 dB cases, our algorithm seems to perform comparably with Nehorai and Porat in estimating the frequency. For $N = 200$, our method performs better at 8 dB; Nehorai and Porat's method exceeds ours at 16 dB. For $N = 500$, Nehorai and Porat perform better than our method at both 8 dB and 16 dB.

In amplitude estimation, again restricting our view to 8 dB and 16 dB because of the excessive number of outliers at 0 dB, we see that we mostly perform better than Nehorai and Porat. This is particularly true for the bias.

In estimating the phase, however, we consistently perform worse than Nehorai and Porat. There is, we believe, a good reason for this. Whereas we worked with a polar signal model, directly estimating the amplitudes and phases, Nehorai and Porat used a rectangular signal model, first estimating the amplitudes of sine and cosine components, then converting from rectangular to polar coordinates to get each amplitude and phase. While our polar signal model allowed us to estimate very accurately the amplitudes, it seems a model with low sensitivity to changes in the phases, hence the poor phase estimates. However, were we to use a rectangular signal model, as outlined in Section 2 of this paper, then, presumably, we could estimate the sine and cosine components with a similar accuracy as we could estimate the amplitudes of the polar model, and thus with at least the accuracy of Nehorai and Porat. Since the phase is derived from sine and cosine components in the same way as Nehorai and Porat, we could therefore achieve similar phase accuracy as their method.

Table 2
Statistical results for the amplitude estimate via EKF (Nehorai and Porat results shown in parentheses)

		r_1			r_2			r_3			r_4			r_5		
N (samples)	SNR (dB)	True	Bias	Standard deviation	True	Bias	Standard deviation	True	Bias	Standard deviation	True	Bias	Standard deviation	True	Bias	Standard deviation
200	0	1.17	−0.12 (−0.08)	0.19 (0.15)	0.58	−0.14 (−0.13)	0.17 (0.13)	0.39	−0.11 (0.13)	0.11 (0.10)	0.29	−0.09 (−0.14)	0.11 (0.09)	0.23	−0.09 (−0.08)	0.08 (0.08)
	8	2.94	−0.039 (−0.13)	0.17 (0.32)	1.47	−0.07 (−0.12)	0.15 (0.24)	0.98	−0.10 (−0.15)	(0.19) (0.19)	0.73	−0.08 (−0.13)		0.59	−0.07 (−0.15)	0.11 (0.15)
0.10 (0.17)																
	16	7.38	−0.007 (−0.06)	0.11 (0.22)	3.69	−0.04 (−0.07)	0.11 (0.18)	2.46	−0.07 (−0.05)	0.17 (0.18)	1.84	−0.08 (−0.07)		1.48	−0.10 (−0.09)	0.15 (0.19)
0.17 (0.21)																
500	0	1.17	−0.053 (−0.04)	0.11 (0.10)	0.58	−0.04 (−0.05)	0.12 (0.09)	0.39	−0.06 (−0.04)	0.11 (0.09)	0.29	−0.06 (−0.06)		0.23	−0.05 (−0.04)	0.09 (0.07)
0.11 (0.08)																
	8	2.94	−0.005 (−0.02)	0.09 (0.08)	1.47	−0.01 (−0.02)	0.09 (0.08)	0.98	−0.02 (−0.03)	0.09 (0.07)	0.73	−0.02 (−0.04)		0.59	−0.02 (−0.04)	0.09 (0.08)
0.09 (0.08)																
	16	7.38	−0.01 (−0.02)	0.07 (0.12)	3.69	−0.01 (−0.02)	0.08 (0.11)	3.46	−0.03 (−0.02)	0.09 (0.11)	1.84	−0.02 (−0.03)		1.48	−0.03 (−0.03)	0.08 (0.10)
0.09 (0.11)																

Table 3
Statistical results for the phase estimates via EKF (Nehorai and Porat results shown in parentheses)

		θ_1		θ_2		θ_3		θ_4		θ_5	
N (samples)	SNR (dB)	Bias $\times 10^{-1}$	Standard deviation $\times 10^{-1}$	Bias $\times 10^{-1}$	Standard deviation $\times 10^{-1}$	Bias $\times 10^{-1}$	Standard deviation $\times 10^{-1}$	Bias $\times 10^{-1}$	Standard deviation $\times 10^{-1}$	Bias $\times 10^{-1}$	Standard deviation $\times 10^{-1}$
200	0	−4.47	3.38	−7.27	6.92	0.58	12.9	5.96	9.49	4.12	7.26
		(0.68)	(4.48)	(0.98)	(7.64)	(−0.02)	(9.04)	(0.48)	(9.21)	(0.83)	(8.66)
	8	−4.97	0.55	−9.84	1.63	−4.55	13.7	10.81	3.20	6.03	2.01
		(−0.00)	(1.25)	(0.56)	(2.88)	(0.09)	(4.49)	(0.76)	(5.42)	(2.13)	(6.21)
	16	−5.07	0.32	−10.01	0.56	−9.46	11.7	11.19	0.90	6.23	1.14
		(−0.08)	(0.56)	(−1.66)	(0.11)	(−0.33)	(0.17)	(−0.37)	(0.21)	(0.19)	(2.42)
500	0	−5.01	1.49	−8.37	4.83	−4.62	12.3	5.07	10.44	6.50	5.69
		(0.06)	(2.42)	(2.99)	(4.29)	(0.22)	(6.33)	(−0.52)	(7.27)	(0.26)	(8.70)
	8	−4.95	0.54	−9.94	0.95	−6.34	13.1	11.54	1.71	6.56	2.10
		(0.01)	(0.51)	(0.56)	(1.05)	(0.01)	(1.45)	(−0.04)	(2.01)	(−0.12)	(2.71)
	16	−5.01	0.33	−10.14	0.52	−10.60	10.5	11.46	1.00	6.21	1.13
		(0.03)	(0.25)	(0.65)	(0.47)	(−0.02)	(0.65)	(0.02)	(0.78)	(0.03)	(0.97)

In summary, the asymptotic results of Tables 1-3 show several points. Our algorithm performs worse at 0 dB. This may or may not be due to poorly selected algorithm parameters. At higher SNR levels, our algorithm performs comparably with Nehorai and Porat's algorithm when estimating amplitudes. On the other hand, Nehorai and Porat perform better than we in estimating phases, although this may be because they use a rectangular signal model while we use a polar model.

We now consider the tracking behaviour of our algorithm, something that other algorithms cannot emulate. Whereas the asymptotic results we presented were tabular and quantitative, the tracking results are graphical and qualitative. Figures 3-8 show the results of a single trial. In this case the signal was initially the same as that in the asymptotic results, with SNR = 8 dB. However, we added some white process noise $v(t)$—see (2.1)—into the signal model, causing the frequency, amplitudes and phases to evolve with time. The process noise had variance

$$\mathrm{E}[v(t)v(t)'] = \mathrm{diag}\{a_1, a_1, a_1, a_1, a_1, a_2, a_1, a_1, a_1, a_1, a_1\},$$

where

$$a_1 = 1 \times 10^{-3}, \qquad a_2 = 3 \times 10^{-7}.$$

Note that the state with variance of a_2 corresponds to the frequency ω. The initialization of the estimator was the same as for the asymptotic trials.

Figure 3 shows how the true frequency and its estimate evolve with time. Note that although the frequency estimate starts at a very inaccurate initial value, it locks onto the true value and tracks it closely after time $t = 50$. Figure 4 is an expanded scale version of Fig. 3, showing the tracking behaviour from $t = 60$ to $t = 200$. Throughout this time, the frequency error never exceeds 0.6% of the actual frequency.

Figure 5 shows the estimated and true amplitudes of the fundamental frequency component r_1.

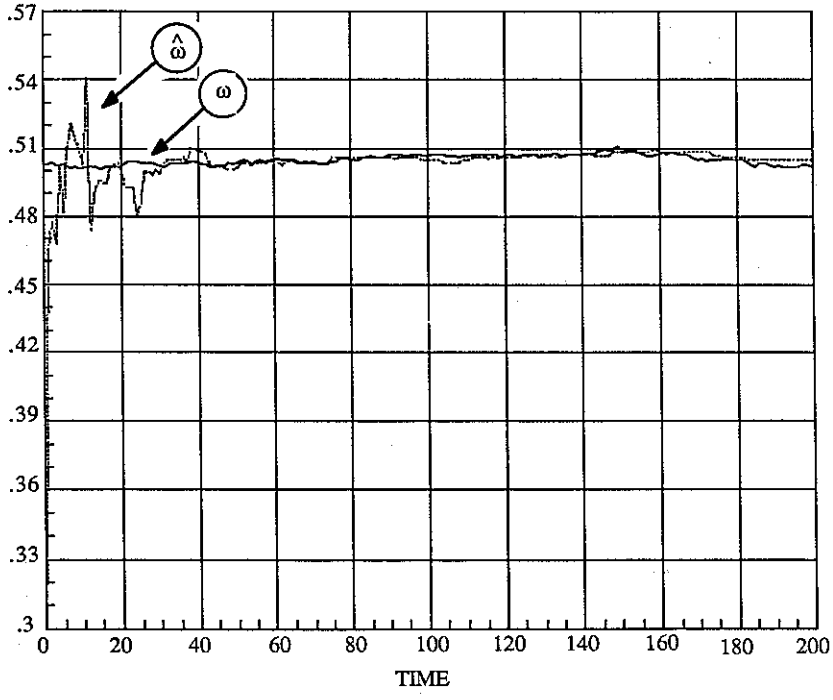

Fig. 3. True frequency and frequency estimate: tracking example.

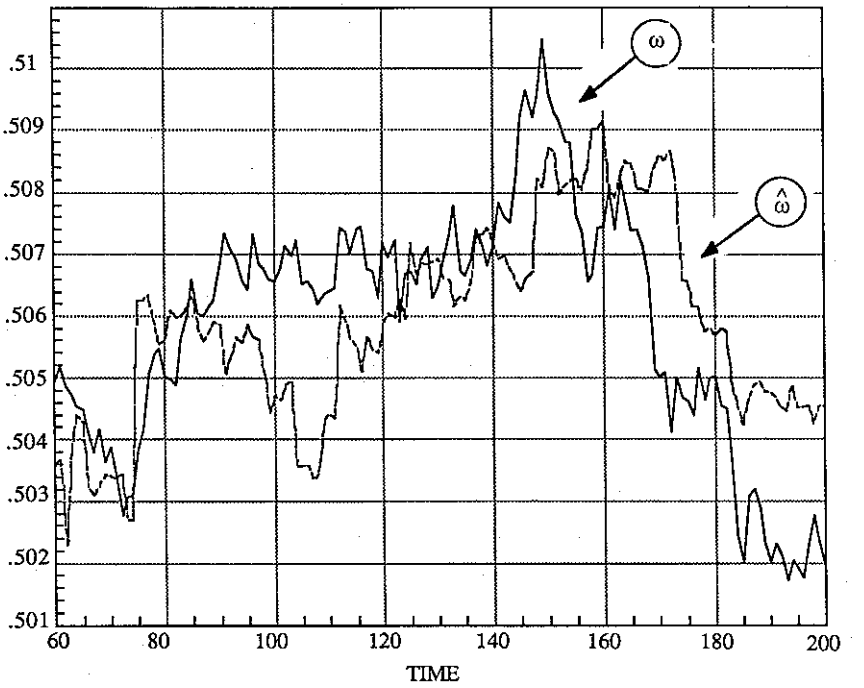

Fig. 4. The frequency and frequency estimate: expanded scale; tracking example.

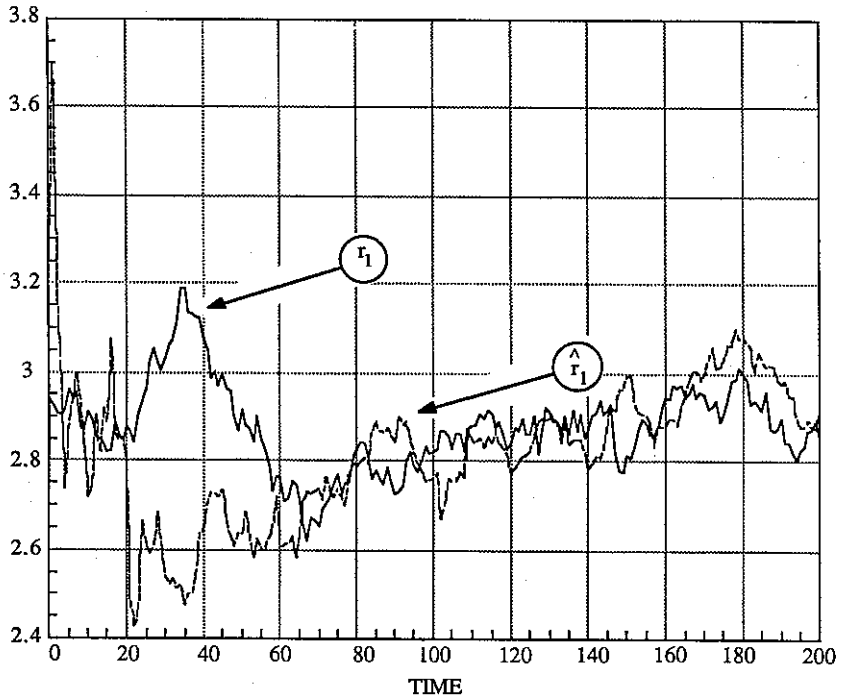

Fig. 5. Amplitude of fundamental component: true value and estimate; tracking example.

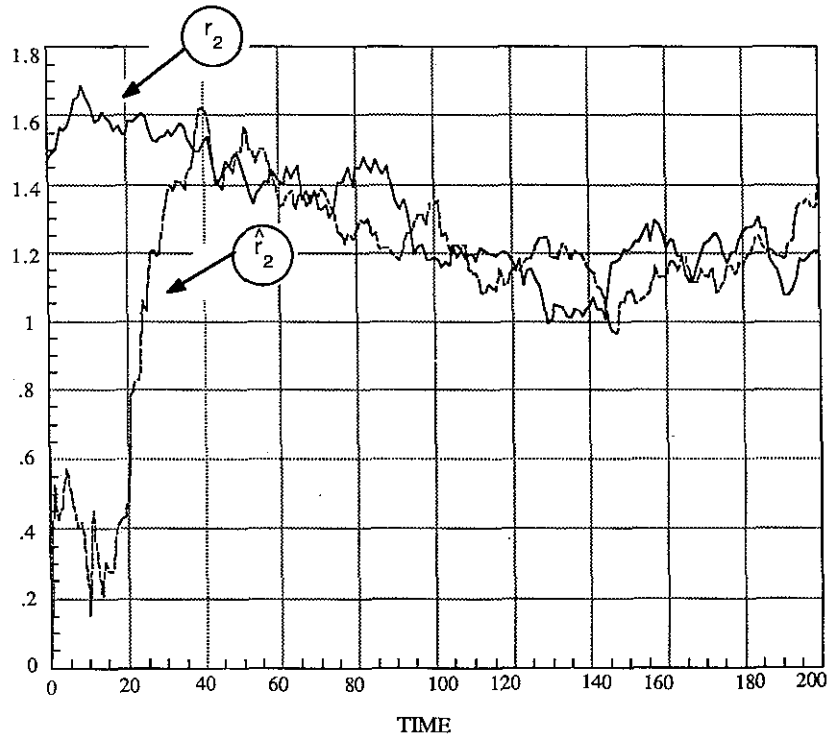

Fig. 6. Amplitude of second harmonic: true value and estimate; tracking example.

Tracking is achieved after $t = 60$. Figure 6, showing r_2 and $\hat{r}_2$, shows a similar behaviour despite an initial estimate $\hat{r}_2(0)$ of zero. The amplitudes of higher harmonics and their estimates exhibited a similar tracking response and are not shown here.

For the phases, Fig. 7 shows the error in the fundamental component $\theta_1 - \hat{\theta}_1$ while Fig. 8 shows the error in the second harmonic $\theta_2 - \hat{\theta}_2$. In both cases, there is quite a vigorous transient response, after which the error wanders fairly closely around zero radians. The transient response is worse for the second harmonic, probably because the initial estimate of the second harmonic's amplitude was zero, thus giving no initial information about the phase. For the third and higher harmonics the phase estimate error showed similar behaviour to the first and second harmonics and is, therefore, not presented here.

The graphs in Figs. 3–8 show clearly that tracking is possible in our algorithm. After an initial transient phase, the estimated parameters settle down and track the true parameters. This is a unique quality of our algorithm. Other algorithms, specifically that of Nehorai and Porat, and that of Barrett and McMahon, cannot simultaneously track frequency, phase and amplitude variations even though such variations do occur in real world applications.

The key comparison with the constant parameter signal results of the earlier part of this section concerns the frequency estimation. As indicated in Table 1, for an SNR of 8 dB and 500 samples, the standard derivation in ω was 10^{-4} (2×10^{-5} for Nehorai and Porat). In percentage terms, this is 0.2% (0.04% for Nehorai and Porat). The frequency error, once tracking has been achieved, with the nonstationary signal never exceeds 0.6%, as noted above, and the standard deviation is obviously less than this. This shows that, at least for this example, the penalty paid in securing tracking capability is modest.

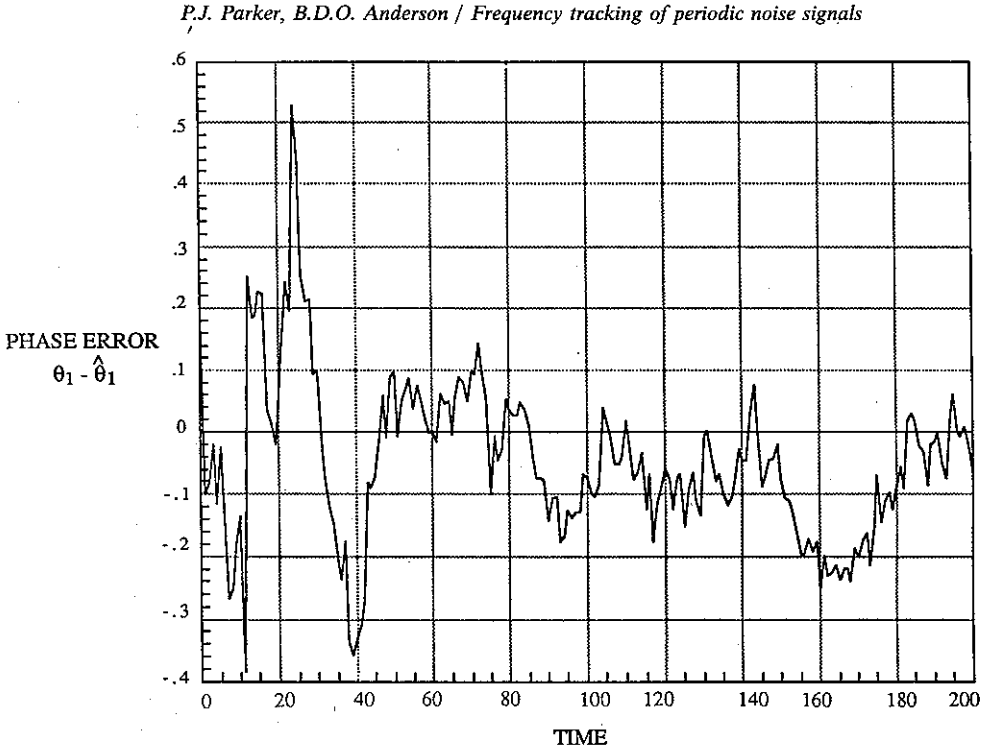

Fig. 7. Error in estimating the phase of the fundamental: tracking example.

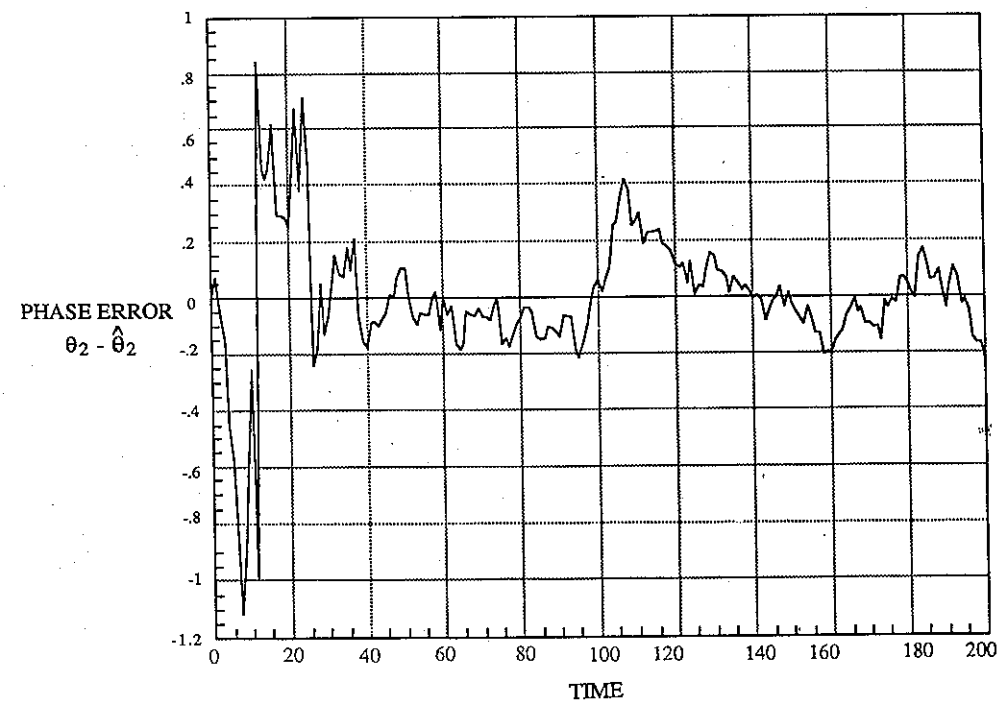

Fig. 8. Error in estimating the phase of the second harmonic: tracking example.

5. Conclusions

This paper has presented a new frequency estimation algorithm with a number of unique properties. The algorithm is based on the extended Kalman filter with its attendant ideas of mean square optimal estimates and recursive calculation. It is the only algorithm known to us that can track simultaneously amplitudes, phase and frequency variations for a signal that is quasi-periodic rather than merely sinusoidal.

With simulations we have demonstrated both the tracking behaviour and asymptotic behaviour of the estimator. In comparison with the algorithms of Nehorai and Porat which cannot easily track changing frequencies, we get good results when estimating the frequency and amplitudes of a truly periodic signal. When estimating the phase, our algorithm performed worse, but this is not an absolute conclusion and should be the subject of further investigation. Similarly, we need to further investigate the poor behaviour of our algorithm at low signal-to-noise ratios.

We have also demonstrated useful theoretical properties. The Riccati difference equation of our estimator has a solution which settles down to a nearly constant value, and this was useful in showing a pleasing separation of the amplitude and phase frequency estimations. The separation makes our algorithm significantly easier to comprehend.

A number of topics immediately suggest themselves for future investigation. First amongst these is a signal with Doppler shift. This should be an easy problem to solve merely by including an extra state in the signal model corresponding to the Doppler shift frequency, or equivalently, to the speed at which the source of the signal is approaching the receiver. It does, however, add an extra level of complexity to the algorithm and would therefore need further investigation. As was the case when estimating the fundamental frequency of the signal, the information contained in all Fourier components should be useful in estimating the Doppler shift frequency.

A second extension is to fixed lag smoothing. By allowing a certain lag in producing the frequency, amplitude and phase estimates, the accuracy of estimates should improve. Kumar, see [8], has done some significant work upon smoothers applied to phase estimation of a sinusoidal signal. His results show promises for applying smoothing to our problem.

A third extension is to the case when more than one signal is present, such as in sonar when two or more ships can be detected. Here the work of Sundresh et al. [19] and Cassara et al. [20] may be of use. These authors looked at the problem of simultaneously demodulating two different FM signals. Their estimator consisted of two interconnected phase-locked loops. For our problem, the same structure could be used, replacing the two phase-locked loops with two of our extended Kalman filter estimators.

Finally, there would certainly be a value in seeking an integrated approach, based on some combinations of the hidden Markov model ideas (with their very low SNR capability) and the ideas of this paper.

Appendix A. Proof of Theorem 3.1

We begin with a lemma.

LEMMA A.1. Suppose that

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

and

$$P(t) = \Sigma(t) - \Sigma(t)H(t)[H^{\mathrm{T}}(t)^{\mathrm{T}}\Sigma(t)H(t) + R]^{-1} \times H(t)^{\mathrm{T}}\Sigma(t). \tag{A.2}$$

Let

$$Z(t) = P(t)^{-1}.$$

Then

$$Z(t+1) = F^{-T}[Z(t) - Z(t)F^{-1}[F^{-T}Z(t)F^{-1} + Q^{-1}]^{-1}F^{-T}Z(t)]F^{-1} + H(t+1)R^{-1}H(t+1)^T. \quad (A.3)$$

REMARK. $\Sigma(t+1)$ is an error covariance associated with a one-step ahead prediction estimate; $P(t)$ is the error covariance associated with a true filtered estimator. The lemma says that $P^{-1}(t)$ satisfies a Riccati equation of the same types as $\Sigma(t)$.

PROOF OF LEMMA A.1

$$\begin{aligned} Z(t+1) &= P(t+1)^{-1} \\ &= \Sigma(t+1)^{-1} + H(t+1)R^{-1}H(t+1)^T \\ &= [FP(t)F^T + Q]^{-1} \\ &\quad + H(t+1)R^{-1}H(t+1)^T \\ &= [FZ(t)^{-1}F^T + Q]^{-1} \\ &\quad + H(t+1)R^{-1}H(t+1)^T \end{aligned}$$

and this becomes (A.3) by the matrix inversion lemma.

In the Riccati equation for $Z(t)$, the only time-varying coefficients are those associated with the additive term $H(t+1)R^{-1}H^T(t+1)$ in (A.3), which is a simpler way for the time-varying coefficients to enter the picture than that in (A.1), and is the ultimate reason for our working (temporarily) with $Z(t)$. In the next lemma, we look at the effect of replacing $H(t+1)R^{-1}H^T(t+1)$ by its low pass equivalent (the idea being that we average out the fast variations occurring at and above frequency ω), and we compare the solution of the resulting equation with $Z(t)$ by finding an equation for the difference of the solution of the two equations.

LEMMA A.2. Let $Z(t)$ be as above, and let $\bar{Z}$ satisfy the steady state equation

$$\bar{Z} = F^{-T}[\bar{Z} - \bar{Z}F^{-1}(F^{-T}\bar{Z}F^{-1} + Q^{-1})^{-1} \times F^{-T}\bar{Z}]F^{-1} + N, \quad (A.4)$$

where

$$N = \text{ave}[H(t+1)R^{-1}H(t+1)^T = \bar{H}(RI)^{-1}\bar{H}^T. \quad (A.5)$$

Let

$$\bar{F}_Z = F^{-T}[I - \bar{Z}F^{-1}(F^{-T}\bar{Z}F^{-1} + Q^{-1})^{-1}F^{-T}]. \quad (A.6)$$

Then

$$\begin{aligned} \bar{Z} - Z(t+1) &= \bar{N} - H(t+1)R^{-1}H(t+1)^T \\ &\quad + \bar{F}_Z(\bar{Z} - Z(t))\bar{F}_Z^T \\ &\quad + \bar{F}_Z(\bar{Z} - Z(t))F^{-1} \\ &\quad \times (F^{-T}Z(t)F^{-1} + Q^{-1})^{-1}F^T \\ &\quad \times (\bar{Z} - Z(t))\bar{F}_Z^T. \end{aligned} \quad (A.7)$$

PROOF

$$\begin{aligned} Z(t+1) &= H(t+1)R^{-1}H(t+1)^T \\ &\quad + F^{-T}[Z(t)^{-1} + F^{-1}QF^{-T}]^{-1}F^{-1}, \\ \bar{Z} &= N + F^{-T}(\bar{Z}^{-1} + F^{-1}QF^{-T})^{-1}F^{-1}, \\ \bar{Z} - Z(t+1) &= N - H(t+1)R^{-1}H(t+1)^T \\ &\quad + F^{-T}\{(\bar{Z}^{-1} + F^{-1}QF^{-T})^{-1} \\ &\quad - (Z(t)^{-1} + F^{-1}QF^{-T})^{-1}\}F^{-1} \\ &= N - H(t+1)R^{-1}H(t+1)^T \\ &\quad + F^{-T}[\bar{Z}^{-1} + F^{-1}QF^{-T}]^{-1} \\ &\quad \times \{Z(t)^{-1} - \bar{Z}^{-1}\} \\ &\quad \times [Z(t)^{-1} + F^{-1}QF^{-T}]^{-1}F^{-1} \\ &= N - H(t+1)R^{-1}H(t+1)^T + F^{-T} \\ &\quad \times [I - \bar{Z}F^{-1}(F^{-T}\bar{Z}F^{-1} \\ &\quad + Q^{-1})^{-1}F^{-T}]\bar{Z} \\ &\quad \times \{Z(t)^{-1} - \bar{Z}^{-1}\}Z(t) \\ &\quad \times [I - F^{-1}(F^{-T}Z(t)F^{-1} \\ &\quad + Q^{-1})^{-1}F^{-T}Z(t)]F^{-1} \\ &= N - H(t+1)R^{-1}H(t+1)^T \end{aligned}$$

$$+\bar{F}_Z(\bar{Z}-Z(t))$$
$$\times[\mathrm{I}-F^{-1}(F^{-\mathrm{T}}Z(t)F^{-1}$$
$$+Q^{-1})^{-1}F^{-\mathrm{T}}Z(t)]F^{-1}. \qquad \text{(A.8)}$$

Now observe that

$$\begin{aligned}
&[\mathrm{I}-F^{-1}(F^{-\mathrm{T}}Z(t)F^{-1}+Q^{-1})^{-1}F^{-\mathrm{T}}(Z(t)-\bar{Z}^{-1})]\\
&\quad\times[\mathrm{I}-F^{-1}(F^{-\mathrm{T}}\bar{Z}F^{-1}+Q^{-1})^{-1}F^{-\mathrm{T}}\bar{Z}]\\
&=\mathrm{I}-F^{-1}(F^{-\mathrm{T}}Z(t)F^{-1}+Q^{-1})^{-1}F^{-\mathrm{T}}(Z(t)-\bar{Z})\\
&\quad-F^{-1}(F^{-\mathrm{T}}\bar{Z}F^{-1}+Q^{-1})^{-1}F^{-\mathrm{T}}\bar{Z}\\
&\quad+F^{-1}(F^{-\mathrm{T}}Z(t)F^{-1}+Q^{-1})^{-1}\\
&\quad\times[F^{-\mathrm{T}}Z(t)F^{-1}+Q^{-1}-F^{-\mathrm{T}}\bar{Z}F^{-1}-Q^{-1}]\\
&\quad\times(F^{-\mathrm{T}}\bar{Z}F^{-1}+Q^{-1})^{-1}F^{-\mathrm{T}}\bar{Z}\\
&=\mathrm{I}-F^{-1}(F^{-\mathrm{T}}Z(t)F^{-1}+Q^{-1})^{-1}F^{-\mathrm{T}}Z(t).
\end{aligned} \qquad \text{(A.9)}$$

Inserting (A.9) in (A.8) and using (A.6) gives

$$\begin{aligned}
\bar{Z}-Z(t+1)&=\bar{N}-H(t+1)R^{-1}H(t+1)^{\mathrm{T}}\\
&\quad+\bar{F}_Z(\bar{Z}-Z(t))\\
&\quad\times[\mathrm{I}-F^{-1}(F^{-\mathrm{T}}Z(t)F^{-1}\\
&\quad+Q)^{-1}F^{-\mathrm{T}}(Z(t)-\bar{Z})]\bar{F}_Z^{\mathrm{T}}\\
&=\bar{N}-H(t+1)R^{-1}H(t+1)^{\mathrm{T}}\\
&\quad+\bar{F}_Z(\bar{Z}-Z(t))\bar{F}_Z^{\mathrm{T}}+\bar{F}_Z(\bar{Z}-Z(t))\\
&\quad\times F^{-1}(F^{-\mathrm{T}}Z(t)F^{-1}+Q^{-1})^{-1}\\
&\quad\times F^{-\mathrm{T}}(\bar{Z}-Z(t))\bar{F}_Z^{\mathrm{T}}.
\end{aligned}$$

In the next lemma, we relate $\bar{F}_Z$ to $\bar{F}$, defined in (3.6).

LEMMA A.3. With quantities defined as earlier, there holds

$$\bar{\Sigma}^{-1}\bar{F}\bar{\Sigma}=\bar{F}_Z.$$

PROOF. With $\bar{P}=\bar{\Sigma}-\bar{\Sigma}\bar{H}(\bar{H}^{\mathrm{T}}\bar{\Sigma}\bar{H}+R\mathrm{I})^{-1}\bar{H}\bar{\Sigma}$, there holds $\bar{\Sigma}=F\bar{P}F^{\mathrm{T}}+Q$ and $\bar{F}\bar{\Sigma}=F\bar{P}$, so that

$$\begin{aligned}
\bar{\Sigma}^{-1}\bar{F}\bar{\Sigma}&=(F\bar{P}F^{\mathrm{T}}+Q)^{-1}\bar{F}\bar{\Sigma}\\
&=(F\bar{Z}^{-1}F^{\mathrm{T}}+Q)^{-1}F\bar{P}\\
&=F^{-\mathrm{T}}[\bar{Z}-\bar{Z}F^{-1}(F^{-\mathrm{T}}\bar{Z}F^{-1}+Q^{-1})^{-1}F^{-\mathrm{T}}Z]F^{-1}F\bar{P}\\
&=F^{-\mathrm{T}}[\mathrm{I}-\bar{Z}F^{-1}(F^{-\mathrm{T}}\bar{Z}F^{-1}+Q^{-1})^{-1}F^{-\mathrm{T}}]\\
&=\bar{F}_Z. \quad \square
\end{aligned}$$

Now we can see that $\bar{Z}-Z(t)$ is small. For observe that (A.7), a recursive equation for $\bar{Z}-Z(t)$, has three terms on the right-hand side. The first is a forcing term and is high pass in character, due to the definition of N. The second is a linear term in $\bar{Z}-Z(t)$, which is multiplied by $\bar{F}_Z$ and $\bar{F}_Z^{\mathrm{T}}$ which in our particular model is a low pass state transition matrix, due to Lemma A.3 and the theorem hypothesis. The third term is nonlinear in $\bar{Z}-Z(t)$, in fact it is quadratic in this term. Were the nonlinear term absent, the response as $t\to\infty$ would necessarily be small, like the response of any low pass system driven by a high pass signal. Now the equation can be represented as in Fig. 9.

The theory which relates nonlinear systems to their linearizations, see e.g. [5], guarantees that if the non-linear feedback does not have too high a gain $\bar{Z}-Z(t)$ will remain small.

Also, provided $\bar{Z}$ and $Z(t)$ are not too near to being singular, $\bar{Z}-Z(t)$ small will give $\bar{P}-P(t)$ small, where $\bar{P}=\bar{Z}^{-1}$. Consequently $\bar{\Sigma}-\Sigma(t)$ will be small. Further, $\bar{\Sigma}-\Sigma(t)$ will tend to be high pass in character. Roughly, neglecting the quadratic term in (A.7) shows that $\bar{Z}-Z(t)$ is high pass in character, and from this one can argue (with another layer of approximation) that $\bar{\Sigma}-\Sigma(t)$ will be high pass.

REMARK. The immediately preceding argument has been somewhat imprecise, but the effort of making it more exact, with a precise definition of norms, seems hardly justified in this case.

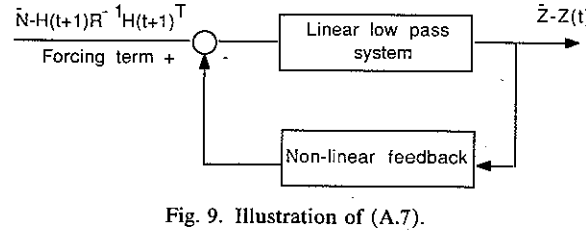

Fig. 9. Illustration of (A.7).

REMARK. Since N depends upon the $\hat{r}_i$, $\bar{\Sigma}$ will also depend upon the $\hat{r}_i$, and will thus vary as the $\hat{r}_i$ vary. However, it is one of our assumptions that the r_i vary slowly with t, and so we would expect, in turn, $\hat{r}_i$ and $\bar{\Sigma}$ to vary slowly.

With the assumptions that Q is block diagonal of the form stated in the theorem, $\bar{\Sigma}$ will also have the required form by virtue of its definition in (3.5).

This completes the proof of Theorem 3.1. □

References

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

[2] R.F. Barrett and D.R.A. McMahon, "ML estimation of the fundamental frequency of a harmonic series", *Proc. ISSPA 87*, August 1987, Brisbane, Australia, pp. 333–336.

[3] D.V. Bhaskar Rao and S.Y. Kung, "Adaptive notch filtering for retrieval of sinusoids in noise", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 32, August 1984, pp. 791–802.

[4] F.A. Cassara, H. Schachter and G.H. Simowitz, "Acquisition behaviour of the cross-coupled phase-locked loop FM demodulator", *IEEE Trans. Comm.*, Vol. 28, June 1980, pp. 897–904.

[5] C.A. Desoer and M. Vidyasagar, *Feedback Systems: Input–Output Properties*, Academic Press, New York, 1975.

[6] B. Friedlander and J.O. Smith, "Analysis and performance evaluation of an adaptive notch filter", *IEEE Trans. Inform. Theory*, Vol. 30, March 1984, pp. 283–295.

[7] C.N. Kelly and S.C. Gupta, "Discrete time demodulation of continuous-time signals", *IEEE Trans. Inform. Theory*, Vol. IT-18, July 1972, pp. 488–493.

[8] R. Kumar, *Optimum filters and smoothers design for carrier phase and frequency tracking, NASA JPL Publication 87–10*, Jet Propulsion Laboratory, Pasadena, California, 1987.

[9] N.K. M'Sirdi et al., "Adaptive comb filters: Implementation for harmonic signal", *Signal Processing: Theory and Applications*, North-Holland, EURASIP, Eusipco, Grenoble, France, September 1988.

[10] N.K. M'Sirdi, H.A. Tjokronegoro and I.E. Landau, "Retrieval of signals in with cascaded notch filters: An RML algorithm", *Proc. Internat. Conf. Acoust. Speech Signal Process. 1988*, New York.

[11] N.K. M'Sirdi and I.D. Landau, "Adaptive evolutionary spectrum analysis for narrow band signals", *Proc. Internat. Conf. Acoust. Speech Signal Process. 1988*, Dallas, Texas.

[12] N.K. M'Sirdi and H.A. Tjokronegoro, "Cascaded adaptive notch filters: An RML algorithm", in: *Signal Processing IV: Theory and Applications*, North-Holland, EURASIP, Eusipco, Grenoble, France, September 1988.

[13] A. Nehorai, "A minimal parameter adaptive notch filter with constrained poles and zeroes", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 33, August 1985, pp. 983–996.

[14] A. Nehorai and B. Porat, "Adaptive comb filtering for harmonic signal enhancement", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 34, October 1986, pp. 1124–1138.

[15] T. S. Ng, "Some aspects of an adaptive digital notch filter with constrained poles and zeroes", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 35, February 1987, pp. 158–161.

[16] V.S. Pisarenko, "The retrieval of harmonics from a covariance function", *Geophys. J. Ray Astron. Soc.* Vol. 33, 1973, pp. 347–366.

[17] D.L. Snyder, *The State Space Approach to Continuous Estimation with Application to Analog Communication Theory*, 1969, MIT Press, Cambridge, MA.

[18] R.L. Streit and R.F. Barrett, "Frequency line tracking using hidden Markov models", *IEEE Trans. Acoust. Speech Signal Process.*, submitted.

[19] T.S. Sundresh, F.A. Cassara and H. Schachter, "Maximum a posteriori estimator for suppression of interchannel interference in FM receivers", *IEEE Trans. Comm.*, Vol. 28, December 1980, pp. 1480–1485.

[20] H.L. Van Trees, *Detection, Estimation and Modulation Theory, Part II*, John Wiley, New York, 1971.

[21] B. Widrow and J.R. Glover Jr. et al. "Adaptive noise cancelling: Principles and applications", *Proc. IEEE*, Vol. 63, December 1975, pp. 1692–1716.